CRISPR-CAS NICKASE SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION IN EUKARYOTES

Applicants: The Broad Institute, Inc., Cambridge, MA (US); Massachusetts Institute of Technology, Cambridge, MA (US)

Inventor: Feng Zhang, Cambridge, MA (US)

Assignees: The Broad Institute Inc., Cambridge, MA (US); Massachusetts Institute of Technology, Cambridge, MA (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

Appl. No.: 14/183,471
Filed: Feb. 18, 2014

Prior Publication Data
US 2014/0234972 A1 Aug. 21, 2014

Related U.S. Application Data
Continuation application No. 14/105,031, filed on Dec. 12, 2013.


Int. Cl.
C12Q 1/68 (2006.01)
C12N 15/85 (2006.01)
C12N 15/63 (2006.01)
C12N 9/14 (2006.01)
C12N 9/22 (2006.01)
C12N 9/52 (2006.01)
C12N 15/00 (2006.01)
C07H 21/02 (2006.01)
C07H 21/04 (2006.01)
A61K 38/43 (2006.01)
A61K 38/46 (2006.01)
A61K 38/47 (2006.01)
C12N 13/10 (2006.01)

U.S. Cl.
CPC C12N 15/85 (2013.01); C12N 15/63 (2013.01); C12N 9/22 (2013.01); C12N 15/1082 (2013.01)
USPC 435/6.1; 1435/6.13; 435/195; 435/199; 435/220; 435/320; 424/94.1; 424/94.16; 424/54.61; 536/22.1; 536/23.1; 536/23.2; 536/23.7; 536/24.1

Field of Classification Search
None
See application file for complete search history.

References Cited
U.S. PATENT DOCUMENTS
2013/0130248 A1 5/2013 Haurwitz et al.

FOREIGN PATENT DOCUMENTS
WO 2008/108989 9/2008
WO 2010/054108 5/2010
WO 2012/165456 12/2012
WO 2013/098244 7/2013
WO 2013/141680 * 9/2013 C12N 15/10
WO 2013/142578 * 9/2013 C12N 15/10
WO 2013/176772 11/2013
WO 2014/065506 5/2014
WO 2014/089290 6/2014
WO 2014/099744 6/2014

OTHER PUBLICATIONS

Primary Examiner — Anne Gusso
Assistant Examiner — Nancy J Leith
Attorney, Agent, or Firm — Vedder Price P.C.; Thomas J. Kowalski; Deborah L. Lu

ABSTRACT
The invention provides for systems, methods, and compositions for manipulation of sequences and/or activities of target sequences. Provided are vectors and vector systems, some of which encode one or more components of a CRISPR complex, as well as methods for the design and use of such vectors. Also provided are methods of directing CRISPR complex formation in eukaryotic cells and methods for selecting specific cells by introducing precise mutations utilizing the CRISPR/Cas system.

30 Claims, 116 Drawing Sheets
References Cited

OTHER PUBLICATIONS


* cited by examiner
FIG. 2B
FIG. 2D
Target locus  
5'...AGCTGGAGGAGGAAGGTCCTGAGTCCGAGCAAGAGAAGGGCTCCCAC...3'  
3'...TCGACCTCCTCCCTCCCGAACCTAGGCTCGTCTCTCTCTCTCTCTCCCAGGGGTG...5'

crRNA  
5'...GAGUCGAGCAGAAAGAAGAGUGUUUUGAGC...3'

indel  
AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAAGAAGAGGTGGCTCCCAGAT

human EMX1 protospacer target (mutation in 5 of 43 sequenced clones = 11.6%)  
WT  
5'...CTGGAGGAGGAAGGGCCCTGAGTCCGAGCAAGAAGAGGGCTCCCACAT...3'  
Δ1  
CTGGAGGAGGAAGGGCCCTGAGTCCGAGCAAGAAGAGGGCTCCCACAT  
+1  
CTGGAGGAGGAAGGGCCCTGAGTCCGAGCAAGAAGAGGGCTCCCACAT  
Δ3  
CTGGAGGAGGAAGGGCCCTGAGTCCGAGCAAGAAGAGGGCTCCCACAT  
m1, Δ6  
CTGGAGGAGGAAGGGCCCTGAGGCCCAGAAGAAGAGGGCTCCCACAT

FIG. 2E-F
**Fig. 4A-C**

### A

**Human EMX1 Locus**

**Protospacer (1)**

<table>
<thead>
<tr>
<th>Human EMX1 Locus</th>
<th>Protospacer (1)</th>
<th>PAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'...CTGAGGAGGAAGGGCTGAGCTCCGAGCGAGAAAAGAGGGCTCCCAT...-3'</td>
<td>5'...AGGUCGAGCAAGAAGGAAGUUGAGAGGC...-3'</td>
<td></td>
</tr>
<tr>
<td>3'...GACCTCCCTTTCCCGGCGAGAGGCTCTCTTTCTTCTTTCGCGGGGTA...-5'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**WT crRNA**

- m1: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m2: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m3: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m4: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m5: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m6: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m7: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m8: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m9: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m10: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC

**Mismatch-Containing Guide Sequences**

- m11: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m12: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m13: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m14: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m15: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC
- m16: GAGUCCGAGCAAGAAGGAAGUUGAGAGGC

### B

**Mismatched Spacers**

- m17
- m15
- m13
- m11
- m9
- m7
- m5
- m3
- m1

**Indel (%)**

- 5.6
- 7.5
- 8.8
- 9.7

### C

**Left TALEN Binding Site**

**Protospacer (1)**

<table>
<thead>
<tr>
<th>Human EMX1 Locus</th>
<th>Protospacer (1)</th>
<th>PAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'...CTGAGGAGGAAGGGCTGAGCTCCGAGCGAGAAAAGAGGGCTCCCAT...-3'</td>
<td>5'...AGGUCGAGCAAGAAGGAAGUUGAGAGGC...-3'</td>
<td></td>
</tr>
<tr>
<td>3'...GACCTCCCTTTCCCGGCGAGAGGCTCTCTTTCTTCTTTCGCGGGGTA...-5'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Right TALEN Binding Site**
FIG. 4D

<table>
<thead>
<tr>
<th></th>
<th>TALEN</th>
<th>Chimeric RNA</th>
<th>crRNA</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>684bp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>367bp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>317bp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>indel (%)</td>
<td>3.6</td>
<td>3.8</td>
<td>4.5</td>
<td>11</td>
</tr>
</tbody>
</table>
C

human EMX1 locus

HR Template

HindIII, Nhel

D

hSpCas9 + - -
hSpCas9n - + -
HR template + + +

2281bp

1189bp

1092bp

HR (%) 0.70 0.46

E

AATGACAAGCTTGCTAGCGGTGGG

HindIII  Nhel

FIG. 5C-E
G

crRNA array design

EMX1(1) DR DR EMX1(8) DR

5'...AAAAAGGAGGAGCCCTGAGTCAGCTGAGAGAGAAAGTT...AACGGAGGAGGAGGCAACAGATGAGAAACATCAAGGTTTAG...3'

human EMX1 locus

protospacer (8)
PAM

3'...TCGGGAAGAAGAGACGAGCTAGTCGGGAAGAGAGG...GAAGGGTTCCTCCTCCTCCTCCTGCTACTCTTTTGAAGCTCTTCGGGGG...5'

deletion result

5'...GGCAATGGGCTACACCGTTGATGATGGAGGCCCTCTCTTAGGAAGGCCCCAGAGCAGCACCCTCTGAGGCTCAACACTCAGGC...3'

FIG. 5G
<table>
<thead>
<tr>
<th>Cas9 species</th>
<th>gene</th>
<th>protospacer ID</th>
<th>protospacer sequence (5' to 3')</th>
<th>PAM</th>
<th>strand</th>
<th>cell line tested</th>
<th>% indel (pre-crRNA + IracrRNA)</th>
<th>% indel (chimeric RNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>EMX1</td>
<td>1</td>
<td>GGAAGGCCCTGAGTCCGAGAGAACGAAAAAGT</td>
<td>GGG</td>
<td>+</td>
<td>293FT</td>
<td>20 ± 1.8</td>
<td>6.7 ± 0.62</td>
</tr>
<tr>
<td></td>
<td>EMX1</td>
<td>2</td>
<td>CATGGAGATGACATCGATGGTCTCCCTCCCAT</td>
<td>TGG</td>
<td>-</td>
<td>293FT</td>
<td>2.1 ± 0.31</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>EMX1</td>
<td>3</td>
<td>GGCCTGCAGACGCTCCCTCCCTCCCTCCCTAG</td>
<td>TGG</td>
<td>+</td>
<td>293FT</td>
<td>14 ± 1.1</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>EMX1</td>
<td>4</td>
<td>CATGGATGACATCGATGGTCTCCCTCCCTCCCTAG</td>
<td>TGG</td>
<td>-</td>
<td>293FT</td>
<td>11 ± 1.7</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>EMX1</td>
<td>5</td>
<td>TCGCTGACAGATGACGCTCCCTCCCTCCCTCCCTAG</td>
<td>TGG</td>
<td>-</td>
<td>293FT</td>
<td>4.3 ± 0.46</td>
<td>2.1 ± 0.51</td>
</tr>
<tr>
<td></td>
<td>EMX1</td>
<td>6</td>
<td>TCGCTGACAGATGACGCTCCCTCCCTCCCTCCCTAG</td>
<td>TGG</td>
<td>-</td>
<td>293FT</td>
<td>4.0 ± 0.66</td>
<td>0.41 ± 0.25</td>
</tr>
<tr>
<td></td>
<td>EMX1</td>
<td>7</td>
<td>TCGCTGACAGATGACGCTCCCTCCCTCCCTCCCTAG</td>
<td>TGG</td>
<td>-</td>
<td>293FT</td>
<td>1.5 ± 0.12</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>EMX1</td>
<td>8</td>
<td>AGGGAGAGGGCAGACGATAACGAAACTCCAGG</td>
<td>AGG</td>
<td>-</td>
<td>293FT</td>
<td>7.8 ± 0.83</td>
<td>2.3 ± 1.2</td>
</tr>
<tr>
<td>S. pyogenes SF370 type II CRISPR</td>
<td>PVALB</td>
<td>9</td>
<td>AGGGAGAGGGCAGACGATAACGAAACTCCAGG</td>
<td>AGG</td>
<td>+</td>
<td>293FT</td>
<td>21 ± 2.6</td>
<td>6.5 ± 0.32</td>
</tr>
<tr>
<td></td>
<td>PVALB</td>
<td>10</td>
<td>AGGGAGAGGGCAGACGATAACGAAACTCCAGG</td>
<td>AGG</td>
<td>+</td>
<td>293FT</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>PVALB</td>
<td>11</td>
<td>AGGGAGAGGGCAGACGATAACGAAACTCCAGG</td>
<td>AGG</td>
<td>-</td>
<td>293FT</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>Th</td>
<td>12</td>
<td>CAACACTGAGATGACATCGATGCTAAATGCT</td>
<td>AGG</td>
<td>-</td>
<td>Neuro2A</td>
<td>27 ± 4.3</td>
<td>4.1 ± 2.2</td>
</tr>
<tr>
<td></td>
<td>Th</td>
<td>13</td>
<td>CAACACTGAGATGACATCGATGCTAAATGCT</td>
<td>GGG</td>
<td>-</td>
<td>Neuro2A</td>
<td>4.8 ± 1.2</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>Th</td>
<td>14</td>
<td>CAACACTGAGATGACATCGATGCTAAATGCT</td>
<td>AGG</td>
<td>+</td>
<td>Neuro2A</td>
<td>11.3 ± 1.3</td>
<td>N.D.</td>
</tr>
<tr>
<td>S. thermophilus Homo sapiens LMD-9 CRISPR</td>
<td>EMX1</td>
<td>15</td>
<td>GAGGGAGGTAAGTATACGAAACACGAACTGAAAT</td>
<td>GTGAAT</td>
<td>-</td>
<td>293FT</td>
<td>14 ± 0.88</td>
<td>N.T.</td>
</tr>
<tr>
<td></td>
<td>EMX1</td>
<td>16</td>
<td>GAGGGAGGTAAGTATACGAAACACGAACTGAAAT</td>
<td>GTGAAT</td>
<td>-</td>
<td>293FT</td>
<td>7.8 ± 0.77</td>
<td>N.T.</td>
</tr>
</tbody>
</table>

**FIG. 6**
A

short tracrRNA (88bp)

northern blot probe target

long tracrRNA (171bp)

B

<table>
<thead>
<tr>
<th></th>
<th>SpCas9</th>
<th>long tracrRNA</th>
<th>short tracrRNA</th>
<th>SpRNase III</th>
<th>DR-EMXf(1)-DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

indel (%)  0.60  14  11

FIG. 7A-B
<table>
<thead>
<tr>
<th></th>
<th>~ SpRNase III</th>
<th>+ SpRNase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpCas9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>long tracrRNA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>short tracrRNA</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>DR-EMX1(1)-DR</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

```

# FIG. 7C

```
% indel = \left( 1 - \sqrt{1 - \frac{(a + b)}{(a + b + c)}} \right) \times 100
FIG. 9A-B
<table>
<thead>
<tr>
<th>Chr</th>
<th>NGG median</th>
<th>NGG mean</th>
<th>NNAGAAC median</th>
<th>NNAGAAC mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>12.8</td>
<td>67</td>
<td>115.8</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>13.0</td>
<td>64</td>
<td>103.8</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8</td>
<td>63</td>
<td>98.5</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>9.0</td>
<td>61</td>
<td>94.5</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>13.0</td>
<td>63</td>
<td>97.9</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>13.0</td>
<td>63</td>
<td>98.5</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>12.4</td>
<td>64</td>
<td>102.0</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>12.8</td>
<td>64</td>
<td>100.9</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>13.9</td>
<td>65</td>
<td>102.0</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>12.1</td>
<td>66</td>
<td>107.0</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>12.0</td>
<td>65</td>
<td>103.5</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>13.6</td>
<td>62</td>
<td>94.6</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>12.4</td>
<td>64</td>
<td>103.5</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>11.5</td>
<td>68</td>
<td>107.7</td>
</tr>
<tr>
<td>15</td>
<td>7</td>
<td>11.7</td>
<td>74</td>
<td>126.8</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>10.3</td>
<td>76</td>
<td>127.9</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>9.4</td>
<td>82</td>
<td>145.4</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
<td>9.4</td>
<td>82</td>
<td>145.4</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>11.1</td>
<td>72</td>
<td>121.8</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>13.4</td>
<td>64</td>
<td>114.0</td>
</tr>
<tr>
<td>21</td>
<td>7</td>
<td>13.4</td>
<td>63</td>
<td>140.3</td>
</tr>
<tr>
<td>22</td>
<td>8</td>
<td>13.2</td>
<td>63</td>
<td>99.0</td>
</tr>
<tr>
<td>23</td>
<td>8</td>
<td>29.2</td>
<td>62</td>
<td>223.7</td>
</tr>
</tbody>
</table>

**FIG. 11A-C**

A: Streptococcus pyogenes SRF70 type II CRISPR PAM occurrence in human genome (NGG)

B: Streptococcus thermophilus MD8 CRISPR PAM occurrence in human genome (NNAGAAC)
A

U6  DR  EMXY(1)  DR

northern blot probe target

5'-.TGAATGGTCCAAAAACGGAATGGGCCTGAGTCCGARCAAGAGAAGTGGTTTTAAGCTATGCTGTCTTGGATGC...-3'

B

<table>
<thead>
<tr>
<th></th>
<th>- SpRNase III</th>
<th>+ SpRNase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpCas9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>long tracrRNA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>short tracrRNA</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>DR-EMXY(1)-DR</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

FIG. 14A-B
FIG. 15 A-B
FIG. 16

human EMX1 locus

5' - TTTCTAGTGCTGAGTTCTGTGACTCCTACATTACTCTCTTGTTCTGTATACCTCCTCC - 3'

3' - AAAAGATCGACCTCAAGAGACTTGAGGATGTAAGCATGAGAGACAAGACATATGATGAGGAGG - 5'

protospacer (16)  PAM  protospacer (15)
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Assay</th>
<th>Genomic Target</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp-EMX1-F</td>
<td>SURVEYOR assay, sequencing</td>
<td>EMX1</td>
<td>AAAACCACCCCTCTCTCTGGC</td>
</tr>
<tr>
<td>Sp-EMX1-R</td>
<td>SURVEYOR assay, sequencing</td>
<td>EMX1</td>
<td>GGAAGATGGAGACACGGAGAG</td>
</tr>
<tr>
<td>Sp-PVALB-F</td>
<td>SURVEYOR assay, sequencing</td>
<td>PVALB</td>
<td>CTGGAAGGCAATGCCTGAC</td>
</tr>
<tr>
<td>Sp-PVALB-R</td>
<td>SURVEYOR assay, sequencing</td>
<td>PVALB</td>
<td>GGCAGCAAAACTCTTTGCTCT</td>
</tr>
<tr>
<td>Sp-Th-F</td>
<td>SURVEYOR assay, sequencing</td>
<td>Tb</td>
<td>GTGCTTTGCAAGAGCCTACC</td>
</tr>
<tr>
<td>Sp-Th-R</td>
<td>SURVEYOR assay, sequencing</td>
<td>Tb</td>
<td>CCTGAGCGCATGCAAGTA</td>
</tr>
<tr>
<td>St-EMX1-F</td>
<td>SURVEYOR assay, sequencing</td>
<td>EMX1</td>
<td>ACCTTCTGTGTTCACCCATTTC</td>
</tr>
<tr>
<td>St-EMX1-R</td>
<td>SURVEYOR assay, sequencing</td>
<td>EMX1</td>
<td>TGGGGAGTGACAGACTTC</td>
</tr>
<tr>
<td>Sp-EMX1-RFLP-F</td>
<td>RFLP, sequencing</td>
<td>EMX1</td>
<td>GCCCTCCCTGGGTCTAAAGTA</td>
</tr>
<tr>
<td>Sp-EMX1-RFLP-R</td>
<td>RFLP, sequencing</td>
<td>EMX1</td>
<td>AGAGGCTGATCTGCTGTA</td>
</tr>
<tr>
<td>Pb_EMX1_sp1</td>
<td>Northern Blot Probe</td>
<td>Not applicable</td>
<td>TAGCTCTAAAACCTTTCTCTGCTGGAC</td>
</tr>
<tr>
<td>Pb_tracrRNA</td>
<td>Northern Blot Probe</td>
<td>Not applicable</td>
<td>CTAGCCTTTTTTAACTTGATGCTGTTT</td>
</tr>
</tbody>
</table>

FIG. 17
FIG. 18A
C

protoscaler 4 (PVALB)  
protoscaler 5 (PVALB)

<table>
<thead>
<tr>
<th>crRNA</th>
<th>+48</th>
<th>+54</th>
<th>+67</th>
<th>+85</th>
<th>crRNA</th>
<th>+48</th>
<th>+54</th>
<th>+67</th>
<th>+85</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

chimeric RNA

18.3  49.9  
25.1  39.7

FIG. 18C
Protospacer 1
(EMX1)

Protospacer 2
(EMX1)

Protospacer 3
(EMX1)

Protospacer 4
(PVALB)

Protospacer 5
(PVALB)

FIG. 21
a) Targeting construct

b) Editing construct

<table>
<thead>
<tr>
<th>Editing Construct</th>
<th>cfu/ug of crR6M DNA</th>
<th>Ed. Un.</th>
<th>Kan^r transformants</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>2.6 ± 0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6(srtA)</td>
<td>14 ± 1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6^{370.1}</td>
<td>25 ± 2.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIG. 23A-B
FIG. 25A-F
FIG. 26A-D
FIG. 27A-D
FIG. 28A-B
FIG. 30A-C
FIG. 31
FIG. 33A-D
FIG. 35A-C
FIG. 36
FIG. 37

Antibiotic-resistant cfu / total cfu

- pCRISPR::Ø
- pCRISPR::rpsL

Strep$^R$ 1.0E-7
Kan$^R$ 8.1E-2

1.4E-5

10^{-8} 10^{-6} 10^{-4} 10^{-2} 10^{0} 1
FIG. 38A-B
FIG. 41B
hSpCas9
5' CGGCTGGAAAAATCTGATGCGCCACGTGCGCGCGGAGAAAGAATGGGCTGTGCTGGCAAC
  720

  ++++++++RLENLIAQLPGEKKNQGLFGN
  221 222 223 224 225 226 229 230 231 232 233 234 235 236 237 238 239 240

5' CTGATT90CCTGAGCCCTGCGGCGCGCGCCACCTCTAGAGCCGACATCGCTGCACCTGGCCGAG
  780

  ++++++++LIALLSLGLTTPNFKNFSNFDFDLAE
  241 242 243 244 245 246 247 248 249 250 252 253 254 256 257 258 259 260

5' GATGCCAACCCTGACGTGAGCAGGCAACGCAACCTACGACGACGACCGGTGCAACCGCCTGGCC
  840

  ++++++++DKLQLSKKDTYDDDDDLONLLAA
  261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280

5' CAGATCGGGCCAGTACGCGCCCTCTCTCTGCGCCGAAAGGACCTCTGCGAGCTGCCCT
  900

  ++++++++GIGDQYADLFLLAANLSDAIL
  281 282 283 284 285 286 287 288 289 290 291 292 294 295 296 297 298 299 300

5' CTGCCTGGGCAGATCCCTGAGGTGARACCGAGTACAACGGGCGCCTCGGTGCGCTCT
  960

  ++++++++LLSDLVNTKAPLSSAS
  301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320

5' AT1GATCAAGAGAATAGACGGACACCAACAGGAGCTGACCTGACCTGAAAGCTCTGCTGGG
  1020

  ++++++++MIKRYDEHHQODILKLKALYR
  321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340

FIG. 41C
FIG. 41D
FIG. 41F
FIG. 41G
FIG. 41J
FIG. 41K
FIG. 41M
FIG. 43A
hEMX1 locus

genomic target ↓ PAM

5'...CAGAAGAAGAAGGGC...CCAATGGGGAGGACATCGATGTCACTCTCCAATGACTTAGGTTGGCGAAC...CTCTGGCCAATCCCT...-3'

3'...GTCTTCTTTCTTCGGG...GGTTACCCCCCTCTGTAGCTACAGTGAGTTACTGATCCACACCACCCTGGTT...GAGACCAGTGAGGGA...-5'

sense 5'...CAGAAGAAGAAGGGC...ACATCGATGTCACTCTCCAATGACAAGCTTGCTAGCGGTTGGGCAACCACAAAO...CTCTGGCCAATCCCT...-3'

antisense 3'...GTCTTCTTTCTTCGGG...TGTAGCTACAGTGAGTTACTGTTCAACGATGCCACCCCGTTGGTGTTT...GAGACCAGTGAGGGA...-5'

FIG. 43C
d

<table>
<thead>
<tr>
<th>Cas9:</th>
<th>HEK 293FT</th>
<th>hESC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wt</td>
<td>wt</td>
</tr>
<tr>
<td></td>
<td>D10A</td>
<td>D10A</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

indel (%): 23 3.8

FIG. 43D
FIG. 44B
FIG. 45
indel formation

homologous recombination

FIG. 47
A

EMX1 target

50bp

hEMX1 genomic locus

50bp left homology

50bp right homology

Antisense or Sense Oligo

HindIII

B

<table>
<thead>
<tr>
<th></th>
<th>SpCas9</th>
<th>-</th>
<th>+</th>
<th>+</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpCas9n</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Antisense-Oligo</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sense-Oligo</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HR template</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

2281bp ▶

1182bp ▶

1099bp ▶

HR (%): 0.41 0.27 1.1 0.43 0.073

FIG. 48A-B
pCAG
ccttaaacatcctgcatgctacatagtaatagtaatcatactagcgcgtcgacgtggctcagacagagtttctgccgtcaacgacagtacgctgatcgtgctgtgaagctcttctctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
FIG. 50B
GTGGACCATATCTGTGCCTACAGACTTCTTCTGAAGGACGACTCCATCGACAAACAAGGTTG
CTGACCAAGAAGCGAAGAAGACTCTCTGGCGGACAGTCGACCTGGAAGCCCAAGCTGATTCC
CAGAGAAGATTGCAGCCTAAGCTGACAGGATGATGACGATCCCTCGCAAGATGAGGATCTG
ATAAGGCGCCGGCCTCTATCAAGAGACGACGACGATGAGGATGATGACGATCGCCCTGGGA
ATGCGACAGATCTGCGACTCCCGGATGAACGACTGAGCAGGAGGACGAGGAGGAACGCCAC
GATCGACCTCGAAGCGGCCTGTGGGAACCGCCCTGTACTAAAAGTACCTTACAGTGG
AAGCGAGGTTCTGTGACGCGCACTAACAAGGTGTACGATCGGGAAGATGATCGCCA
AGAGCGAGCAAGAAATCGGCAGAAGCTACGCCAAGATCTTTTCCTACTAGGAAAAAT
ATGAACCTTTCAAGAGCAGAATTATCCCGCAAGCGCAATCGGGAGAGTGCCTGGAACAGGCGCGA
TCTGATACGAGAAGACGGGGAACCGGGGAGATGCTGTCGGATGAAAGGCGGGGATT
TTCGACCGTTGCGGAAAGTGCTGAGAAGATGCTGAGATCCCCAAAGTGAATATGCTGAAAAGAC
GAGGGTGCAACAGGCCTCCTTACGAAAGATGCTATCCTCTGCGCAAGAGAAGCTG
GCTTCAGGAAAGAATCCCAATCCAGCTTCTTGGAAGCCCAAGGCTACAAAAGAAAGTGAAA
AAGGACCTGATCATCAGTGCTCTCCTAGATCTCCCTCTGCTGACGTTGAACACGGCC
GAAGAGAAATGCTGGCTTCGCTCCCGCCGCAACTGGAAGGAAACGAACTTGGCCCTGC
CCTCCAAATATGGAACCTCTCTGCATCTGCGACGCACTATGGAAGAGGCGAGCT
CCCCGAGGATATGAGCAGAACAGCTGTTCTGTGAACAGCACAACGACTACCTG
GAGCAGATCTGATCAGTCAGCGATGTCTCTCTCCAAAGAAGGATGATGTCGTCG
TAAATCTGGCAAAAAGTTGCTGTTCGCCCTACAAACAGCAGGCAGATGACCCATACAG
AGCAAGCCGAGAATATCCATCCTACCTCGTTATCCCTCGACAAATCTGGGAACGGCCCGCC
CTTTCAAGATCTTTAGACACACCACACTGCGACGGAAAGAGGTACACCCAGACCAAAAGAG
GCTGCTGGACAGCGCCCTGACACTCAGACAGACACTCCGCGCTGACGACAGGAT
CGACCTGCTGCTGAGGCGCACAAGAAAGGCCCGCCGCGCCAGAAAAGGCGCGGC
CAGGCAAAAAAGAAAAG

P2A-EGFP

FIG. 50C
aaagcaatagcatcatacttacatcatttttcatcctctctagtggaggctgtgcttcagcatgttgattgtcgctcaaacactcataaacatatcatttagcagttgggtactcactccatcactccatttcatccttgtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
FIG. 50F
FIG. 50G
FIG. 50H
Validation of Cas9 nuclease activity by Surveyor

FIG. 53
<table>
<thead>
<tr>
<th>pCAG-loxP(pA)loxP-NLS-hSpCas9-NLS-2A-GFP</th>
<th>Average</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 1</td>
<td>32.1</td>
<td>7.1</td>
</tr>
<tr>
<td>Clone 2</td>
<td>27.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Clone 3</td>
<td>35.9</td>
<td>1.4</td>
</tr>
<tr>
<td>Clone 4</td>
<td>39.0</td>
<td>4.7</td>
</tr>
</tbody>
</table>

| pCAG-NLS-hSpCas9-NLS-2A-GFP            |         |       |
| Clone 1                                | 26.9    | 1.3   |
| Clone 2                                | 33.1    | 2.7   |

FIG. 54
left guide RNA target (L)

5' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
3' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

spacer (0 - n BP)

right guide RNA target (R)

5' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
3' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

processing by nickase
(to improve NHEJ efficiency, can also co-express with TREX1)

5' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
3' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

FIG. 58
Day 0: *in silico* design

Day 2.5: Reagent construction

Day 5-8: Functional validation

Day 9-28: Clonal expansion

Diagram:

- `sgRNA` Cas9
- PX330
- `pCRISPR` (sgRNA in PX330)
- Repair template (optional)
- Transfect
- Genotype
- Isolate clonal lines
- Expand

FIG. 60
Fig. 61A-C
FIG. 62A-D
FIG. 63A-C
U6 Promoter

U6 forward primer 5' GAGGGCCCTATTTCCATGATTCC-3'
5' - GAGGGCCCTATTTCCATGATTCC -3'

- guide sequence - direct repeat -
5' - GAAACACCTTTCTGCTTTTTG -3'

5' - AAAAAAAGCACCGACTCGGCTGCCCACTTTTTCAAACTGTTGATACCGACTAATTGCATATTTATTTCTGCTTTTTG -3'

tracrRNA

U6 Promoter

5' - GAGGGCCCTATTTCCATGATTCC -3'

- guide sequence - direct repeat -
5' - GAAACACCTTTCTGCTTTTTG -3'

5' - AAAAAAAGCACCGACTCGGCTGCCCACTTTTTCAAACTGTTGATACCGACTAATTGCATATTTATTTCTGCTTTTTG -3'

tracrRNA

CBh
NLS SpCas9 bGH pA

U6

chimeric RNA

PCR or gel purification

FIG. 64
<table>
<thead>
<tr>
<th>Gene</th>
<th>gRNA1</th>
<th>gRNA2</th>
<th>gRNA3</th>
<th>None</th>
<th>Avg % Cutting</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCN2A</td>
<td></td>
<td></td>
<td></td>
<td>4.6</td>
<td>ND 2.2</td>
</tr>
<tr>
<td>KATNAL2</td>
<td>6.0</td>
<td>2.3</td>
<td>5.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCTD13</td>
<td></td>
<td></td>
<td>3.8</td>
<td>10.4</td>
<td>3.8</td>
</tr>
<tr>
<td>CHD8</td>
<td></td>
<td></td>
<td></td>
<td>NO</td>
<td>2.0 8.1*</td>
</tr>
<tr>
<td>CACNA1C</td>
<td></td>
<td></td>
<td>9.4</td>
<td>3.9</td>
<td>6.3</td>
</tr>
<tr>
<td>CACNB2</td>
<td></td>
<td></td>
<td>5.1</td>
<td>1.2</td>
<td>3.0</td>
</tr>
<tr>
<td>GIT1</td>
<td></td>
<td></td>
<td>6.1</td>
<td>2.8</td>
<td>6.3</td>
</tr>
<tr>
<td>MECP2</td>
<td>2.2</td>
<td>2.7</td>
<td>6.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHANK3</td>
<td>20.2</td>
<td>9.8</td>
<td>7.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- For injection, unique target
- * Suboptimal gRNA, 1 off-target prediction
- ** Suboptimal gRNA, 7 off-target predictions
- ND Not Determined

FIG. 65
gRNA sequences for Chd8 targeting:

Chd8.1 - agctgttttactgtccgct
Chd8.2 - aatggata cacctgtc gaa
Chd8.3 - caat ggata cacctgtc gaa
FIG. NLS1

FIG. 67
Cas9-NLS Test (EMX1 protospacer 1)

Percentage of Endogenous Genome Cleavage

Construct NLS positions

FIG. 68
NLS architecture optimization for SpCas9

FIG. 70
CRISPR-CAS NICKASE SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION IN EUKARYOTES

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE


The foregoing applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, and all documents cited or referenced herein ("herein cited documents"), and all documents cited or referenced in herein cited documents, together with any manufacturer’s instructions, descriptions, product specifications, and product sheets for any products mentioned herein or in any document incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. More specifically, all referenced documents are incorporated by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

This invention was made with government support under the NIH Pioneer Award DP1MH100706, awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE INVENTION

The present invention generally relates to systems, methods and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that may use vector systems related to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Aug. 28, 2014, is named 44790.06.2003_SL.txt and is 321,104 bytes in size.

BACKGROUND OF THE INVENTION

Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome engineering technologies that are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome.

SUMMARY OF THE INVENTION

There exists a pressing need for alternative and robust systems and techniques for sequence targeting with a wide array of applications. This invention addresses this need and provides related advantages. The CRISPR/Cas or the CRISPR-Cas system (both terms are used interchangeably throughout this application) does not require the generation of customized proteins to target specific sequences but rather a single Cas enzyme can be programmed by a short RNA molecule to recognize a specific DNA target, in other words the Cas enzyme can be recruited to a specific DNA target using said short RNA molecule. Adding the CRISPR-Cas system to the repertoire of genome sequencing techniques and analysis methods may significantly simplify the methodology and accelerate the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. To utilize the CRISPR-Cas system effectively for genome editing without deleterious effects, it is critical to understand aspects of engineering and optimization of these genome engineering tools, which are aspects of the claimed invention.

In one aspect, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell; wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; and (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence; wherein components (a) and (b) are located on the same or different vectors of the system. In some embodiments, component (a) further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct specific binding of a CRISPR complex to a different target sequence in a
eukaryotic cell. In some embodiments, the system comprises the tracr sequence under the control of a third regulatory element, such as a polymerase III promoter. In some embodiments, the tracr sequence exhibits at least 50%, 60%, 70%, 80%, 90%, 95%, or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. Determining optimal alignment is within the purview of one of skill in the art. For example, there are publically and commercially available alignment algorithms and programs such as, but not limited to, ClustalW, Smith-Waterman in matlabb, Bowtie, Geneious, BioPython and SeqMan. In some embodiments, the CRISPR complex comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR complex in a detectable amount in the nucleus of an eukaryotic cell. With this, it is well understood that a nuclear localization sequence is not necessary for CRISPR complex activity in eukaryotes, but that including such sequences enhances stability of the system, especially as to targeting nucleic acid molecules in the nucleus. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is S. pneumoniae, S. pyogenes, or S. thermophilus Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in an eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length. In general, and throughout this specification, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other variations of polynucleotides known in the art. One type of vector is a “plasmid,” which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA segments are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as “expression vectors.” Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, “operably linked” is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term “regulatory element” is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g. transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMEOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g. liver, pancreas), or particular cell types (e.g. lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoters (e.g. 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g. 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g. 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol I promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV)LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al., Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. Also encompassed by the term “regulatory element” are enhancer elements, such as WPRE; CMV enhancers; the R-U5 segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.).

Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

In one aspect, the invention provides a vector comprising a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising one or
more nuclear localization sequences. In some embodiments, said regulatory element drives transcription of the CRISPR enzyme in a eukaryotic cell such that said CRISPR enzyme accumulates in a detectable amount in the nucleus of the eukaryotic cell. In some embodiments, the regulatory element is a polymerase II promoter. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is S. pneumoniae, S. pyogenes or S. thermophilus Cas9, and may include mutated Cas9 derived from these organisms. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity.

In one aspect, the invention provides a CRISPR enzyme comprising one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is S. pneumoniae, S. pyogenes or S. thermophilus Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length. In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments may be an animal, for example a mammal. Also, the organism may be an insect. The organism also may be a plant. Further, the organism may be a fungus.

In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence. In some embodiments, the host cell comprises components (a) and (b). In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein each of the two or more guide sequences directs sequence-specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the eukaryotic host cell further comprises a third regulatory element, such as a polymerase III promoter, operably linked to said tracr sequence. In some embodiments, the tracr sequence exhibits at least 50%, 60%, 70%, 80%, 90%, 95%, or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. In some embodiments, the CRISPR enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is S. pneumoniae, S. pyogenes or S. thermophilus Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a
eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase II promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length.

In one aspect, the invention provides a method of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene.

In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence. In some embodiments, said vectors are delivered to the eukaryotic cell in a subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein said CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence.

In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises: introducing one or more vectors into the eukaryotic cell (s), wherein the one or more vectors drive expression of one or more of: a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, and a tracr sequence; and (b) allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said disease gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized to the tracr sequence, thereby generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

In one aspect, the invention provides a method for developing a biologically active agent that modulates a cell signaling event associated with a disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises: contacting a test compound with a model cell of any one of the described embodiments; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with said mutation in said disease gene, thereby developing said biologically active agent that modulates said cell signaling event associated with said disease gene.

In one aspect, the invention provides a recombinant polynucleotide comprising a guide sequence upstream of a tracr mate sequence, wherein the guide sequence when expressed directs sequence-specific binding of a CRISPR complex to a corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

In one aspect the invention provides for a method of selecting one or more prokaryotic cell(s) by introducing one or more mutations in a gene in the one or more prokaryotic cell(s), the method comprising: introducing one or more vectors into the prokaryotic cell(s), wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, a guide sequence linked to a tracr mate sequence, a tracr sequence, and a editing template; wherein the editing template comprises the one or more mutations that abolish CRISPR enzyme cleavage; allowing homologous recombination of the editing template with the target polynucleotide in the cell(s) to be selected; allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized to the tracr sequence, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more prokaryotic cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment, the CRISPR enzyme is Cas9. In another aspect of the invention the cell to be selected may be a eukaryotic cell. Aspects of the invention
allow for selection of specific cells without requiring a selection marker or a two-step process that may include a counterselection system.

Accordingly, it is an object of the invention not to encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or method of making the product, or method of using the product.

It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention. These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

FIG. 1 shows a schematic model of the CRISPR system. The Cas9 nuclease from Streptococcus pyogenes (yellow) is targeted to genomic DNA by a synthetic guide RNA (sgRNA) consisting of a 20 nt guide sequence (blue) and a scaffold (red). The guide sequence base-pairs with the DNA target (blue), directly upstream of a requisite 5′-NGG protospacer adjacent motif (PAM; magenta), and Cas9 mediates a double-stranded break (DSB) ~3 bp upstream of the PAM (red triangle).

FIG. 2 shows an exemplary CRISPR system, a possible mechanism of action, an example adaptation for expression in eukaryotic cells, and results of tests assessing nuclear localization and CRISPR activity. FIG. 2C discloses SEQ ID NOS 281-282, respectively, in order of appearance. FIG. 2D discloses SEQ ID NOS 283-285, respectively, in order of appearance. FIG. 2E discloses SEQ ID NOS 286-290, respectively, in order of appearance.

FIG. 3 shows an exemplary expression cassette for expression of CRISPR system elements in eukaryotic cells, predicted structures of example guide sequences, and CRISPR system activity as measured in eukaryotic and prokaryotic cells (SEQ ID NOS 291-300, respectively, in order of appearance).

FIG. 4A-D shows results of an evaluation of SpCas9 specificity for an example target. FIG. 4A discloses SEQ ID NOS 301, 284 and 302-312, respectively, in order of appearance. FIG. 4C discloses SEQ ID NO: 301.

FIG. 5 shows an exemplary vector system and results for its use in directing homologous recombination in eukaryotic cells. FIG. 5E discloses SEQ ID NO: 313. FIG. 5F discloses SEQ ID NOS 314-315, respectively, in order of appearance. FIG. 5G discloses SEQ ID NOS 316-320, respectively, in order of appearance.

FIG. 6 provides a table of protospacer sequences (SEQ ID NOS 33, 32, 31, 324-329, 35, 34 and 332-336, respectively, in order of appearance) and summarizes modification efficiency results for protospacer targets designed based on exemplary S. pyogenes and S. thermophilus CRISPR systems with corresponding PAMs against loci in human and mouse genomes. Cells were transfected with Cas9 and either pre-crRNA/tracrRNA or chimeric RNA, and analyzed 72 hours after transfection. Percent indels are calculated based on Surveyor assay results from indicated cell lines (N=3 for all protospacer targets, errors are S.E.M., N.D. indicates not detectable using the Surveyor assay, and N.T. indicates not tested in this study).

FIG. 7A-C shows a comparison of different tracrRNA transcripts for Cas9-mediated gene targeting. FIG. 7A discloses SEQ ID NOS 337-338, respectively, in order of appearance. FIG. 8 shows a schematic of a surveyor nuclease assay for detection of double strand break-induced micro-insertions and -deletions.

FIG. 9 shows exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells. FIG. 9A discloses SEQ ID NOS 339-341, respectively, in order of appearance. FIG. 9B discloses SEQ ID NOS 342-344, respectively, in order of appearance.

FIGS. 10A-10D show (A) SEQ ID NOS 343-345, respectively, in order of appearance, (B) a bacterial plasmid transformation interference assay, (C) expression cassettes and plasmids used therein, and (D) transformation efficiencies of cells used therein.

FIG. 11 shows histograms of distances between adjacent S. pyogenes SF370 locus 1 PAM (NGG) (FIG. 10A) and S. thermophilus LM90 locus 2 PAM (NNAGA), (FIG. 10B) in the human genome; and distances for each PAM by chromosome (Chr) (FIG. 10C).

FIG. 12A-C shows an exemplary CRISPR system, an example adaptation for expression in eukaryotic cells, and results of tests assessing CRISPR activity. FIG. 12B discloses SEQ ID NOS 348-349, respectively, in order of appearance. FIG. 12C discloses SEQ ID NO: 350.

FIG. 13 shows exemplary manipulations of a CRISPR system for targeting of genomic loci in mammalian cells. FIG. 13A discloses SEQ ID NOS 343-354, respectively, in order of appearance. FIG. 14A-B shows the results of a Northern blot analysis of crRNA processing in mammalian cells. FIG. 14A discloses SEQ ID NO: 355.

FIG. 15 shows an exemplary selection of protospacers in the human PVALB and muss Th loci. FIG. 15A discloses SEQ ID NO: 356. FIG. 15B discloses SEQ ID NO: 357.

FIG. 16 shows example protospacer and corresponding PAM sequence targets of the S. thermophilus CRISPR system in the human EMX1 locus (SEQ ID NO: 350).

FIG. 17 provides a table of sequences for primers and probes (SEQ ID NOS 36-39 and 358-365, respectively, in order of appearance) used for Surveyor, RLFP, genomic sequencing, and Northern blot assays.

FIG. 18A-C shows exemplary manipulation of a CRISPR system with chimeric RNAs and results of SURVEYOR assays for system activity in eukaryotic cells. FIG. 18A discloses SEQ ID NO: 366, respectively, in order of appearance.
FIG. 19 shows a graphical representation of the results of SURVEYOR assays for CRISPR system activity in eukaryotic cells.

FIG. 20 shows an exemplary visualization of some S. pyogenes Cas9 target sites in the human genome using the UCSC genome browser (SEQ ID NO. 367-445, respectively, in order of appearance).

FIG. 21 shows predicted secondary structures for exemplary chimeric RNAs comprising a guide sequence, tracr mate sequence, and tracr sequence (SEQ ID NO. 446-465, respectively, in order of appearance).

FIG. 22 shows exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells (SEQ ID NO. 466 and 343-344, respectively, in order of appearance).

FIG. 23 shows that Cas9 nuclease activity against endogenous targets may be exploited for genome editing. (a) Concept of genome editing using the CRISPR system. The CRISPR targeting construct directed cleavage of a chromosomal locus and was co-transformed with an editing template that recombined with the target to prevent cleavage. Kanamycin-resistant transformants that survived CRISPR attack contained modifications introduced by the editing template. (b) Transformation of crR6M DNA in R6232.5 cells with no editing template, the R6 wild-type stR vector or the R6370.1 editing templates. Recombination of either R6 stR or R6370.1 prevented cleavage by Cas9. Transformation efficiency was calculated as colony forming units (cfu) per μg of crR6M DNA; the mean values with standard deviations from at least three independent experiments are shown. PCR analysis was performed on 8 clones in each transformation. "Un." indicates the unedited stR locus of strain R6232.5; "Ed." shows the editing template. R6232.5 and R6370.1 targets are distinguished by restriction with Eae.

FIG. 24 shows analysis of PAM and seed sequences that eliminate Cas9 cleavage. (a) PCR products with randomized PAM sequences or random seed sequences were transformed in R6 cells (SEQ ID NO. 467-471, respectively, in order of appearance). These cells expressed Cas9 loaded with a crRNA that targeted a chromosomal region of R6232.5 cells (highlighted in pink) that is absent from the R6 genome. More than 2×105 chloramphenicol-resistant transformants, carrying inactive PAM or seed sequences, were combined for amplification and deep sequencing of the target region. (b) Relative proportion of number of reads after transformation of the random PAM constructs in crR6 cells (compared to number of reads in R6 transformants). The relative abundance for each 3-nucleotide PAM sequence is shown. Severely underrepresented sequences (NGG) are shown in red; partially underrepresented one in orange (NAG). (c) Relative proportion of number of reads after transformation of the random seed sequence constructs in crR6 cells (compared to number of reads in R6 transformants). The relative abundance of each nucleotide for each position of the first 20 nucleotides of the protospacer sequence is shown (SEQ ID NO. 472). High abundance indicates lack of cleavage by Cas9, i.e. a CRISPR inactivating mutation. The grey line shows the level of the WT sequence. The dotted line represents the level above which a mutation significantly disrupts cleavage (see section "Analysis of deep sequencing data" in Example 5).

FIG. 25 shows introduction of single and multiple mutations using the CRISPR system in S. pneumoniae. (a) Nucleotide and amino acid sequences of the wild-type and edited (green nucleotides; underlined amino acid residues) bgaA. The protospacer, PAM and restriction sites are shown (SEQ ID NO. 473-477 and 474, respectively, in order of appearance). (b) Transformation efficiency of cells transformed with targeting constructs in the presence of an editing template or control. (c) PCR analysis for 8 transformants of each editing experiment followed by digestion with NsilI (R→G) and Tsp50I (NE→AA). Deletion of bgaA was revealed as a smaller PCR product. (d) Miller assay to measure the β-galactosidase activity of WT and edited strains. (e) For a single-step, double deletion the targeting construct contained two spacers (in this case matching stR and bgaA) and was co-transformed with two different editing templates (f) PCR analysis for 8 transformants to detect deletions in stR and bgaA loci. 6/8 transformants contained deletions of both genes.

FIG. 26 provides mechanisms underlying editing using the CRISPR system. (a) A stop codon was introduced in the erythromycin resistance gene ermAM to generate strain JEN53. The wild-type sequence can be restored by targeting the stop codon with the CRISPR:ermAM(stop) construct, and using the ermAM wild-type sequence as an editing template. (b) Mutant and wild-type ermAM sequences (SEQ ID NO. 478-481, respectively, in order of appearance). (c) Fraction of erythromycin-resistant (ermAM) cfu calculated from total or kanamycin-resistant (kanR) cfu. (d) Fraction of total cells that acquire both the CRISPR construct and the editing template. Co- transformation of the CRISPR targeting construct produced more transformants (t-test, p=0.011). In all cases the values show the mean±sd. for three independent experiments.

FIG. 27 illustrates genome editing with the CRISPR system in E. coli. (a) A kanamycin-resistant plasmid carrying the CRISPR array (pCRISPR) targeting the gene to edit may be transformed in the EHEC-3 recombineering strain containing a chloramphenicol-resistant plasmid harboring cas9 and tracr (Cas9), together with an oligonucleotide specifying the mutation. (b) A K42F mutation conferring streptomycin resistance was introduced in the rpsL gene (SEQ ID NO. 482-485, respectively, in order of appearance). (c) Fraction of streptomycin-resistant (stepR) cfu calculated from total or kanamycin-resistant (kanR) cfu. (d) Fraction of total cells that acquire both the pCRISPR plasmid and the editing oligonucleotide. Co-transformation of the pCRISPR targeting plasmid produced more transformants (t-test, p=0.004). In all cases the values showed the mean±sd. for three independent experiments.

FIG. 28 illustrates the transformation of crR6 genomic DNA leads to editing of the targeted locus (a) The IS1167 element of S. pneumoniae R6 was replaced by the CRISPR01 locus of S. pyogenes SF370 to generate crR6 strain. This locus encodes for the Cas9 nuclease, a CRISPR array with six spacers, the tracrRNA that is required for crRNA biogenesis and Cas1, Cas2 and Cas3, proteins not necessary for targeting. Strain crR6M contains a minimal functional CRISPR system without cas1, cas2 and cas3. The aphaA-1 gene encodes kanamycin resistance. Protospacers from the streptococcal bacteriophages f8232.5 and f370.1 were fused to a chloramphenicol resistance gene (cat) and integrated in the stR gene of strain R6 to generate strains R6232.5 and R6370.1. (b) Left panel: Transformation of crR6 and crR6M genomic DNA in R6232.5 and R6370.1. As a control of cell competence a streptomycin resistant gene was also transformed. Right panel: PCR analysis of 8 R6232.5 transformants with crR6 genomic DNA. Primers that amplify the stR locus were used for PCR. 7/8 genotyped colonies replaced the R68232.5 stR locus by the WT locus from the crR6 genomic DNA.

FIG. 29 provides chromatograms of DNA sequences of edited cells obtained in this study. In all cases the wild-type and mutant protospacer and PAM sequences (or their reverse complement) are indicated. When relevant, the amino acid
sequence encoded by the protospacer is provided. For each editing experiment, all strains for which PCR and restriction analysis corroborated the introduction of the desired modification were sequenced. A representative chromatogram is shown. (a) Chromatograms for the introduction of a PAM mutation into the R6 target (FIG. 23b) (SEQ ID NOS 486-487, respectively, in order of appearance). (b) Chromatograms for the introduction of the R>A and N/E>A A mutations into β-galactosidase (β-galA) (FIG. 23c) (SEQ ID NOS 473-477 and 474, respectively, in order of appearance). (c) Chromatograms for the introduction of a 6664 bp deletion within β-galA ORF (FIGS. 25c and 25f). The dotted line indicates the limits of the deletion (SEQ ID NOS 488-490, respectively, in order of appearance). (d) Chromatograms for the introduction of a 729 bp deletion within srA ORF (FIG. 25f). The dotted line indicates the limits of the deletion (SEQ ID NOS 491-493, respectively, in order of appearance). (e) Chromatograms for the generation of a premature stop codon within ermAM (FIG. 33) (SEQ ID NOS 494-497, respectively, in order of appearance). (f) rpsL editing in E. coli (FIG. 27) (SEQ ID NOS 492-485, respectively, in order of appearance).

FIG. 30 illustrates CRISPR immunity against random S. pneumoniae targets containing different PAMs. (a) Position of the 10 random targets on the S. pneumoniae R6 genome. The chosen targets have different PAMs and are on both strands. (b) Spacers corresponding to the targets were cloned in a minimal CRISPR array on plasmid pLZ12 and transformed into strain crR6Rc, which supplies the processing and targeting machinery in trans. (c) Transformation efficiency of the different plasmids in strain R6 and crR6Rc. No colonies were recovered for the transformation of pDB99-108 (T1-T10) in crR6Rc. The dashed line represents limit of detection of the assay.

FIG. 31 provides a general scheme for targeted genome editing. To facilitate targeted genome editing, crR6M was further engineered to contain tracrRNA, Cas9 and only one repeat of the CRISPR array followed by kanamycin resistance marker (aphA-3), generating strain crR6Rk. DNA from this strain is used as a template for PCR with primers designed to introduce a new spacer (green box designated with N). The left and right PCRAs are assembled using the Gibson method to create the targeting construct. Both the targeting and editing constructs are then transformed into strain crR6Rc, which is a strain equivalent to crR6Rk but has the kanamycin resistance marker replaced by a chloramphenicol resistance marker (cat). About 90% of the kanamycin-resistant transformants contain the desired mutation.

FIG. 32 illustrates the distribution of distances between PAMs. NGG and CCN that are considered to be valid PAMs. Data is shown for the S. pneumoniae R6 genome as well as for a random sequence of the same length and with the same GC-content (39.7%). The dotted line represents the average distance (12) between PAMs in the R6 genome.

FIG. 33 illustrates CRISPR-mediated editing of the ermAM locus using genomic DNA as targeting construct. To use genomic DNA as targeting construct it is necessary to avoid CRISPR autoimmunity, and therefore a spacer against a sequence not present in the chromosome must be used (in this case the ermAM erythromycin resistance gene). (a) Nucleotide and amino acid sequences of the wild-type and mutated (red letters) ermAM gene. The protospacer and PAM sequences are shown (SEQ ID NOS 494-497, respectively, in order of appearance). (b) A schematic for CRISPR-mediated editing of the ermAM locus using genomic DNA. A construct carrying an ermAM-targeting spacer (blue box) is made by PCR and Gibson assembly, and transformed into strain crR6Rc, generating strain JEN37. The genomic DNA of JEN37 was then used as a targeting construct, and was co-transformed with the editing template into JEN38, a strain in which the srA gene was replaced by a wild-type copy of ermAM. Kanamycin-resistant transformants contain the edited genotype (JEN43). (c) Number of kanamycin-resistant cells obtained after co-transformation of targeting and editing control templates. In the presence of the control template 5.4x10^4 cfu/ml were obtained, and 4.3x10^3 cfu/ml when the editing template was used. This difference indicates an editing efficiency of about 99% ([4.3x10^5]/[5.4x10^4]/4.3x10^3]. (d) To check for the presence of edited cells seven kanamycin-resistant clones and JEN38 were streaked on agar plates with (erm+) or without (erm-) erythromycin. Only the positive control displayed resistance to erythromycin. The ermAM mut genotype of one of these transformants was also verified by DNA sequencing (FIG. 29e). FIG. 34 illustrates sequential introduction of mutations by CRISPR-mediated genome editing. (a) A schematic for sequential introduction of mutations by CRISPR-mediated genome editing. First, R6 is engineered to generate crR6Rk. crR6Rk is co-transformed with a srA-targeting construct fused to cat for chloramphenicol selection of edited cells, along with an editing construct for a ΔsrA in-frame deletion. Strain crR6 ΔsrA is generated by selection on chloramphenicol. Subsequently, the ΔsrA strain is co-transformed with a β-galA-targeting construct fused to aphA-3 for kanamycin selection of edited cells, and an editing construct containing a Δβ-galA in-frame deletion. Finally, the engineered CRISPR locus can be erased from the chromosome by first co-transforming R6 DNA containing the wild-type IS1167 locus and a plasmid carrying a β-galA protosporer (pDB97), and selection on spectinomycin. (b) PCR analysis for 8 chloramphenicol (CmR)-resistant transformants to detect the deletion in the srA locus. (c) β-galactosidase activity as measured by Miller assay. In S. pneumoniae, this enzyme is anchored to the cell wall by sortase A. Deletion of the srA gene results in the release of β-galactosidase into the supernatant. Δβ-galA mutants show no activity. (d) PCR analysis for 8 spectinomycin (Spec)-resistant transformants to detect the replacement of the CRISPR locus by wild-type IS1167.

FIG. 35 illustrates the background mutation frequency of CRISPR in S. pneumoniae. (a) Transformation of the CRISPR::O or CRISPR::erm(stop) targeting constructs in JEN53, with or without the ermAM editing template. The difference in kan^2 CFU between CRISPR::O and CRISPR::erm(stop) indicates that Cas9 cleavage kills non-edited cells. Mutants that escape CRISPR interference in the absence of editing template are observed at a frequency of 3x10^{-5}. (b) PCR analysis of the CRISPR locus of escapers shows that 7/8 have a spacer deletion. (c) Escaper #2 carries a point mutation in cas9 (SEQ ID NOS 498-501, respectively, in order of appearance).

FIG. 36 illustrates that the essential elements of the S. pyogenes CRISPR locus 1 are reconstituted in E. coli using pCas9. The plasmid contained tracrRNA, Cas9, as well as a leader sequence driving the crRNA array. The pCRISPR plasmids contained the leader and the array only. Spacers may be inserted into the crRNA array between Bsal sites using annealed oligonucleotides (SEQ ID NOS 345, 502 and 127, respectively, in order of appearance). Oligonucleotide design is shown at bottom. pCas9 carried chloramphenicol resistance (CmR) and is based on the low-copy pACYC184 plasmid backbone. pCRISPR is based on the high-copy number pZE21 plasmid. Two plasmids were required because a pCRISPR plasmid containing a spacer targeting the E. coli chromosome may not be constructed using this organism as a cloning host if Cas9 is also present (it will kill the host).
FIG. 37 illustrates CRISPR-directed editing in E. coli MG1655. An oligonucleotide (W542) carrying a point mutation that both confers streptomycin resistance and abolishes CRISPR immunity, together with a plasmid targetting rpsL (pCRISPR-rpsL) or a control plasmid (pCRISPR::O) were co-transformed into wild-type E. coli strain MG1655 containing pCas9. Transformants were selected on media containing either streptomycin or kanamycin. Dashed line indicates limit of detection of the transformation assay.

FIG. 38 illustrates the background mutation frequency of CRISPR in E. coli HME63. (a) Transformation of the pCRISPR::O or pCRISPR::rpsL plasmids into HME63 competent cells. Mutants that escape CRISPR interference were observed at a frequency of 2.6x10^-4. (b) Amplification of the CRISPR array of escapees showed that 8/8 have deleted the spacer.

FIG. 39 shows a circular depiction of the phylogenetic analysis revealing five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

FIG. 40 shows the linear depiction of the phylogenetic analysis revealing five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

FIG. 41A-M shows sequences where the mutation points are located within the SpCas9 gene (SEQ ID Nos 503-504, respectively, in order of appearance).

FIG. 42 shows a schematic construct in which the transcriptional activation domain (VP64) is fused to Cas9 with two mutations in the catalytic domains (D10 and H840).

FIG. 43-A-D shows genome editing via homologous recombination. (a) Schematic of SpCas9 nickase, with D10A mutation in the RuvC I catalytic domain. (b) Schematic representing homologous recombination (HR) at the human EMX1 locus using either sense or antisense single stranded oligonucleotides as repair templates. Red arrow above indicates sgRNA cleavage site; PCR primers for genotyping (Tables 1 and K) are indicated as arrows in right panel. (c) Sequence of region modified by HR. d. SURVEYOR assay for wildtype (wt) and nickase (D10A) SpCas9-mediated indels at the EMX1 target 1 locus (n=3) (SEQ ID Nos 505-507, 505, 508 and 507, respectively, in order of appearance). Arrows indicate positions of expected fragment sizes.

FIG. 44 shows single vector designs for SpCas9. FIG. 44A discloses SEQ ID Nos 322-323 and 330, respectively, in order of appearance. FIG. 44B discloses SEQ ID NO: 331.

FIG. 45 shows quantification of cleavage of NLS-Csm1 constructs NLS-Csn1, Csn1, Csn1-NLS, NLS-Csn1-NLS, NLS-Csn1-GFP-NLS and NLS-Cas9-NLS.

FIG. 46 shows index frequency of NLS-Cas9, Cas9, Cas9-NG and NLS-Cas9-NLS.

FIG. 47 shows a gel demonstrating that SpCas9 with nickase mutations (individually) do not induce double strand breaks.

FIGS. 48A-48B show (A) a design of the oligo DNA used as a Homologous Recombination (HR) template in this experiment and (B) a comparison of HR efficiency induced by different combinations of Cas9 protein and HR template.

FIG. 49A shows the Conditional Cas9, Rosas26 targeting vector map. FIG. 49B shows the Constitutive Cas9, Rosas26 targeting vector map.

FIG. 50A-H show the sequences of each element present in the vector maps of FIGS. 49A-B (SEQ ID Nos 509-518, respectively, in order of appearance).

FIG. 51 shows a schematic of the important elements in the Constitutive and Conditional Cas9 constructs.
of the Type IIIs restriction sites (blue outline) with direct insertion of annealed oligos. It is worth noting that an extra G was placed before the first base of the guide sequence. Applicants have found that an extra G in front of the guide sequence does not adversely affect targeting efficiency. In cases when the 20-nt guide sequence of choice does not begin with guanine, the extra guanine will ensure the sgRNA is efficiently transcribed by the U6 promoter, which prefers a guanine in the first base of the transcript (SEQ ID NOS 322-323 and 330, respectively, in order of appearance).

FIG. 62A-D shows the anticipated results for multiplex NHEJ. (a) Schematic of the SURVEYOR assay used to determine indel percentage. First, genomic DNA from the heterogeneous population of Cas9-targeted cells is amplified by PCR. Amplicons are then annealed slowly to generate heteroduplexes. The renaturated heteroduplexes are cleaved by SURVEYOR nuclease, whereas homoduplexes are left intact. Cas9-mediated cleavage efficiency (% indel) is calculated based on the fraction of cleaved DNA, as determined by integrated intensity of gel bands. (b) Two sgRNAs (orange and blue bars) are designed to target the human GRIN2B and DYNK1A loci. SURVEYOR gel shows modification at both loci in transfected cells. Colored arrows indicated expected fragment sizes for each locus. (c) A pair of sgRNAs (light blue and green bars) are designed to excise an exon (dark blue) in the human EMX1 locus. Target sequences and PAMs (red) are shown in respective colors, and sites of cleavage indicated by red triangle. Predicted junction is shown below. Individual clones isolated from cell populations transfected with sgRNA 3, 4, or both are assayed by PCR (OUT Twd, OUT Rev), reflecting a deletion of ~270-bp. Representative clones with no modification (12/23), mono-allelic (10/23), and bi-allelic (1/23) modifications are shown. IN Fwd and IN Rev primers are used to screen for inversion events (FIG. 6d) (SEQ ID NOS 522-524, respectively, in order of appearance). (d) Quantification of clonal lines with EMX1 exon deletions. Two pairs of sgRNAs (3.1, 3.2 left-flanking sgRNAs; 4.1, 4.2, right flanking sgRNAs) are used to mediate deletions of variable sizes around one EMX1 exon. Transfected cells are clonally isolated and expanded for genotyping analysis for deletions and inversion events. Of the 105 clones are screened, 51 (49%) and 11 (10%) carrying heterozygous and homozygous deletions, respectively. Approximate deletion sizes are given since junctions may be variable.

FIG. 63A-C shows the application of sSODNs and targeting vector to mediate HR with both wildtype and nickase mutant of Cas9 in HEK293FT and HUES9 cells with efficiencies ranging from 1.0-27%. FIG. 63B discloses SEQ ID NOS 505-507, 505, 508 and 507, respectively, in order of appearance.

FIG. 64 shows a schematic of a PCR-based method for rapid and efficient CRISPR targeting in mammalian cells. A plasmid containing the human RNA polymerase III promoter U6 is PCR-amplified using a U6-specific forward primer and a reverse primer, to yield the complementary part of the U6 promoter, the sgRNA (+85) scaffold with guide sequence, and 7 T nucleotides for transcriptional termination. The resulting PCR product is purified and co-delivered with a plasmid carrying Cas9 driven by the CBh promoter (SEQ ID NOS 519, 525, 520 and 526-527, respectively, in order of appearance).

FIG. 65 shows SURVEYOR Mutation Detection Kit from Transgenomics results for each gRNA and respective controls. A positive SURVEYOR result is one large band corresponding to the genomic PCR and two smaller bands that are the product of the SURVEYOR nuclease making a double-strand break at the site of a mutation. Each gRNA was validated in the mouse cell line, Neuro-N2a, by liposomal transient co-transfection with hSpCas9. 72 hours post-transfection genomic DNA was purified using QuickExtract DNA from Epicentre. PCR was performed to amplify the locus of interest.

FIG. 66 shows Surveyor results for 38 live pups (lanes 1-38) 1 dead pup (lane 39) and 1 wild-type pup for comparison (lane 40). Pups 1-19 were injected with gRNA Chdh-2 and pups 20-38 were injected with gRNA Chdh-3. Of the 38 live pups, 13 were positive for a mutation. The one dead pup also had a mutation. There was no mutation detected in the wild-type sample. Genomic PCR sequencing was consistent with the SURVEYOR assay findings (SEQ ID NOS 528-530, respectively, in order of appearance).

FIG. 67 shows a design of different Cas9 NLS constructs. All Cas9s were the human-codon-optimized version of the Sp Cas9. NLS sequences are shown for the human cas9 gene, either N-terminals or C-terminals. All Cas9 variants with different NLS designs were cloned into a backbone vector containing so it is driven by EF1a promoter. On the same vector there is a chimeric RNA targeting human EMX1 locus driven by U6 promoter, together forming a two-component system.

FIG. 68 shows the efficiency of genomic cleavage induced by Cas9 variants bearing different NLS designs. The percentage indicates the portion of human EMX1 genomic DNA that were cleaved by each construct. All experiments are from 3 biological replicates. n=3, error indicates S.E.M.

FIG. 69A shows a design of the CRISPR-TF (Transcription Factor) with transcriptional activation activity. The chimeric RNA is expressed by U6 promoter, while a human-codon-optimized, double-mutant version of the Cas9 protein (hSpCas9m), is driven by triple NLS and V646 functional domain is expressed by a EF1a promoter. The double mutations, D10A and H840A, renders the cas9 protein unable to introduce any cleavage but maintained its capacity to bind to target DNA when guided by the chimeric RNA.

FIG. 69B shows transcriptional activation of the human SOX2 gene with CRISPR-TF system (Chimeric RNA and the Cas9-NLS-V646 fusion protein). 293FT cells were transfected with plasmids bearing two components: (1) U6-driven different chimeric RNAs targeting 20-bp sequences within or around the human SOX2 genomic locus, and (2) EF1a-driven hSpCas9m (double mutant)-NLS-V646 fusion protein. 96 hours post transfection, 293FT cells were harvested and the level of activation is measured by the induction of mRNA expression using a qRT-PCR assay. All expression levels are normalized against the control group (grey bar), which represents results from cells transfected with the CRISPR-TF backbone plasmid without chimeric RNA. The qRT-PCR probes used for detecting the SOX2 mRNA is Taqman Human Gene Expression Assay (Life Technologies). All experiments represents data from 3 biological replicates, n=3, error bars show s.e.m.

FIG. 70 depicts NLS architecture optimization for SpCas9.

FIG. 71 shows a QQ plot for NGGNN sequences.

FIG. 72 shows a histogram of the data density with fitted normal distribution (black line) and 0.99 quantile (dotted line).

FIG. 73A-C shows RNA-guided repression of bga expression by dgrRNA::cas9*-a. The Cas9 protein binds to the tracrRNA, and to the precursor CRISPR RNA which is processed by RNaseIII to form the crRNA. The crRNA directs binding of Cas9 to the bga promoter and represses transcription. b: The targets used to direct Cas9** to the bga promoter are represented (SEQ ID NO: 531). Putative -35, -10 as well as the bgaA start codon are in bold. c: Beta-galac-
tosidase activity as measure by Miller assay in the absence of targeting and for the four different targets. FIG. 7A-E shows characterization of Cas9** mediated repression. a. The gfpmut2 gene and its promoter, including the −35 and −10 signals are represented together with the position of the different target sites used in the study. b. Relative fluorescence upon targeting of the coding strand. c. Relative fluorescence upon targeting of the non-coding strand. d. Northern blot with probes B477 and B478 on RNA extracted from T5, T10, B10 or a control strain without a target. e. Effect of an increased number of mutations in the 5′ end of the crRNA of B1, T5 and B10.

The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE INVENTION

The terms “polynucleotide”, “nucleotide”, “nucleotide sequence”, “nucleic acid” and “oligonucleotide” are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched nucleo-

5 toides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleo-

10 tide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

In aspects of the invention the terms “chimeric RNA”, “chimeric guide RNA”, “guide RNA”, “single guide RNA” and “synthetic guide RNA” are used interchangeably and refer to the polynucleotide sequence comprising the guide sequence, the tracr sequence and the tracr mate sequence. The term “guide sequence” refers to the about 20 bp sequence within the guide RNA that specifies the target site and may be used interchangeably with the terms “guide” or “spacer”. The term “tracr mate sequence” may also be used interchangeably with the term “direct repeat(s)”.

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

As used herein the term “variant” should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature.

The terms “non-naturally occurring” or “engineered” are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature.

“Complementarity” refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base pairing or other non-

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

As used herein, “stringent conditions” for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent, and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijsen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hy-

20 bridization With Nucleic Acid Probes Part I, Second Chapter “Overview of principles of hybridization and the strategy of nucleic acid probe assay”, Elsevier, N.Y.

“Hybridization” refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogsteen binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PCR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the “complement” of the given sequence.

As used herein, “expression” refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as “gene product.” If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

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

The terms “subject,” “individual,” and “patient” are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals,
sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

The terms “therapeutic agent”, “therapeutic capable agent” or “treatment agent” are used interchangeably and refer to a molecule or compound that confers some beneficial effect upon administration to a subject. The beneficial effect includes elimination of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition.

As used herein, “treatment” or “treating,” or “palliating” or “ameliorating” are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested.

The term “effective amount” or “therapeutically effective amount” refers to the amount of an agent that is sufficient to effect beneficial or desired results. The therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The term also applies to a dose that will provide an image for detection by any one of the imaging methods described herein. The specific dose may vary depending on one or more of: the particular agent chosen, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, the tissue to be imaged, and the physical delivery system in which it is carried.


Several aspects of the invention relate to vector systems comprising one or more vectors, or vectors as such. Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as Escherichia coli, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using 17 promoter regulatory sequences and 17 polymerases.

Vectors may be introduced and propagated in a prokaryote. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in Escherichia coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. Gene 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose 8 binding protein, or protein A, respectively, to the target recombinant protein.


In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). When used in mammalian cells, the vector expression vector’s control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and in the known art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the
nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987, Genes Dev. 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988, Adv. Immunol. 43: 225-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989, EMBO J. 8: 729-733) and immunoglobulins (Baneiji, et al., 1983, Cell 33: 729-740; Queen and Baltimore, 1983, Cell 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989, Proc. Natl. Acad. Sci. U.S.A. 86: 5473-5477), pancreas-specific promoters (Edlund et al., 1985, Science 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990, Science 249: 374-379) and the α-fetoprotein promoter (Campos and Tilghman, 1989, Genes Dev. 3: 537-546).

In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system. In general, CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats), also known as SPINDRs (Spacer Interspersed Direct Repeats), constitute a family of DNA loci that are usually specific to a particular bacterial species. The CRISPR locus comprises a distinct class of interspersed short sequence repeats (SSRs) that were recognized in E. coli (Ishino et al., J. Bacteriol., 169:5429-5433 [1987]; and Nakata et al., J. Bacteriol., 171:3553-3556 [1989]), and associated genes. Similar interspersed SSRs have been identified in Haloferax mediterranei, Streptococcus pyogenes, Anaerococcus, and Mycobacterium tuberculosis (Szymczak et al., Mol. Microbiol., 10:1057-1065 [1993]; Hoeschele et al., Mol. Microbiol., 5:254-263 [1999]; Masepohl et al., Biophys. Acta 1307:26-30 [1996]; and Mojica et al., Mol. Microbiol., 17:85-93 [1995]). The CRISPR loci typically differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SSSRs) (Jansen et al., OMICS J. Integr. Biol., 6:23-33 [2002]; and Mojica et al., Mol. Microbiol., 36:244-246 [2000]). In general, the repeats are short sequences that occur in clusters that are regularly spaced by unique intervening sequences with a substantially constant length (Mojica et al., [2000], supra). Although the repeat sequences are highly conserved between strains, the number of interspersed repeats and the sequences of the spacer regions typically differ from strain to strain (van Embden et al., J. Bacteriol., 182:2539-2540 [2000]). CRISPR loci have been identified in more than 40 prokaryotes (See e.g., Jansen et al., Mol. Microbiol., 43:1565-1575 [2002]; and Mojica et al., [2005]) including, but not limited to Aeropyrum, Pyrococcus, Sulfurolobus, Archaeoglobus, Halocarcula, Methanobacterium, Methanothermobacter, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thermoplasma, Corynebacterium, Mycobacterium, Streptomycetes, Aquifex, Porphyromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myxococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylobacterium, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema, and Thermotoga.

In general, “CRISPR system” refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated (“Cas”) genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (comprising a “direct repeat” and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a “spacer” in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus. In some embodiments, one or more elements of a CRISPR system is derived from a type I, type II, or type III CRISPR system. In some embodiments, one or more elements of a CRISPR system is derived from a particular organism comprising an endogenous CRISPR system, such as Streptococcus pyogenes. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, “target sequence” refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. In some embodiments, the target sequence may be within an organelle of a eukaryotic cell, for example, mitochondrion or chloroplast. A sequence or template that may be used for recombination into the targeted locus comprising the target sequences is referred to as an “editing template” or “editing polynucleotide” or “editing sequence”. In aspects of the invention, an exogenous template polynucleotide may be referred to as an editing template. In an aspect of the invention the recombination is homologous recombination.

Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence. In some embodiments, the tracr sequence has sufficient complementarity to a tracr-mate sequence to hybridize and participate in formation of a CRISPR complex. As with the target sequence, it is believed that complete complementarity is not needed, provided there is sufficient to be functional. In some embodiments, the tracr sequence has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. In some embodiments, one or more vectors driving expression of one or more elements of a CRISPR system are introduced into a host cell such that expression of the elements of the CRISPR system direct formation of a CRISPR complex at one or more target sites. For example, a Cas enzyme, a guide sequence linked to a tracr-mate sequence, and a tracr sequence could each be operably linked to separate regulatory elements on separate vectors. Alternatively, two or more of the elements expressed
from the same or different regulatory elements, may be combined in a single vector, with one or more additional vectors providing any components of the CRISPR system that are not included in the first vector. CRISPR system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to (“upstream” of) or 3’ with respect to (“downstream” of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a CRISPR enzyme and one or more of the guide sequence, tracr mate sequence (optionally operably linked to the guide sequence), and a tracr sequence embedded within one or more intron sequences (e.g. each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the CRISPR enzyme, guide sequence, tracr mate sequence, and tracr sequence are operably linked to and expressed from the same promoter.

In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a “cloning site”). In some embodiments, one or more insertion sites (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. In some embodiments, a vector comprises an insertion site upstream of a tracr mate sequence, and optionally downstream of a regulatory element operably linked to the tracr mate sequence, such that following insertion of a guide sequence into the insertion site and upon expression the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell. In some embodiments, a vector comprises two or more insertion sites, each insertion site being located between two tracr mate sequences so as to allow insertion of a guide sequence at each site. In such an arrangement, the two or more guide sequences may comprise two or more copies of a single guide sequence, two or more different guide sequences, or combinations of these. When multiple different guide sequences are used, a single expression construct may be used to target CRISPR activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell.

In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, such as a Cas protein. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Cas11 and CasX12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Cas5, Cas2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Cx17, Csx14, Csx10, Csx6, CasX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologs thereof, or modified versions thereof. These enzymes are known; for example, the amino acid sequence of *S. pyogenes* Cas9 protein may be found in the SwissProt database under accession number Q992/W2. In some embodiments, the unmodified CRISPR enzyme has DNA cleavage activity, such as Cas9. In some embodiments the CRISPR enzyme is Cas9, and may be Cas9 from *S. pyogenes* or *S. pneumoniae*. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, a vector encodes a CRISPR enzyme that is mutated to with respect to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from *S. pyogenes* converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9a nickase include, without limitation, H840A, N854A, and N863A. In some embodiments, a Cas9 nickase may be used in combination with guide sequence(s), e.g., two guide sequences, which target respectively sense and antisense strands of the DNA target. This combination allows both strands to be nicked and used to induce NHEJ. Applicants have demonstrated (data not shown) the efficacy of two nickase targets (i.e., sgRNAs targeted at the same location but to different strands of DNA) in inducing mutagenic NHEJ. A single nickase (Cas9-D10A with a single sgRNA) is unable to induce NHEJ and create indels but Applicants have shown that double nickase (Cas9-D10A and two sgRNAs targeted to different strands at the same location) can do so in human embryonic stem cells (hESCs). The efficiency is about 50% of nucleosome (i.e., regular Cas9 without D10 mutation) in hESCs.

As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III) may be mutated to produce a mutated Cas9 substantially lacking all DNA cleavage activity. In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity. In some embodiments, a CRISPR enzyme is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is less than about 25%, 10%, 5%, 1%, 0.1%, or lower with respect to its non-mutated form. Other mutations may be useful; where the Cas9 or other CRISPR enzyme is from a species other than *S. pyogenes*, mutations in corresponding amino acids may be made to achieve similar effects.

In some embodiments, an enzyme coding sequence encoding a CRISPR enzyme is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, mouse, rat, rabbit, dog, or non-human primate. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily avail-
able, for example, at the “Codon Usage Database”, and these tables can be adapted in a number of ways. See Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence databases: status for the year 2000” Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as GeneForge (Aptagen; Jacobias, P. A.), are also available. In some embodiments, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or all, or more, or all codons) in a sequence encoding a CRISPR enzyme correspond to the most frequently used codon for a particular amino acid.

In some embodiments, a vector encodes a CRISPR enzyme comprising one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In some embodiments, the CRISPR enzyme comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g., one or more NLSs at the amino-terminus and one or more NLSs at the carboxy-terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In a preferred embodiment of the invention, the CRISPR enzyme comprises at most 6 NLSs. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. Typically, an NLS consists of one or more short sequences of positively charged lysines or arginines exposed on the protein surface, but other types of NLSs are known. Non-limiting examples of NLSs include NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKRKKRV (SEQ ID NO: 1); the NLS of nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPATKAKQOGAKK (SEQ ID NO: 2)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO: 3) or RQRNRELKRSP (SEQ ID NO: 4); the hRNP1 M9 NLS having the sequence NQSSNSFGMGPNGFSGGSGGPPPQQTAKPQQGQY (SEQ ID NO: 5); the sequence RMRIZRKKNGGKDTAELRRRVEVVELKAKKDEQLKRRNV (SEQ ID NO: 6) of the IBB domain from importin-alpha; the sequences VSRRKKPRP (SEQ ID NO: 7) and PPKKARED (SEQ ID NO: 8) of the myosin T protein; the sequence PQPKKPKL (SEQ ID NO: 9) of human p53; the sequence SAIRKKKKKAMP (SEQ ID NO: 10) of mouse e-abl IV; the sequences DLRK (SEQ ID NO: 11) and PKQKKRR (SEQ ID NO: 12) of the influenza virus NS1; the sequence RKLKKKKKL (SEQ ID NO: 13) of the Hepatitis virus delta antigen; the sequence RKKKFLKRR (SEQ ID NO: 14) of the mouse Mx1 protein; the sequence KRRKQDKDVGDVEAVKSSK (SEQ ID NO: 15) of the human poly(ADP-ribose) polymerase; and the sequence RKFLQAGMNLARKTTK (SEQ ID NO: 16) of the steroid hormone receptors (human) glucocorticoid.

In general, the one or more NLSs are of sufficient strength to drive accumulation of the CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the CRISPR enzyme, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the CRISPR enzyme, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI). Examples of detectable markers include fluorescent proteins (such as Green fluorescent proteins, or GFP; RFP; CFP), and epitope tags (HIA tag, flag tag, SNAP tag). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry. Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of CRISPR complex formation (e.g., assay for DNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by CRISPR complex formation and/or CRISPR enzyme activity), as compared to a control no exposed to the CRISPR enzyme or complex, or exposed to a CRISPR enzyme lacking the one or more NLSs.

In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybrize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome. For example, for the S. pyogenes Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMUMMMNNNNNNNNNNNNXGG (SEQ ID NO: 534) where XGG (SEQ ID NO: 535) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an S. pyogenes Cas9 target site of the
form MMMMNMNMMMMNNXGG (SEQ ID NO: 536) where NNNNNNNNNNNXGGX (SEQ ID NO: 537) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. For the *S. thermophilus* CRISPR1Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMNNNNNNNNNNNNXGGXG (SEQ ID NO: 17) where NNNNNNNNNNNXXAGAA (SEQ ID NO: 18) (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. thermophilus* CRISPR1 Cas9 target site of the form MMMMNMNMMMMNNNNNNNNXAGAA (SEQ ID NO: 20) (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. The *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMNMNMMMMNNNNNNXGGXG (SEQ ID NO: 538) where NNNNNNNNNNNXGGXG (SEQ ID NO: 539) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. pyogenes* Cas9 target site of the form MMMMNMNMMMMNNNNNNNXXGXXG (SEQ ID NO: 540) where NNNNNNNNNNNNXXGXXG (SEQ ID NO: 541) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. In each of these sequences “M” may be A, G, T, or C, and need not be considered in identifying a sequence as unique.

In some embodiments, a guide sequence is selected to reduce the degree of secondary structure within the guide sequence. Secondary structure may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mfold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148).

Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g. A. R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62). Further algorithms may be found in U.S. application Ser. No. 61/836,080; incorporated herein by reference.

In general, a tracr sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr sequences in a cell containing the corresponding tracr sequence; and (2) formation of a CRISPR complex comprises the tracr sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence. In some embodiments, the degree of complementarity between the tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. Example illustrations of optimal alignment between a tracr sequence and a tracr mate sequence are provided in FIGS. 12B and 13B. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. Preferred loop forming sequences for use in hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA. However, longer or shorter loop sequences may be used, as may alternative sequences. The sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In some embodiments, the single transcript further includes a transcription termination sequence; preferably this is a poly-T sequence, for example six T nucleotides. An example illustration of such a hairpin structure is provided in the lower portion of FIG. 13B, where a portion of the sequence 5′ of the final “N” and upstream of the loop corresponds to the tracr mate sequence, and the portion of the sequence 3′ of the loop corresponds to the tracr sequence. Further non-limiting examples of single polynucleotides comprising a guide sequence, a tracr mate sequence, and a tracr sequence are as follows (listed 5′ to 3′), where “N” represents a base of a guide sequence, the first block of lower case letters represent the tracr mate sequence, and the second block of lower case letters represent the tracr sequence, and the final poly-T sequence represents the transcription terminator: (1) NNNNNNNNNNNNNNNNTTGGT-tctacttcagatttGAAATAaactttggaaactcagaagtctagc-gattcatgeac-gaactcagctat-cgcatttggttcggatggtttagttatTTTTTT TT (SEQ ID NO: 21); (2) NNNNNNNNNNNNNNTTTTGTT-ctacttcGAAAGtgcaagcacttcagaataaagtctgctgcctgtctgatTTTTTT TT (SEQ ID NO: 22); (3) NNNNNNNNNNNNNNNNTTTTGTT-ctacttcGAAATgcaagcacttcagaataaagtctgctgcctgtctgatTTTTTTTT (SEQ ID NO: 23); (4) NNNNNNNNNNNNNNNNTTTTGTT-tcacttcGAAATAaactttggaaactcagaagtctagc-gattcatgeac-gaactcagctat-cgcatttggttcggatggtttagttatTTTTTT (SEQ ID NO: 24); (5) NNNNNNNNNNNNNNNNTTTTGTT-tcacttcGAAATAaactttggaaactcagaagtctagc-gattcatgeac-gaactcagctat-cgcatttggttcggatggtttagttatTTTTTTT (SEQ ID NO: 25); and (6) NNNNNNNNNNNNNNNNTTTTGTT-tcacttcGAAATAaactttggaaactcagaagtctagc-gattcatgeac-gaactcagctat-cgcatttggttcggatggtttagttatTTTTTTTTT (SEQ ID NO: 26). In some embodiments, sequences (1) to (3) are used in combination with Cas9 from *S. thermophilus* CRISPR1. In some embodiments, sequences (4) to (6) are used in combination with Cas9 from *S. pyogenes*. In some embodiments, the tracr sequence is a separate transcript from a transcript comprising the tracr mate sequence (such as illustrated in the top portion of FIG. 13B). In some embodiments, a recombination template is also provided. A recombination template may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. In some embodiments, a recombination template is designed to serve as a template in homologous recombination, such as within or near a target sequence nicked or cleaved by a CRISPR enzyme as a part of a CRISPR complex. A template polynucleotide may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In some embodiments,
the template polynucleotide is complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a template polynucleotide might overlap with one or more nucleotides of a target sequences (e.g. about or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more nucleotides). In some embodiments, when a template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

In some embodiments, the CRISPR enzyme is part of a fusion protein comprising one or more heterologous protein domains (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the CRISPR enzyme). A CRISPR enzyme fusion protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains. Examples of protein domains that may be fused to a CRISPR enzyme include, without limitation, epitope tags, reporter gene sequences, and protein domains having one or more of the following activities: methylation activity, demethylation activity, transcriptional activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-galuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). A CRISPR enzyme may be fused to a gene sequence encoding a protein or a fragment of a protein that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. Additional domains that may form part of a fusion protein comprising a CRISPR enzyme are described in US20110059502, incorporated herein by reference. In some embodiments, a tagged CRISPR enzyme is used to identify the location of a target sequence.

In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the invention further provides methods produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. In some embodiments, a CRISPR enzyme in combination with (and optionally complexed with) a guide sequence is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a CRISPR system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-813 (1992), Nabel & Feigner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perriard, British Medical Bulletin 51(3):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology, Doerfler and Böhm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycaiton or lipid: nucleic acid conjugates, naked DNA, artificial virosions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for effective receptor-recognition lipofection of polynucleotides include those of Feigner, WO 91/17424, WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

The preparation of lipid-nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, Science 270:404-410 (1995); Blaeser et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:65-74 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,253,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

The use of DNA or RNA viral based systems for the delivery of nucleic acids takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene delivery system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-actingLTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GalV), Simian Immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700). In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are
capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adenos-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virolology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and 2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also be infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, CR161, CCRF-CEM, MOLT, mLHCD-3, NHDF, HeLa S3, Huh7, Huh4, Huh7, HUVEC, HASMC, HeK9n, HeK9a, MiaPaCll, Panc1, PC-3, Tf1, CTLL-2, CIR, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT1, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J450.1, LRMb, Bel-1, BC-3, IC21, DL2, RAV264.7, NRK, NRK-52E, MRCS, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryonic fibroblasts, 3T3 Swiss, 3T3-L1, 3T3 Mouse fetal fibroblasts, 10.1 mouse fibroblasts, 293-1, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A549, ALC, B16, B55, BP1-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, DsIc5, CHI-1011/2, C16/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhr<sup>+</sup>, COR-L23, COR-L23/CP, COR-L23/50I0, COR-L23/R23, COS-7, CON-434, CML T1, CMT, CT126, D17, DH82, DU145, DcAp, EL4, EM2, EM3, EM76/A1, EM76/AR10, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalci17, H-60, HIMEC, HT-29, Jurkat, JV cells, K562 cells, Ks012, KCT2.2, K,G1, KY01, LCAP, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR-0.2R, MONO-MAC 6, MTX-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NHI-3T3, NALM1-1, NW-145, OPCN/OPTC cell lines, Peer, PNT-1A/PNT-2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, SF9, SKBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic species thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassas, Va.)). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. In certain embodiments, the organism or subject is a plant. In certain embodiments, the organism or subject is a plant. Methods for producing transgenic plants and animals are known in the art, and generally begin with a method of cell transfection, such as described herein.

In one aspect, the invention provides methods for modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such a binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

With recent advances in crop genetics, the ability to use CRISPR-Cas systems to perform efficient and cost effective gene editing and manipulation will allow the rapid selection and comparison of single and multiplexed genetic manipulations to transform such genomes for improved production and enhanced traits. In this regard reference is made to US patents and publications: U.S. Pat. No. 6,603,061; Agrobacterium-Mediated Plant Transformation Method; U.S. Pat. No. 7,686,149—Plant Genome Sequences and Uses Thereof and US
2009/0100536—Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al. "Crop genomics: advances and applications" Nat Rev Genet. 2011 Dec; 15(12): 85-96 are also herein incorporated by reference in their entirety. An advantageous embodiment of the invention, the CRISPR/Cas9 system is used to engineer microalgae (Example 15). Accordingly, reference herein to animal cells may also apply, mutans mutans, to plant cells unless otherwise apparent.

In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be in vivo, ex vivo or in vitro. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal or plant (including microalgae), and modifying the cell or cells. Culturing may occur at any stage ex vivo. The cell or cells may even be re-introduced into the non-human animal or plant (including micro-algae).

In plants, pathogens are often host-specific. For example, *Fusarium oxysporum* f. sp. *lycopersici* causes tomato wilt, but attacks only tomato, and *F. oxysporum* f. sp. *dianthic Puccinia graminis* causes wheat rust. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible. There can also be Horizontal Resistance, e.g., partial resistance against all races of a pathogen, typically controlled by many genes and Vertical Resistance, e.g., complete resistance to some races of a pathogen but not to other races, typically controlled by a few genes. In a Gene-for-Gene level, plants and pathogens evolve together, and the genetic changes in both are balanced in favor of each other. Accordingly, using Natural Variability, breeders combine most useful genes for Yield, Quality, Uniformity, Hardiness, Resistance. The sources of resistance genes include native or foreign Varieties, Heirloom Varieties, Wild Plant Relatives, and Induced Mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can analyze the genome of sources of resistance genes, and in Varieties having desired characteristics or traits employ the present invention to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a target mate sequence and one or more insertion sites for inserting a guide sequence upstream of the target mate sequence, wherein the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the target mate sequence that is hybridized to the target sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence. Elements may be provide individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, or a tube. In some embodiments, the kit includes instructions in one or more languages, for example in more than one language.

In some embodiments, a kit comprises one or more reagents for use in a process utilizing one or more of the elements described herein. Reagents may be provided in any suitable container. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (e.g., in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In some embodiments, the buffer has a pH from about 7 to about 10. In some embodiments, the kit comprises one or more oligonucleotides corresponding to a guide sequence for insertion into a vector so as to enable link the guide sequence and a regulatory element. In some embodiments, the kit comprises a homologous recombination template polynucleotide.

In one aspect, the invention provides methods for using one or more elements of a CRISPR system. The CRISPR complex of the invention provides an effective means for modifying a target polynucleotide. The CRISPR complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types. As such the CRISPR complex of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide. The guide sequence is linked to a target mate sequence, which in turn hybridizes to a tracr sequence.

The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR complex.

The target polynucleotide of a CRISPR complex may include a number of disease-associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides as listed in U.S. provisional patent applications 61/736,527 and 61/748,427 having Broad reference BI-2011/008/WSGR and BI-2011/008/WSGR respectively, both entitled SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION filed on Dec. 12, 2012 and Jan. 2, 2013, respectively, the contents of all of which are herein incorporated by reference in their entirety.

Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease
associated gene or polynucleotide. A “disease-associated” gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcripts or translated products may be known or unknown, and may be at a normal or abnormal level.

Examples of disease-associated genes and polynucleotides are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

Examples of disease-associated genes and polynucleotides are listed in Tables A and B. Disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web. Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in Table C.

Mutations in these genes and pathways can result in production of improper proteins or proteins in improper amounts which affect function. Further examples of genes, diseases and proteins are hereby incorporated by reference from U.S. Provisional applications 61/736,527 filed on Dec. 12, 2012 and 61/748,427 filed Jan. 2, 2013. Such genes, proteins and pathways may be the target polynucleotide of a CRISPR complex.

### TABLE A

<table>
<thead>
<tr>
<th>DISEASE/DISORDERS</th>
<th>GENE(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplasia</td>
<td>PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; NOTCH1; NOTCH2; NOTCH3; AKT;</td>
</tr>
<tr>
<td></td>
<td>AKT2; AKT3; HIF1a; HIF1b; HIF2; HIF2a; PPAR alpha; PPAR gamma; WT1;</td>
</tr>
<tr>
<td></td>
<td>WTI (Wilms Tumor); FGF Receptor Family members (5 members: 1, 2, 3, 4, 5)</td>
</tr>
<tr>
<td>Age-related Macular</td>
<td>Vldlr; Cer2</td>
</tr>
<tr>
<td>Degeneration</td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Neuregulin1 (Ng1); Erb-b (receptor for Neuregulin); Complexin1 (Cpx1);</td>
</tr>
<tr>
<td></td>
<td>Tph1; Tryptophan hydroxylase; Tph2; Tryptophan hydroxylase 2; Neurox1;</td>
</tr>
<tr>
<td></td>
<td>GSK3b; GSK3b; GSK3b; GSK3b; GSK3b;</td>
</tr>
<tr>
<td>Disorders</td>
<td>SHT-1 (Slc6a4); COMT; DRD (Drla); SLC6A3; DAOD;</td>
</tr>
<tr>
<td></td>
<td>DYNKB1; Dna (Dn1)</td>
</tr>
<tr>
<td>Trinucleotide Disorders</td>
<td>HTT (Huntington’s D)</td>
</tr>
<tr>
<td>Repeat Disorders</td>
<td>FXN/X25 (Friedreich’s Ataxia); ATX3 (Machado-Joseph’s D); ATXN1;</td>
</tr>
<tr>
<td></td>
<td>ATXN2 (spinocerebellar ataxias); DMPK (myotonic dystrophy); Atrophin-1;</td>
</tr>
<tr>
<td></td>
<td>Atal (DRPLA D); CBP (Ceb-BP - global instability); VDLR (</td>
</tr>
<tr>
<td></td>
<td>Alzheimer’s); Ataxa10</td>
</tr>
<tr>
<td>Fragile X Syndrome</td>
<td>FMR2; FTX1; FTX2; nGLUR3</td>
</tr>
<tr>
<td>Secretive Related</td>
<td>AGP-1 (alpha and beta); Presenilin (Psen1); nicastrin</td>
</tr>
<tr>
<td>Disorders</td>
<td>Nestin; PEN-2</td>
</tr>
<tr>
<td>Others</td>
<td>Nost1; Parp1; Nat1; Nat2</td>
</tr>
<tr>
<td>Prion - related disorder</td>
<td>Prp</td>
</tr>
<tr>
<td>ALS</td>
<td>SOD1; ALS2; STX; FUS; TARDBP; VEGF (VEGF-a); VEGF-b; VEGF-c</td>
</tr>
<tr>
<td>Drug addiction</td>
<td>Prkce (alcohol); Drd2; Drd4; ACHAT (alcohol); GrhA2;</td>
</tr>
<tr>
<td></td>
<td>Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Grin1 (alcohol)</td>
</tr>
<tr>
<td>Autism</td>
<td>MeCP2; BZRAP1; MDAA2; Sema5A; Neurox1; Fragile X (</td>
</tr>
<tr>
<td></td>
<td>FMR2 (AFTF); FTX1; FTX2; Mqhp5)</td>
</tr>
<tr>
<td>Alzheimer’s Disease</td>
<td>E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterin; PS1;</td>
</tr>
<tr>
<td></td>
<td>SORL1; CR1; Vldlr; Ubata; Ub3a; CHID4 (Aap1, Aqpaporin 1); Uchi1;</td>
</tr>
<tr>
<td></td>
<td>Uchh; APP</td>
</tr>
<tr>
<td>Inflammation</td>
<td>IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA4)); IL-17b;</td>
</tr>
<tr>
<td></td>
<td>IL-17c; IL-17d; IL-17f; IL-23; Cx3cl1; ptpi22; TNFα;</td>
</tr>
<tr>
<td></td>
<td>NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b);</td>
</tr>
<tr>
<td></td>
<td>CTLA4; Cx3cl1</td>
</tr>
<tr>
<td>Parkinson’s Disease</td>
<td>α-Synuclein; DJ-1; LRRK2; Parkin; PINK1</td>
</tr>
</tbody>
</table>

### TABLE B

<table>
<thead>
<tr>
<th>Blood and coagulation diseases and disorders</th>
<th>Anemia (CDAN1; CDAN1; CDAP; RB519; DBA; PCKR; PK1; NTSC3; UMPH1; PNS1; RIHAG; RII90A; NRAMP2; SPTB; ALAS2; ANHI1; ASB; ABCB7, ABC7; ASAT); Bare lymphocyte syndrome (TAPBP; TPSN; TAP2; ABC3; PSF2; RING11; MHC2TA; C2TA; RFX3; RFXAP; RFX5); Bleeding disorders (T7BAA2; P2RX1; P2Y1); Factor H and</th>
</tr>
</thead>
</table>
**Table B-continued**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell dysregulation and oncology diseases and disorders</td>
<td></td>
</tr>
<tr>
<td>B-cell non-Hodgkin lymphoma (BC17A, BCL7); Leukemia (TAL1)</td>
<td></td>
</tr>
<tr>
<td>Tcell, SC1, TAL2, FLI1, NBS1, NBS2, ZNFSN1A, IJKI, LYT1</td>
<td></td>
</tr>
<tr>
<td>HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMP9, AF10, AHI1</td>
<td></td>
</tr>
<tr>
<td>CHEFB, CHEF, CHIC2, BTL, FLI1, KIT, PBT, LPI, NPM1, NUP214</td>
<td></td>
</tr>
<tr>
<td>D9S46E, CAN; CAIN, RENXII, CBEAB, AML1, WHSC1L1, NDDX1</td>
<td></td>
</tr>
<tr>
<td>FLI3, AFIP1, NPM1, ZNF145, PLZF, PML, MYL, STAT5B</td>
<td></td>
</tr>
<tr>
<td>AF10, CALM, CLTH, ARLI, ARS1, PIK3R7, P2X7, BCR, CML, PHL, ALL, GABF, EF1</td>
<td></td>
</tr>
<tr>
<td>NCI, AMBL1, ABK1, DRI1, ING1, CXCL12, SDF1L1</td>
<td></td>
</tr>
<tr>
<td>FAS, CD95, AIPL1; Combined immunodeficiency, (IL-2R, SCID, XCI, SDIX, IMD4)</td>
<td></td>
</tr>
<tr>
<td>HIV-1 (CCL5, SCYA5, D1S1336, TFC228)</td>
<td></td>
</tr>
<tr>
<td>HIV susceptibility or infection (IL-10, CS1, CMKRB2, CCR2, CMKRB3,</td>
<td></td>
</tr>
<tr>
<td>CCR2, CMKRB5, CCR3, CCR5, CCR6, CCR7)</td>
<td></td>
</tr>
<tr>
<td>Immunodeficiencies (CD4, CD3G, AICDA, AID, HIVG, TNPFSF5, C040, UNG,</td>
<td></td>
</tr>
<tr>
<td>DGL, BCIGM4, TNFSF5, CD40, UNG, DGL, BCIGM4, TNPFSF5, CD40, UNG, DGL,</td>
<td></td>
</tr>
<tr>
<td>BCIGM4, CCR5, CD40, LG, HOM1, IGM, FOXP3, PI3K, AID, XPID, PDX, TNPFSF5,</td>
<td></td>
</tr>
<tr>
<td>TAC1; Inflammation (IL-10, IL-1, IL-1b, IL-13, IL-17a, IL-17b, IL-17c,</td>
<td></td>
</tr>
<tr>
<td>IL-17d, IL-23, CCL4O)</td>
<td></td>
</tr>
<tr>
<td>Severe combined immunodeficiencies (SCID1, SCID2, SCID3)</td>
<td></td>
</tr>
<tr>
<td>SCID1, SCID2, SCID3, SCID4, SCID5, SCID6</td>
<td></td>
</tr>
<tr>
<td>Inflammation and autoimmune related diseases and disorders</td>
<td></td>
</tr>
<tr>
<td>Amyloid neuropathy (TTR, PALB), Amyloidosis (APOA1, A2, AAA, CVAR, ADI,</td>
<td></td>
</tr>
<tr>
<td>ASN, FGA, LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8, KRT11A, NAIC, TESC292,</td>
<td></td>
</tr>
<tr>
<td>KIAA1988)</td>
<td></td>
</tr>
<tr>
<td>Cystic fibrosis (CFTR, ABCB1, CYF, MARP1, G6PD, G6PD1)</td>
<td></td>
</tr>
<tr>
<td>Metabolic, liver, kidney and protein diseases and disorders</td>
<td></td>
</tr>
<tr>
<td>Becker muscular dystrophy (DMD, BMD, MFS)</td>
<td></td>
</tr>
<tr>
<td>Duchenne Muscular Dystrophy (DMD, BMD), myotonic muscular dystrophy</td>
<td></td>
</tr>
<tr>
<td>(MMD, MND), Bechler-Duhamel muscular dystrophy (LMNA, LMN1, LMN1, DMD2,</td>
<td></td>
</tr>
<tr>
<td>FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, SNM1, EMD2, FPLD, CMD1A, HGPS,</td>
<td></td>
</tr>
<tr>
<td>LGMD1B, CMD1A; Facioscapulohumeral muscular dystrophy (FSHD1A, FSHD1A)</td>
<td></td>
</tr>
<tr>
<td>Neurological and neuronal diseases and disorders</td>
<td></td>
</tr>
<tr>
<td>ALS (SOD1, ALS2, STEF, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c);</td>
<td></td>
</tr>
<tr>
<td>Alzheimer disease (APP, AAV, CAV1, AD1, ADPE, AD2, PSEN2,        AD4, STM2,</td>
<td></td>
</tr>
<tr>
<td>APBB2, F65L1, NO3, PLAT, URE, ARB, DCPI, ACE1, MPD, PAC1P1, PAXIP1, PTP</td>
<td></td>
</tr>
<tr>
<td>1, BM13, BMH, PSEN1, AD3); Autism (MECP2, BZRAP1, MDX2A, SMAS3, NEXN</td>
<td></td>
</tr>
<tr>
<td>N1L1, BM13, BMX7, NOS3, NOS4, KIAA1206, AUTOX2); Fragile X Syndrome</td>
<td></td>
</tr>
<tr>
<td>(FMR2, FXR1, FXR2, mGLUR5); Huntington's disease and disease like disorders (HD, IT15,</td>
<td></td>
</tr>
<tr>
<td>PKR, PRLP, FRZ1, JPD1, HLD2, TBP, SCA17); Parkinson disease</td>
<td></td>
</tr>
<tr>
<td>(NQ-A2, NURR1, NOT, TINUR, SNAIP7, TBP, SCA17, SNA,</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE C

<table>
<thead>
<tr>
<th>CELLULAR FUNCTION</th>
<th>GENES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PI3K/AKT Signaling</strong></td>
<td>PRKCE, ITGA5, ITGA6, IRAK1, PRKAA2, EIF2AK2, PTEN, EIF4E, PRKCE, GRK6, MAPK1, TSC2, PLK1, AKT2, IKBK; PIK3CA; CDK8, CDKN1B, NFKB2; BCL2, PIK3CB, PPP2R1A, MAPK8, BCL2L1, MAPK3, TSC2, ITGA1, Kras, EIF1EB2, REL-A, PIK3C2; NOS3, PRKAA1, MAPK9, CDK2; PPP3CA, PIM1; ITGB7, YWHAZ, ILK, TP53, RAF1, IKKG, REL, DYRK1A; CDKN1A; ITGB1; MAPK2, AKT1; JAK2; PIK3R1; CHUK; PDK1; PPP3CA, CTDN1, MAPK1, NFKB1; PAK7, ITGAM3, CCND1, GSK3A; FRAP1, SFN, TGAM2, TK1; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK; HSP90AA1; RP56KB1</td>
</tr>
<tr>
<td><strong>ERK/MAPK Signaling</strong></td>
<td>PRKCE, ITGA5; ITGA6; HSPB1; IRAK1; PRKAA2, EIF2AK2, RAC1, RAP1A, TNN1, EIF4E, ELK1, GRK6, MAPK1, RAC2, PLK1, AKT2, PIK3CA, CDK8, CREB1, PRKCI, PTK2, ROS1, PRKCA, PIK3CB, PPP2R1A, PIK3C3, MAPK8, MAPK3, ITGA1, ET51, Kras, MYCN, EIF4EBP1, PPARG, PIK3CD, PRKAA1, MAPK9, SRC, CDK2, PPP3CA, PIM1, PIK3CA2, ITGB7, YWHAZ, PPP1CC, KSR1, PXN, RAF1, FYN, DYRK1A, ITGB1, MAPK2, AK4, PIK3R1, TSG101, PPP3CA, MAPK1, PAK5, ITGB3, IRS1, ITGAM2, MYC, TT, CSNK1A1, CRKL, BRAF, ATF4, PRKCA, SRC, STAT1, SGK</td>
</tr>
<tr>
<td><strong>Glucocorticoid Receptor Signaling</strong></td>
<td>RAC1, TAF4B, EP100, SMAD2, TRAF6, PCEA, ELK1, MAPK1, SMAD3, AKT2, IKBK; NOS2A, UBE2L1, PIK3CA, CREB1, ROS1, HSPA5, NFKB2, BCL2, MAP3K14, STAT5B, PIK3CB, PIK3C3, MAPK8, BCL2L1, MAPK3, TSC22D3, MAPK10, NRP1, Kras, MAPK13, RELA, STAT5A, MAPK8, NOS2A, PIK3, NRC1, PIK3C2A, CDKN1C, TRAF2, SERpine1, NCOA3, MAPK14, TNF, RAF1, IKKGB, MAP3K7, CREBBP, CDKN1A, MAP2K2, JAK1, IL6, NCOA2, AKT1, JAK2, PIK3R1, CHUK, STAT3, MAPK11, NFKB1, TGFBR1, ESRI, SMAD4, CEBPB, JUN, AR, AKT3, CCL2, MMP1, STAT1, IL6, HSP90AA1</td>
</tr>
<tr>
<td><strong>Axonal Guidance Signaling</strong></td>
<td>PRKCE, ITGA5; ROCK1; ITGAV, CXC4, ADAM12, IGFI, RAC1, RAP1A, EIF4E, PRKCE, NRP1, NTRK2, IGFI, RAC1, RAP1A, EIF4E, PRKCE, NRP1, NTRK2</td>
</tr>
<tr>
<td>CELLULAR FUNCTION</td>
<td>GENES</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Ephrin Receptor Signaling</td>
<td>ARIHGEF7; SMO; ROCK2; MAPK3; PGE; RAC2; PTPN11; GNAS; AKT2; PI3KCA; ERBB2; PRKCK1; PTK2; CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11; PRKD1; GN2B2; ABL1; MAPK3; ITGA1; Kras; RHOA; PRKCD; PI3KCA; ITGB7; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAPK2; PAK4; ADAM17; AKT1; PIK3R1; GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CD42; VEFGA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA</td>
</tr>
<tr>
<td>Actin Cytoskeleton Signaling</td>
<td>PRKCE; ITGAM; ROCK1; ITGAV; CXCR4; IRR1; AKR1A; PRRK2A2; EGFRA; RAC1; RAS; ARHGEF7; GRK6; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14; CXCL12; MAPK8; GN2B2; ABL1; MAPK3; ITGA1; Kras; RHOA; PRKCD; PI3KCA; ITGA1; MAPK9; SRC; CDK2; PI3K; ITGB7; PXN; RAF1; FYN; DRIK1A; ITGB1; MAP2K2; PAK4; AKT1; ITGAV; ADAM10; MAP2K1; PAK3; ITGB2; CD42; VEFGTA; EPHD2; EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4; AKT3; SGK</td>
</tr>
<tr>
<td>Huntington’s Disease Signaling</td>
<td>ACTN4; PRKCE; ITGAM; ROCK1; ITGAV; IRR1; PRRK2A2; EGFRA; RAC1; RAS; ARHGEF7; GRK6; ROCK2; MAPK1; PGF; RAC2; PLK1; AKT2; PI3KCA; CDK8; PTK2; CFL1; PI3KCB; MYB10; DAPAP1; PI3K3; MAPK8; FSR3; MAP3K; SLCA11; ITGA1; Kras; RHOA; PRKCD; PI3KCA; ITGB7; PP1CC; PXN; VIL2; RAF1; GSN; DRIK1A; ITGB1; MAP2K2; PAK4; PI3KCA; PIK3R1; MAP2K1; PAK3; ITGB6; CD42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRAF; VAV3; SGK</td>
</tr>
<tr>
<td>Apoptosis Signaling</td>
<td>PRKCE; ROCK1; BID; IRR1; PRRK2A2; EGFRA; BAK1; BIRC4; GRK6; MAPK1; CAPN1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8; BCL2L1; CAPN1; MAPK3; CASP8; RELA; ITGAV; PRKCD; PI3KCA; MAPK9; ITGAV; CASP8; PI3KCA; ITGAV; MAP3K; ITGAV; CASP8; РAZ; REL; RAF1; IKBKG; RELJ; CASP9; DRIK1A; MAP2K2; CHUK; IKBAP1; MAP2K1; NFKB1; PAK3; LMNA; CASP2; BIRC2; TTK; CSNK1A1; BRAF; VAX; PI3KCA; SGK; CASP3; BIRC3; PARP1</td>
</tr>
<tr>
<td>B Cell Receptor Signaling</td>
<td>RAC1; PITPN; LYN; ELK1; MAPK1; RAC2; PTPN11; AKT2; IKBKB; PI3KCA; CREB1; SYK; NFKB2; CAMK2A; MAP3K14; PI3KCB; PI3KAP; BCL2L1; ADI1; MAPK3; ETS1; Kras; MAPK3; RELA; PTPN6; MAPK9; EGFR; PI3KCA; BTK; MAP3K14; RAF1; IKBKG; RELB; MAPK7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NFKB1; CD42; GSK3A; FRA1; BCL2; BCL10; JUN; GSK3B; AT4; AKT3; VAV3; PIP6K1B</td>
</tr>
<tr>
<td>Leukocyte Extravasation Signaling</td>
<td>ACTN4; CD44; PRKCE; ITGAM; OCM1; CXCR4; CYBA; RAC1; RAP1A; PRKCE; ROCK2; RAC2; PTPN11; MMP14; PI3KCA; PRKCE; PTK2; PI3KCA; CXCL12; PI3K3; MAPK8; PRKCD; ANB1; MAPK10; CYBB; MAPK5; RHOA; PRKCD; MAPK8; SRC; PI3KCA; BTB; MAP3K14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2; CTNNB1; PI3K1; CTNNB1; CLDN1; CD42; F11R; ITK; CRKL; VAV3; CTN; PI3KCA; MMP1; MMP9</td>
</tr>
<tr>
<td>Integrin Signaling</td>
<td>ACTN4; ITGAM; ROCK1; RAC1; ITGAV; RAC1; PTK2; IKBKB; PI3KCA; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2; CTNNB1; PI3K1; CTNNB1; CLDN1; CD42; F11R; ITK; CRKL; VAV3; CTN; PI3KCA; MMP1; MMP9</td>
</tr>
<tr>
<td>Acute Phase Response Signaling</td>
<td>IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11; AKT2; IKBKB; PI3KCA; FOS; NFKB2; MAP3K14; PI3KCB; MAPK8; RIPK1; MAPK3; IL18; Kras; MAPK3; IL6; RELA; SOCS3; MAPK9; FTL; NRC1; TRAF2; SERPINE1; MAP3K14; TNP; RAF1; PIK1;</td>
</tr>
<tr>
<td>Cellular Function</td>
<td>Genes</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>PTEN Signaling</td>
<td>ITGA1, ITGA3, AC1, PTEN, PFKC2, BCL2L11, MAPK1, RAC1, AKT1, JAK2, PI3K1, CHEK1, STAT3, MAPK1, NRFB1, FRA1, CEBP, JUN, AKT3, IL1R1, IL6</td>
</tr>
<tr>
<td>p53 Signaling</td>
<td>PTEN, TP53, BCL2, TP53, PDK1, PI3K1, CHEK1, TGFBI, BCL2L11, PMAIP1, CHEK2, TNFRSF1B, TP63, RB1, HICAD2, CD2, PDK2, MAPK1, TP53, IRE2D, CDKN1A, HIPK2, AKT1, RIP3, APAF1, CTNNB1, SIRT1, CCND1, PDKC, ATM, SNS, CDKN2A, TNF, SNAI2, GSK3B, BAX, AKT3</td>
</tr>
<tr>
<td>Aryl Hydrocarbon Receptor Signaling</td>
<td>HSPB1, EP300, PAFN, TGF2, RXRA, MAPK1, NQO1, NCOA2, SP1, ARNT, CDKN1B, FOS, CHEK1, SMAC4, NFKB2, MAPK8, ALDH1A1, ATR, ER2F1, MAPK3, NRIP1, CHEK2, RELA, TP73, GSTP1, RB1, SRC, CDK2, AHR, NFE2L2, NCOA3, TP3, TNF, CDKN1A, NCOA2, APTF1, NFKB1, CCND1, ATN, ERF1, CDK2NA, MYC, JUN, ERB2, BAX, CYC1B1, HIS90AA1</td>
</tr>
<tr>
<td>Xenobiotic Metabolism Signaling</td>
<td>PRKCE, EP300, PFK2, RXRA, MAPK1, NQO1, NCOA2, P53CA, ARNT, P53CL, NFKB2, CAMK2A, PIK3C3B, PPP2R1A, PIK3C3, MAPK8, PIK3D1, ALDH1A1, MAPK3, NRIP1, KRAS, MAPK13, PIK3CD, GSTP1, MAPK9, NOS2A, ABCB1, AHR, PPP2CA, FTL, NFE2L2, P53C2A, PPARG1A, MAPK4, TNF, RAPI, CHREBP, MAPK2, PI3K1, PPP2CA, MAPK21, NFKB1, KEAP1, PRKCA, EIF2AK3, JUN, CYP1B1, HIS90AA1</td>
</tr>
<tr>
<td>SAPK/JNK Signaling</td>
<td>PRKCE, IRAK1, PFRK2A, EIF2AK2, RAC1, ELK1, GRK6, MAPK1, GADD45A, AC1, PLK1, AKT2, PIK3CA, FADD, CD8, PIK3CB, PIK3C3, MAPK8, RIP1, GN1L2, IRS1, MAPK3, MAPK4, DAXX, KRAS, PRKCE, PIRKA1, MAPK1, CDK2, PIK3C2A, TRAF5, TP53, LCK, MAPK7, DRYK1A, MAPK2, PIK3R1, MAPK2, PAK3, CDCA4, JUN, CT, CSNK1A1, CRL1, Braf, NDK</td>
</tr>
<tr>
<td>PPARα/RXR Signaling</td>
<td>PPARα2A, EP300, INS, SMAD2, TRAF6, PPARA, FASN, RXRA, MAPK1, SMAD3, GNAS, IKBKB, NCOA2, ABCA1, GNAQ, NFKB2, MAPK14, STAT5B, MAPK8, IRS1, MAPK3, KRAS, RELA, PIK3CA, IKBG, RELB, MAPK7, CREBBP, AKT1, PIK3R1, CHEK1, PDGFRα, NFKB1, TL2B, BCL10, GSK3B, AKT3, TNFAIP3, IL1R1</td>
</tr>
<tr>
<td>NF-κB Signaling</td>
<td>IRAK1, EIF2AK2, EP300, INS, MYD88, PRKCE, TRAF6, TBK1, AKT2, EGFR, IKBKB, PIK3A, BTRC, NFKNB2, MAPK14, PIK3C3B, PIK3C3, MAPK8, RIPK1, HDMI2, KRAS, RELA, PIK3CA, TRAF2, IL2, PDGFRB, TNF, INSR, LCK, IKKβ, RELB, MAPK7, CREBBP, AKT1, PIK3R1, CHEK1, PDGFRα, NFKB1, TL2B, BCL10, GSK3B, AKT3, TNFAIP3, IL1R1</td>
</tr>
<tr>
<td>Neuregulin Signaling</td>
<td>ERBB4, PRKCE, ITGA5, ITGA5, PTEN, PPKC2, ELK1, MAPK1, PPNTN1, AKT2, EGFR, ERBB2, PIK3C2, CDK1B, STAT5B, PIK3D1, MAPK3, ITGA1, KRAS, PRKCD, STAT5A, SRC, ITGB7, RAP1, ITGB1, MAPK2, ADAM17, AKT1, PIK3R1, PDK1K2, MAPK1, ITGB1, ERBB4, TRAP1, PSEN1, ITGAI, MYC, NRG1, CRKL, AKT3, PRKCA, HSPA90AA1, RPS6KB1</td>
</tr>
<tr>
<td>Wnt &amp; Beta catenin Signaling</td>
<td>CD44, EP300, LRP6, DVLT1, CSNK1E, GSK3B, SMO, AKT2, FN1, CDH1, BTRC, GNAQ, MAPK2, PPP2R1A, WNT11, SRC, DKK1, PPP2CA, SOX9, SFRP2, ILK, LEF1, SOX8, TP53, MAPK7, CREBBP, TCF7L2, AKT1, PPP2R5C, WNT5A, LRP5, CTNNB1, TCF7L1, CCND1, GSK3B, DVL1, APOC, CDK2, MYC, CSNK1A1, GSK3B, AKT3, SOX2</td>
</tr>
<tr>
<td>Insulin Receptor Signaling</td>
<td>PTEN, INS, EIF4E, PPNT1, PRKCE, MAPK1, TSC1, PPNTN1, AKT2, CBL, PIK3CA, PIK3C1, PIK3C3, MAPK8, IRS1, MAPK3, TSC2, KRAS, EIF4EB1, SLCC2A4, PIK3C2A, PPP1CC, INSR, RAF1, FYN</td>
</tr>
<tr>
<td>Cellular Function</td>
<td>Genes</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>IL-6 Signaling</td>
<td>HSPB1, TRAF6, MAP3K5, ELK1, MAPK1, PTPN11, IKKα, FOS, NFκB2, MAP3K14, MAPK8, MAP3K3, MAPK6, IL-6ST, Kras, MAPK13, IL-6R, RELA, SOCS1, MAPK9, ABCB1, TRAF2, MAPK14, TNF, RAF1, IKKγ, RELB, MAP3K7, MAPK2, IL-1α, JAK2, CHUK, STAT3, MAP2K1, NFKB1, CEBPB, JUN, IL1R1, SRE, IL6</td>
</tr>
<tr>
<td>Hepatic Cholestasis</td>
<td>PRKCE, IRAK1, IRS, MYD88, PRKCC, TRAF6, PPARα, RXRA, IKKβ, PRKCε, NFKB2, MAP3K14, MAPK8, PRKDI1, MAP10, RELA, PRKCD, PDK6, ABCB1, TRAF2, TLR4, TNF, NBS1, IKKβ, RELMβ, MAPK7, IL-6, CHUK, NRKβ, TTP, NFKB1, ESRR, SREBF1, FGFR4, JUN, IL1R1, PRKCA, IL6</td>
</tr>
<tr>
<td>IGF-1 Signaling</td>
<td>IGF1, PRKCC, ELK1, MAPK1, PTPN11, NEDD4, AKT2, PI3KCA, PRKCC, PT2, FOS, PI3KCB, PIK3C3, MAPK6, IKKβ, IRS1, MAPK3, IGFBP7, KRA5, PIK3CA, YWHAZ, PNN, RAF1, CASP9, MAP2K2, AKT1, PIK3R1, PDK1, MAP2K1, IGF2BP2, SNF, JUN, CYR61, AKT3, FOSO1, SRE, CTGF, RPS6KB1</td>
</tr>
<tr>
<td>NRF2-mediated Oxidative Stress Response</td>
<td>PRKCE, EP300, SOD2, PRKCC, MAPK1, SQSTM1, NQ01, PIK3CA, PRKCI, FOS, PI3KCB, PIK3C3, MAPK8, PRKDI1, MAP3K3, KRA5, PRKCD, GSTP1, MAPK9, FTL, NFE2L2, PIK3C2A, MAPK14, RAF1, MAP3K7, CREBBP, MAP3K2, AKT1, PIK3R1, MAP2K1, PI3R1, JUN, KEAP1, GSK3B, ATF4, EIF2AK3, HSP90AA1</td>
</tr>
<tr>
<td>Hepatic Fibrosis/Hepatic Stellate Cell Activation</td>
<td>EDN1, IGF1, KDR, FLJ1, SMAD2, FGFR1, MET, PGF, SMAD3, EGFR, FAS, CSE1, NFκB2, BCL2, MYC10, IKKβ, IL6R, RELA, TRIM4, PDGFRB, TNF, RELB, IL8, PDGFRB, NFκB1, TGFBR1, SMAD4, VEGFA, BAX, IL1R1, CCL2, HGF, MAP1, STAT1, IL6, CTGF, MMP9</td>
</tr>
<tr>
<td>PPAR Signaling</td>
<td>EP300, NBS, TRAF6, PPARα, RXRA, MAPK1, IKKβ, NCOR2, FOS, NFκB2, MAP3K14, STAT5B, MAPK3, NRP1, KRA5, PPARG, RELA, STAT5A, TRAF2, PPARC, PIK3CB, TNF, NBS1, RAF1, IKKγ, RELB, MAP3K7, CREBBP, MAP3K2, CHUK, PIK3R1, PDK1, MAP2K1, AKT3, VAV3, PRKCA</td>
</tr>
<tr>
<td>Fe Erosion RI Signaling</td>
<td>PRKCE, RAC1, PRKCC, LYN, MAPK1, RAC2, PTPN11, AKT2, PI3KCA, SYK, PRKCI, PI3KCB, PI3K3, MAPK8, PRKDI1, MAP3K3, KRA5, PRKCD, MAPK9, PIK3CB2, BTK, MAPK14, TNF, RAF1, FYN, MAP2K2, AKT1, PIK3R1, CHUK, PDK1, STAT3, MAP2K1, NFKB1, BRAF, ATF4, AKT3, PRKCA</td>
</tr>
<tr>
<td>G-Protein Coupled Receptor Signaling</td>
<td>PRKCE, RAPI, A, RGS16, MAPK1, NGS, AKT2, JNK8, PIK3CA, CREBL1, GNAS, NFκB2, CAMK2A, PI3KCB, PIK3C3, MAPK3, KRA5, RELA, SRC, PI3KCB2, RAF1, IKKβ, RELB, FYN, MAP2K2, AKT1, PIK3R1, CHUK, PDK1, STAT3, MAP2K1, NFKB1, BRAF, ATF4, AKT3, PRKCA</td>
</tr>
<tr>
<td>Inositol Phosphate Metabolism</td>
<td>PRKCE, IRAK1, PRKAA2, EL2AK2, PTEN, GRK6, MAPK1, PIK1, AKT2, PI3KCA, CDK4, PI3KCB, PI3K3, MAPK8, MAP3K3, PRKCD, PI3KCA1, MAPK9, CDK2, PI1, PIK3C2A, D2R, MAP2K2, PI3KIA, PIK3R1, MAP2K1, PK3, ATM, TTK, CSNK1A1, BRF1, SGK</td>
</tr>
<tr>
<td>PDGF Signaling</td>
<td>EIF2AK2, ELK1, ABL1, MAPK1, PIK3CA, FOS, PI3KCB, PIK3C3, MAPK8, CAVI, ABL1, AKT1, KRA5, SRC, PIK3CB2, PDGFRB, RAF1, MAP2K2, JAK1, JAK2, PIK3R1, PIK3C, STAT3, SPH1, MAP2K1, MYC, JUN, CRKL, PRKCA, SRF, STAT1, PI3K</td>
</tr>
<tr>
<td>VEGF/Platelet Signaling</td>
<td>ACTN4, ROK1, KDR, FUT1, ROCK2, MAPK1, PDP1, AKT2, PI3KCA, AKT1, PRKCD, BCL2L1, MAPK3, KRA5, HIF1A, NOX3, PI3K2A, PAI1, RAS, MAP2K2, ELAVL1, AKT1, PIK3R1, MAP2K1, AKT3, VAV3, PI3K</td>
</tr>
<tr>
<td>Natural Killer Cell Signaling</td>
<td>PRKCE, RAC1, PRKCC, MAPK1, RAC2, PTPN11, KIR2DL3, AKT2, PI3KCA, SYK, PRKCI, PI3KCB, PIK3C3, PRKDI1, MAPK3, KRA5, PRKCD, PTPN6, PIK3C2A, LCK, RAF1, FYN, MAP2K2, AKT1, PIK3R1, MAP2K1, PAK3, AKT3, VAV3, PI3K</td>
</tr>
<tr>
<td>Cell Cycle: G1/S Checkpoint Regulation</td>
<td>HDAC4, SMAD5, SUVR31H1, HDAC5, CDKN1B, BRCC, ATR, ABL1, E2F1, HDAC2, RBL1, HDAC11, HDAC9, CDK2, E2F2, HDAC3, TP53, CDKN1A, CCND1, E2F4, ATM, RBL2, SMAD4, CDKN2A, MYC, NRG1, GSK3B, RBL1, HDAC6</td>
</tr>
<tr>
<td>T Cell Receptor Signaling</td>
<td>RAC1, ELK1, MAPK1, IKKβ, CBL, PI3KCA, FOS, NFκB2, PI3KCB, PI3K3, MAPK8, MAP3K3, KRA5</td>
</tr>
<tr>
<td>Cellular Function</td>
<td>Genes</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Death Receptor Signaling</td>
<td>CRADD; HSPH1; BID; BIRC4; TBK1; IKKB; FADD; FAS; NFKB2; BCL2; MAPK14; MAPK8; RIP1; CASP9; DAXX; TANK; TIF1B; RELA; TRAF2; TNF; IKKB; RELB; CASP9; CHUK; APE1; NFKB1; CASP2; BIRC2; CASP3; BIRC3</td>
</tr>
<tr>
<td>FGF Signaling</td>
<td>RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11; AKT2; PI3KCA; CREB1; PI3CB; PI3KC3; MAPK8; MAPK3; MAPK13; PTPN6; PI3KCA; MAPK14; RAF1; AKT1; PI3K1; STAT3; MAP2K1; FGFR4; CRKL; ATF4; AKT3; PRKCA; IGF</td>
</tr>
<tr>
<td>GM-CSF Signaling</td>
<td>LYN; ELK1; MAPK1; PTPN11; AKT2; PI3KCA; CAMK2A; STAT5B; PI3CB; PI3KC3; GNB2L1; BCL2L1; MAPK3; ETS1; K Ras; RUNX1; PIM1; PI3KCA; RAF1; MAP2K2; AKT1; AK2; PI3K1; STAT3; MAP2K1; CCND1; AKT3; STAT1</td>
</tr>
<tr>
<td>Amyotrophic Lateral Sclerosis Signaling</td>
<td>BID; IGF1; RAC1; BIRC4; P4G; CAPN5; CAPN2; PI3KCA; BCL2; PI3CB; PI3KC3; BCL2L1; CAPN1; PI3KCA; PI3CB; TP53; CASP9; PI3K1; RAB5A; CASP1; APE1; TGFp; BIRC2; BAAX; AKT3; CASP3; BIRC3</td>
</tr>
<tr>
<td>JAK/Stat Signaling</td>
<td>PTPN1; MAPK1; PTPN11; AKT2; PI3KCA; STAT5B; PI3CB; PI3KC3; K Ras; SOCS1; STAT3A; PTPN6; PI3KCA; RAF1; CDKN1A; MAP2K2; JAK1; AK1; AK2; PI3K1; STAT3; MAP2K1; FRA1; AKT3; STAT1</td>
</tr>
<tr>
<td>Nicotinate and Nicotinamide Metabolism</td>
<td>PRKCE; IRAK1; AKAA2; EIF2AK2; GRS; MAPK1; ELK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1; P53; MAPK5; CDK2; PIM1; DYRK1A; MAPK2; MAPK1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK</td>
</tr>
<tr>
<td>Chemokine Signaling</td>
<td>CXCR4; ROCK2; MAPK1; PTK2; FOS; C/EBP; GNAQ; CAMK2A; CXCL1; MAPK; PTPN6; MAPK3; Kras; MAPK13; RHOA; C3; SRC; PIP1C; MAPK14; NOX1; RAF1; MAPK2; MAPK1; JUN; CCL2; PRKCA</td>
</tr>
<tr>
<td>IL-2 Signaling</td>
<td>ELK1; MAPK1; PTPN11; AKT2; PI3KCA; SYK; FOS; STAT5B; PI3CB; PI3KC3; MAPK8; MAPK3; Kras; SOCS1; STAT5A; PI3KCA; LCK; RAF1; MAP2K2; JAK1; PI3K1; MAP2K1; JUN; AKT3</td>
</tr>
<tr>
<td>Synaptic Long Term Depression</td>
<td>PRKCE; IGF1; PRKCE; PRDX5; LYN; MAPK1; GNAS; PRKCI; GNAQ; PIP31A; IGF1R; PRKD1; DAB1; Kras; GNAS; PRKCD; NO53; NO52A; PIP2CA; YWHAA; RAF1; MAP2K2; PIP2P3C; MAP2K1; PRKCA</td>
</tr>
<tr>
<td>Estrogen Receptor Signaling</td>
<td>TAF4B; EP300; CARM1; PCAF; MAPK1; NCOA2; SMARCA4; MAPK3; NRP1; Kras; SRC; NCOA1; HDAC3; PPARGCA1A; RBM9; NCOA3; RAF1; CREB1; MAPK2; NCOA2; MAPK1; PRKDC; ESR1; ESR2</td>
</tr>
<tr>
<td>Protein Ubiquitination Pathway</td>
<td>TRAF6; SMURF1; BIRC4; BRC1; UCHL1; NEU1D4; CBL; UBE2L; BTRC; HSPA5; USP7; USP10; FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP9; USP2; VH1L1; HS90AA1; BIRC3</td>
</tr>
<tr>
<td>IL-10 Signaling</td>
<td>TRAF6; CCR1; ELK1; IKKIB; SP1; FOS; NFKB1; MAPK14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKKB; RELB; MAPK7; JAK1; CHUK; STAT3; NFKB1; JUN; IL1R1; IL6</td>
</tr>
<tr>
<td>VDR/RXR Activation</td>
<td>PRKCE; EP300; PRKCB; RXRA; GAD45A; HES1; NCO2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD; RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1; LRP5; CEIBP; FOXXO1; PRKCA</td>
</tr>
<tr>
<td>TGF-beta Signaling</td>
<td>E300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; Kras; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREB1; MAPK2; MAPK1; TGFBR1; SMAD4; JUN; SMAD5</td>
</tr>
<tr>
<td>Toll-like Receptor Signaling</td>
<td>IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1; IKKBI; FOS; NFKB2; MAPK14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKKGB; RELB; MAPK7; CHUK; NFKB1; TLR2; JUN</td>
</tr>
<tr>
<td>p38 MAPK Signaling</td>
<td>HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS; CREB1; DDT3; RPS8KA4; DAXX; MAPK13; TRAF2; MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1; SRF; STAT1</td>
</tr>
<tr>
<td>Neurotrophin/TRK Signaling</td>
<td>NTRK2; MAPK1; PTPN11; PI3KCA; CREB1; FOS; PI3CB; PI3KC3; MAPK8; MAPK3; Kras; PI3KC3A; RAF1; MAP2K2; AKT1; PI3K1; PDK1; MAP2K1; CDK4; JUN; ATF4</td>
</tr>
<tr>
<td>CELLULAR FUNCTION</td>
<td>GENES</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>FXR/RXR Activation</td>
<td>INS; PPPAR; FASN; RXRA; AKT2; SDC1; MAPK8; AP03B; MAPK10; PPARG; MTTP; MAPK5; PPARC1A; TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1</td>
</tr>
<tr>
<td>Synaptic Long Term Potentiation</td>
<td>PRKCE; RAPA1; EP300; PRKCKZ; MAPK1; CREB1;</td>
</tr>
<tr>
<td>Calcium Signaling</td>
<td>RAP1A; EP300; HADC4; MAPK1; HADC5; CREB1; CAMK2A; MYB9; MAPK3; HADC2; HADC7A; HADC11; HADC5; CREBBP; CALR; CAMKK2; ATF4;</td>
</tr>
<tr>
<td>EGF Signaling</td>
<td>ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; PIK3CA; AKT1; AK4; PIK3R1; STAT3; MAP2K1; JUN; PIK3CA; SRF; STAT1</td>
</tr>
<tr>
<td>Hypoxia Signaling in the Cardiovascular System</td>
<td>EDN1; PTEN; EP300; NQO1; UBE2L2; CREB1; AKT1;</td>
</tr>
<tr>
<td>Lipid/IL-1 Mediated Inhibition of RXX Function</td>
<td>HIF1A; SLC2A4; NOTCH1; TP53; LDLH; AKT1; ATM; VEGFA; JUN; ATF4; VHL; IKB</td>
</tr>
<tr>
<td>LXR/RXR Activation</td>
<td>IRAK1; MYD88; TAK1; PPARA; RXRA; ABCA1;</td>
</tr>
<tr>
<td>Amyloid Processing</td>
<td>PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;</td>
</tr>
<tr>
<td>IL-4 Signaling</td>
<td>AKT2; PIK3C3A; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;</td>
</tr>
<tr>
<td>Cell Cycle; G2/M DNA Damage Checkpoint</td>
<td>CHEK1; ATR; CHEK2; WWA; TP53; CDKN1A;</td>
</tr>
<tr>
<td>Regulation</td>
<td>PRKD1; ATM; SFI; CDKN2A</td>
</tr>
<tr>
<td>Nitric Oxide Signaling in the Cardiovascular System</td>
<td>RKB; FLT1; POF; AKT2; PIK3CA; PIK3CB; PIK3C3;</td>
</tr>
<tr>
<td>Purine Metabolism</td>
<td>NM21; SMARC4; MYBP; RRM2; ADAR; EIF2AK4;</td>
</tr>
<tr>
<td>AMPK-mediated cAMP Signaling</td>
<td>RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;</td>
</tr>
<tr>
<td>Mitochondrial Dysfunction</td>
<td>SRC2; RAI1; MAP2K2; STAT3; MAP2K1; BRF; ATF4</td>
</tr>
<tr>
<td>Notch Signaling</td>
<td>HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;</td>
</tr>
<tr>
<td>Endoplasmic Reticulum Stress Pathway</td>
<td>IKB; TIRE5; MAPK8; KBP; TRAF2;</td>
</tr>
<tr>
<td>Pyrimidine Metabolism</td>
<td>NME2; AICDA; RRM2; EIF2AK4;</td>
</tr>
<tr>
<td>Parkinson’s Signaling</td>
<td>UCL1; MAPK8; AKT1;</td>
</tr>
<tr>
<td>Cardiac &amp; Beta Adrenergic Signaling</td>
<td>GNAS; GNAQ; PPP2R1A; GB2L1; PPP2CA; PPP1CC;</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>HK2; GCK; G6P; ALDH1A1; PM2; LDIH;</td>
</tr>
<tr>
<td>Glucose Homeostasis</td>
<td>IKB1; SOCS1; JAK1; AK2; IT1M1; STAT1; IT1M3</td>
</tr>
<tr>
<td>Interferon Signaling</td>
<td>AKR2B; SMO; GLI2; DRYK1A; GLI1; GS3K; DRYK1B</td>
</tr>
<tr>
<td>Sonic Hedgehog Signaling</td>
<td>GPCR; GRP; YWHAZ; PHK1;</td>
</tr>
<tr>
<td>Glycerocephalosiphidolipid Metabolism</td>
<td>PLD1; GNP; GPM; YWHAZ; PHK1;</td>
</tr>
<tr>
<td>Phospholipid Degradation</td>
<td>PRDX6; PLD1; GRN; YWHAZ; PHK1;</td>
</tr>
<tr>
<td>Tryptophan Metabolism</td>
<td>SIA12; PRMT5; NEDD4; ALDH1A1; CYTP31; SIA1H</td>
</tr>
<tr>
<td>Lysine Degradation</td>
<td>SUV39H1; EHMT2; NSE1; SETD7; PPP2R5C</td>
</tr>
<tr>
<td>Nucleotide Excision Repair Pathway</td>
<td>ERCC5; ERCC4; XPA; XPC; ERCC1</td>
</tr>
<tr>
<td>Starch and Sucrose Metabolism</td>
<td>UCL1; HK2; GCK; G6P;</td>
</tr>
<tr>
<td>Aminoacyl-tRNA Metabolism</td>
<td>NQO1; HK2; GCK; HK1</td>
</tr>
<tr>
<td>Arachidonic Acid Metabolism</td>
<td>PRDX6; GRN; YWHAZ; CYTP31</td>
</tr>
<tr>
<td>Circadian Rhythm Signaling</td>
<td>CSNK1E; CREB1; ATF4; NR1D1</td>
</tr>
<tr>
<td>Coupling System</td>
<td>BKBR1; F2R; SERPIN1; F3</td>
</tr>
<tr>
<td>Dopamine Receptor Signaling</td>
<td>PPP2R1A; PPP2CA; PPP1CC; PPP2R5C</td>
</tr>
<tr>
<td>Glutathione Metabolism</td>
<td>IDH2; GSTP1; ANPEP; IDH1</td>
</tr>
<tr>
<td>Glycerolipid Metabolism</td>
<td>ALDH1A1; GPM; PHK1;</td>
</tr>
<tr>
<td>Metabolism</td>
<td>PLD1; GNP; YWHAZ; PHK1;</td>
</tr>
<tr>
<td>CELULAR FUNCTION</td>
<td>GENES</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Linoleic Acid Metabolism</td>
<td>PRDX6; GRN; YWHAZ; CYP1B1</td>
</tr>
<tr>
<td>Methionine Metabolism</td>
<td>DNMT1; DNMT3B; AIC5; DNMT3A</td>
</tr>
<tr>
<td>Pyruvate Metabolism</td>
<td>GLD1; ALDH1A1; PKM2; LDHA</td>
</tr>
<tr>
<td>Arginine and Proline Metabolism</td>
<td>ALDH1A1; NO83; NO8A</td>
</tr>
<tr>
<td>Eicosanoid Signaling</td>
<td>PRDX6; GRN; YWHAZ</td>
</tr>
<tr>
<td>Fructose and Mannose Metabolism</td>
<td>HK2; GCK; HK1</td>
</tr>
<tr>
<td>Galactose Metabolism</td>
<td>HK2; GCK; HK1</td>
</tr>
<tr>
<td>Sulfuric, Cysteine and Lignin Biosynthesis</td>
<td>PRDX6; PRDX1; TYR</td>
</tr>
<tr>
<td>Antigen Presentation Pathway</td>
<td>CALR; B2M</td>
</tr>
<tr>
<td>Biosynthesis of Steroids</td>
<td>NQ01; DHCR7</td>
</tr>
<tr>
<td>Butanoate Metabolism</td>
<td>ALDH1A1; NLGN1</td>
</tr>
<tr>
<td>Citrate Cycle</td>
<td>IDH1; IDH1</td>
</tr>
<tr>
<td>Fatty Acid Metabolism</td>
<td>ALDH1A1; CYP1B1</td>
</tr>
<tr>
<td>Glycerophospholipid Metabolism</td>
<td>PRDX6; CHKA</td>
</tr>
<tr>
<td>Histidine Metabolism</td>
<td>PRMT5; ALDH1A1</td>
</tr>
<tr>
<td>Iodothyronine Metabolism</td>
<td>ERO1L1; APEX1</td>
</tr>
<tr>
<td>Metabolism of Xenobiotics by Cytochrome p450</td>
<td>GSTP1; CYP1B1</td>
</tr>
<tr>
<td>Methane Metabolism</td>
<td>PRDX6; PRDX1</td>
</tr>
<tr>
<td>Phenylalanine Metabolism</td>
<td>PRDX6; PRDX1</td>
</tr>
<tr>
<td>Propanoate Metabolism</td>
<td>ALDH1A1; LDHA</td>
</tr>
<tr>
<td>Selenoamino Acid Metabolism</td>
<td>PRMT3; AIHCY</td>
</tr>
<tr>
<td>Sphingolipid Metabolism</td>
<td>SPHK1; SPHK2</td>
</tr>
<tr>
<td>Aminophosphonate Metabolism</td>
<td>PRMT5</td>
</tr>
<tr>
<td>Mitochondrial Metabolism</td>
<td>ALDH1A1</td>
</tr>
<tr>
<td>Androgen and Estrogen Metabolism</td>
<td>PRMT5</td>
</tr>
<tr>
<td>Ascorbate and Aldarate Metabolism</td>
<td>ALDH1A1</td>
</tr>
<tr>
<td>Bile Acid Biosynthesis</td>
<td>ALDH1A1</td>
</tr>
<tr>
<td>Cysteine Metabolism</td>
<td>LDHA</td>
</tr>
<tr>
<td>Fatty Acid Biosynthesis</td>
<td>EASN</td>
</tr>
<tr>
<td>Gliotransmitter Receptor Signaling</td>
<td>GNB2L1</td>
</tr>
<tr>
<td>NRF2-mediated Oxidative Stress Response</td>
<td>PRDX1</td>
</tr>
<tr>
<td>Pentose Phosphate Pathway</td>
<td>GPI</td>
</tr>
<tr>
<td>Pentose and Glucuronic Interconversions</td>
<td>UCHL1</td>
</tr>
<tr>
<td>Retinol Metabolism</td>
<td>ALDH1A1</td>
</tr>
<tr>
<td>Riboflavin Metabolism</td>
<td>TYR</td>
</tr>
<tr>
<td>Tyrosine Metabolism</td>
<td>PRMT5; TYR</td>
</tr>
<tr>
<td>Ubiquinone Biosynthesis</td>
<td>PRMT5</td>
</tr>
<tr>
<td>Valine, Leucine and Isoleucine Degradation</td>
<td>ALDH1A1</td>
</tr>
<tr>
<td>Glycine, Serine and Threonine Metabolism</td>
<td>CHKA</td>
</tr>
<tr>
<td>Lysine Degradation</td>
<td>ALDH1A1</td>
</tr>
</tbody>
</table>

Embodiments of the invention also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct. 13, 2011—Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (New insights into repeat instability: role of RNA-DNA hybrids. McVor E I, Polak U, Napierska M. RNA Biol. 2010 September-October; 7(5):551-8). The CRISPR-Cas system may be harnessed to correct these
defects of genomic instability. A further aspect of the invention relates to utilizing the CRISPR-Cas system for correcting defects in the EMPI2A and EMPI2B genes that have been identified to be associated with Lafora disease. Lafora disease is an autosomal recessive condition which is characterized by progressive myoclonus epilepsy which may start as epileptic seizures in adolescence. A few cases of the disease may be caused by mutations in genes yet to be identified. The disease causes seizures, muscle spasms, difficulty walking, dementia, and eventually death. There is currently no therapy that has proven effective against disease progression. Other genetic abnormalities associated with epilepsy may also be targeted by the CRISPR-Cas system and the underlying genetics is further described in Genetics of Epilepsy and Genetic Epilepsies, edited by Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology: 20; 2009).

In various conditions, the CRISPR-Cas system may be used to correct ocular defects that arise from several genetic mutations further described in Genetic Diseases of the Eye, Second Edition, edited by Elias I. Traboulsi, Oxford University Press, 2012.

Several further aspects of the invention relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers’ Disease, Alzheimer’s Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry’s Disease, Gerstmann-Straussler-Scheinker Disease, Huntington’s Disease and other Triplet Repeat Disorders, Leigh’s Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

In some embodiments, the condition may be neoplasia. In some embodiments, where the condition is neoplasia, the genes to be targeted are any of those listed in Table A (in this case PTEN and so forth). In some embodiments, the condition may be Age-related Macular Degeneration. In some embodiments, the condition may be a Schizophrenic Disorder. In some embodiments, the condition may be a Trinucleotide Repeat Disorder. In some embodiments, the condition may be Fragile X Syndrome. In some embodiments, the condition may be a Secretase Related Disorder. In some embodiments, the condition may be a Prion-related disorder. In some embodiments, the condition may be ALS. In some embodiments, the condition may be a drug addiction. In some embodiments, the condition may be Autism. In some embodiments, the condition may be Alzheimer’s Disease. In some embodiments, the condition may be inflammation. In some embodiments, the condition may be Parkinson’s Disease.

Examples of proteins associated with Parkinson’s disease include but are not limited to α-synuclein, DJ-1, LRRK2, PARK1, Parkin, UCHL1, Synphilin-1, and NURR1.

Examples of addiction-related proteins may include ABAT for example.

Examples of inflammation-related proteins may include the monocyte chemotactic protein-1 (MCP1) encoded by the Ccr2 gene, the C — C chemokine receptor type 5 (CCR5) encoded by the Ccr5 gene, the IgG receptor IIb (FCGR2b, also termed CD32) encoded by the Fcg2r2 gene, or the Fc epsilon R1g (FCER1g) protein encoded by the Fcgr1g gene, for example.

Examples of cardiovascular diseases associated proteins may include IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase), MB (myoglobin), IL4 (interleukin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), or CTSK (cathepsin K), for example.

Examples of Alzheimer’s disease-associated proteins may include the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, or the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBE3A gene, for example.

Examples of proteins associated Autism Spectrum Disorder may include the benzodiazepine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, or the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, for example.

Examples of proteins associated Macular Degeneration may include the ATP-binding cassette, sub-family A (ABCA1) member 4 protein (ABCA4) encoded by the ABCR gene, the apolipoprotein E protein (APOE) encoded by the APOE gene, or the chemokine (C-C motif) 11g and 2 protein (CCL2) encoded by the CCL2 gene, for example.

Examples of proteins associated Schizophrenia may include NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, and combinations thereof.

Examples of proteins involved in tumor suppression may include ATM (ataxia telangiectasia mutated), ATR (ataxia telangiectasia and Rad3 related), EGFR (epidermal growth factor receptor),erbB2 (erb-b2 erythroblastic leukemia viral oncogene homolog 2), erbB3 (erb-b2 erythroblastic leukemia viral oncogene homolog 3), erbB4 (erb-b2 erythroblastic leukemia viral oncogene homolog 4), Notch 1, Notch2, Notch 3, or Notch 4, for example.

Examples of proteins associated with a secretase disorder may include PSENEN (presenilin enhancer 2 homolog (C. elegans)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APOE1 (apoE7) cholesterol detective 1 homolog B (C. elegans), PSEN2 (presenilin 2 (Alzheimer disease 4)), or BACE1 (beta-site APP-cleaving enzyme 1), for example.

Examples of proteins associated with Amyotrophic Lateral Sclerosis may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGEB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

Examples of proteins associated with Parkinson’s disease may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

Examples of proteins related to neurodegenerative conditions in prion disorders may include A2M (Alpha-2-Macroglobulin), AATF (Apoptosis antagonizing transcription factor), ACP2 (Acid phosphatase prostate), ACTA2 (Actin alpha 2 smooth muscle aorta), ADAM22 (ADAM metallopeptidase
domain). ADORA3 (Adenosine A3 receptor), or ADRA1D (Alpha-1D adrenergic receptor for Alpha-1D adrenoceptor), for example.

Examples of proteins associated with Immunodeficiency may include AZM [alpha-2-macroglobulin]; A-NAT [arylalkylamine N-acetyltransferase]; ABCA1 [ATP-binding cassette, sub-family A (ABCA1), member 1]; ABCA2 [ATP-binding cassette, sub-family A (ABCA1), member 2]; or ABCA3 [ATP-binding cassette, sub-family A (ABCA1), member 3]; for example.

Examples of proteins associated with Trinucleotide Repeat Disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), or DMPK (dystrophia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2), for example.

Examples of proteins associated with Neurotransmission Disorders include SST (somatostatin), NOS1 (nitric oxide synthase 1 (neuronal)), ADRA2A (adrenergic, alpha-2A-, receptor), ADRA2C (adrenergic, alpha-2C-, receptor), TACR1 (tachykinin receptor 1), or HTR2c (5-hydroxytryptamine (serotonin) receptor 2C), for example.

Examples of neuregulational-associated sequences include A2BP1 [ataxin 2-binding protein 1]; AADAT [ammonium aminotransferase], AANAT [arylalkylamine N-acetyltransferase], ABAT [4-aminobutyrate aminotransferase], ABCA1 [ATP-binding cassette, sub-family A (ABCA1), member 1], or ABCA13 [ATP-binding cassette, sub-family A (ABCA1), member 1], for example.

Further examples of preferred conditions treatable with the present system include may be selected from: Aicardi-Goutieres Syndrome; Alexander Disease; Allan-Herndon-Dudley Syndrome; POLG-Related Disorders; Alpha-Mannosidosis (Type II and III); Alström Syndrome; Angelman Syndrome; Ataxin-Telangiectasia; Neuronal Ceroid-Lipofuscinosis; Beta-Thalassemia; Bilateral Optic Atrophy and (Infantile) Optic Atrophy Type 1; Retinoblastoma (bilateral); Canavan Disease; Cerebrocorticalisocortical Syndrome 1 [COFS1]; Cerebrotendinous Xanthomatosis; Cornelia de Lange Syndrome; MAPT-Related Disorders; Genetic Prion Diseases; Dravet Syndrome; Early-Onset Familial Alzheimer Disease; Friedreich Ataxia [FRDA]; Fryns Syndrome; Fucosidosis; Fukuyama Congenital Muscular Dystrophy; Galactosialidosis; Gaucher Disease; Organic Acidemias; Hemophagocytic Lymphohistiocytosis; Hutchinson-Gilford Progeria Syndrome; Mucolipidosis II; Infantile Free Sialic Acid Storage Disease; PL2G6-Associated Neuropathogenesis; Jervell and Lange-Nielsen Syndrome; Junctional Epidermolysis Bullosa; Huntington Disease; Krabbe Disease (Infantile); Mitochondrial DNA-Associated Leigh Syndrome and NARP; Lesch-Nyhan Syndrome; LISH-Associated Lissencephaly; Lowe Syndrome; Maple Syrup Urine Disease; MECP2 Duplication Syndrome; ATP7A-Related Copper Transport Disorders; LAMA2-Related Muscular Dystrophy; Arylsulfatase A Deficiency; Mucopolysaccharidosis Types I, II or III; Peroxisome Biogenesis Disorders; Zellweger Syndrome Spectrum; Neurodegeneration with Brain Iron Accumulation Disorders; Acid Sphingomyelinase Deficiency; Niemann-Pick Disease Type C; Glycine Encephalopathy; ARX-Related Disorders; Urea Cycle Disorders; COL1A1/2-Related Osteogenesis Imperfecta; Mitochondrial DNA Deletion Syndromes; PIPL-Related Disorders; Perry Syndrome; Phelan-McDermid Syndrome; Glycogen Storage Disease Type II (Pompe Disease) (Infantile); MAPT-Related Disorders; MESP2-Related Disorders; Rhizomelic Chondrodysplasia Punctata Type 1; Roberts Syndrome; Sandhoff Disease; Schindler Disease—Type 1; Adenosine Deaminase Deficiency; Smith-Lemli-Optiz Syndrome; Spinal Muscular Atrophy; Infantile-Onset Spino cerebellar Ataxia; Hexosaminidase A Deficiency; Thanatophoric Dysplasia Type 1; Collagen Type VI-Related Disorders; Usher Syndrome Type I; Congenital Muscular Dystrophy; Wolf-Hirschhorn Syndrome; Lysosomal Acid Lipase Deficiency; and Xeroderma Pigmentosum.

As will be apparent, it is envisaged that the present system can be used to target any polymolecule sequence of interest. Some examples of conditions or diseases that might be usefully treated using the present system are included in the Tables above and examples of genes currently associated with those conditions are also provided there. However, the genes exemplified are not exhaustive.

EXAMPLES

The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therefor and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

Example I

CRISPR Complex Activity in the Nucleus of a Eukaryotic Cell

An example type II CRISPR system is the type II CRISPR locus from Streptococcus pyogenes SF370, which contains a cluster of four genes Cas9, Cas1, Cas2, and Cas1, as well as two non-coding RNA elements, tracrRNA and a characteristic array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences (spacers, about 30 bp each). In this system, targeted DNA double-strand break (DSB) is generated in four sequential steps (FIG. 2A). First, two non-coding RNAs, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the direct repeats of pre-crRNA, which is then processed into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the DNA target consisting of the protospacer and the corresponding PAM via heteroduplex formation between the spacer region of the crRNA and the protospacer DNA. Finally, Cas9 mediates cleavage of target DNA upstream of PAM to create a DSB within the protospacer (FIG. 2A). This example describes an example process for adapting this RNA-programmable nucleic acid system to direct CRISPR complex activity in the nuclei of eukaryotic cells.

Cell Culture and Transfection

Human embryonic kidney (HEK) cell line HEK 293FT (Life Technologies) was maintained in Dulbecco’s modified Eagle’s Medium (DMEM) supplemented with 10% fetal bovine serum (HyClone), 2 mM GlutaMAX (Life Technologies), 100 μM l-penicillin, and 100 μM streptomycin at 37°C with 5% CO2 incubation. Mouse neuro2A (N2A) cell line (ATCC) was maintained with DMEM supplemented with 5% fetal bovine serum (HyClone), 2 mM GlutaMAX (Life Technologies), 100 μM l-penicillin, and 100 μM streptomycin at 37°C with 5% CO2.

HEK 293FT or N2A cells were seeded into 24-well plates (Corning) one day prior to transfection at a density of 200,000 cells per well. Cells were transfected using Lipofectamine™
2000 (Life Technologies) following the manufacturer’s recommended protocol. For each well of a 24-well plate a total of 800 ng of plasmids were used.

Surveyor Assay and Sequencing Analysis for Genome Modification

HEK 293FT or N2A cells were transfected with plasmid DNA as described above. After transfection, the cells were incubated at 37°C for 72 hours before genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA extraction kit (Epicentre) following the manufacturer’s protocol. Briefly, cells were resuspended in QuickExtract solution and incubated at 65°C for 15 minutes and 98°C for 10 minutes. Extracted genomic DNA was immediately processed or stored at −20°C.

The genomic region surrounding a CRISPR target site for each gene was PCR amplified, and products were purified using QiaQuick Spin Column (Qiagen) following manufacturer’s protocol. A total of 400 ng of the purified PCR products were mixed with 2 μl, 10x Taq polymerase PCR buffer (Enzymatics) and ultrapure water to a final volume of 20 μl, and subjected to a re-annealing process to enable heteroduplex formation: 95°C for 10 min, 95°C to 85°C ramping at −2°C/s, 85°C to 25°C at −0.25°C/C/s, and 25°C hold for 1 minute. After re-annealing, products were treated with Surveyor nuclease and Surveyor enhancer S (Transgenomics) following the manufacturer’s recommended protocol, and analyzed on 4-20% Novex TBE poly-acrylamide gels (Life Technologies). Gels were stained with SYBR Gold DNA stain (Life Technologies) for 50 minutes and imaged with a Gel Doc gel imaging system (Bio-rad). Quantification was based on relative band intensities, as a measure of the fraction of cleaved DNA. Fig. 8 provides a schematic illustration of this Surveyor assay.

Restriction fragment length polymorphism assay for detection of homologous recombination

HEK 293FT and N2A cells were transfected with plasmid DNA, and incubated at 37°C for 72 hours before genomic DNA extraction as described above. The target genomic region was PCR amplified using primers outside the homology arms of the homologous recombination (HR) template. PCR products were separated on a 1% agarose gel and extracted with MinElute Gel Extraction Kit (Qiagen). Purified products were digested with HindIII (Fermentas) and analyzed on a 6% Novex TBE poly-acrylamide gel (Life Technologies).

RNA secondary structure prediction and analysis

RNA secondary structure prediction was performed using the online webserver RNAfold developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g. A. R. Gruber et al., 2008, Cell 106(1): 23-24; and P. A. Carr and G. M. Church, 2009, Nature Biotechnology 27(12): 1151-62).

Bacterial Plasmid Transformation Interference Assay

Elements of the S. pyogenes CRISPR locus 1 sufficient for CRISPR activity were reconstituted in E. coli using pCRISPR plasmid (schematically illustrated in Fig. 10A). pCRISPR contained tracrRNA, SpCas9, and a leader sequence driving the crRNA array. Spacers (also referred to as “guide sequences”) were inserted into the crRNA array between Bsal sites using annealed oligonucleotides, as illustrated. Challenge plasmids used in the interference assay were constructed by inserting the protospacer (also referred to as a “target sequence”) sequence along with an adjacent CRISPR motif sequence (PAM) into pUC19 (see Fig. 10B). The challenge plasmid contained ampicillin resistance. Fig. 10C provides a schematic representation of the interference assay. Chemically competent E. coli strains already carrying pCRISPR and the appropriate spacer were transformed with the challenge plasmid containing the corresponding protospacer-PAM sequence. pUC19 was used to assess the transformation efficiency of each pCRISPR-carrying competent strain. CRISPR activity resulted in cleavage of the plasmid carrying the protospacer, precluding ampicillin resistance otherwise conferred by pUC19 lacking the protospacer. Fig. 10D illustrates competence of each pCRISPR-carrying E. coli strain used in assays illustrated in Fig. 4C.

RNA Purification

HEK 293FT cells were maintained and transfected as stated above. Cells were harvested by trypan blue followed by washing in phosphate buffered saline (PBS). Total cell RNA was extracted with TRI reagent (Sigma) following manufacturer’s protocol. Extracted total RNA was quantified using NanoDrop (Thermo Scientific) and normalized to the same concentration.

Northern Blot Analysis of crRNA and tracrRNA Expression in Mammalian Cells

RNA was prepared with equal volumes of 2x loading buffer (Ambion), heated to 95°C for 5 min, chilled on ice for 5 min, and then loaded onto 8% denaturing polyacrylamide gels (Sequagel, National Diagnostics) after pre-running the gel for at least 30 minutes. The samples were electrophoresed for 1.5 hours at 40 W/lane. Afterwards, the RNA was transferred to Hybond N+ membrane (GE Healthcare) at 300 mA in a semi-dry transfer apparatus (Bio-rad) at room temperature for 1.5 hours. The RNA was crosslinked to the membrane using autocrosslink button on Stratagene UV Crosslinker the Stratalinker (Stratagene). The membrane was pre-hybridized in ULTRAhyb-Oligo Hybridization Buffer (Ambion) for 30 min with rotation at 42°C, and probes were then added and hybridized overnight. Probes were ordered from IDT and labeled with [γ-32P] ATP (Perkin Elmer) with T4 polynucleotide kinase (New England Biolabs). The membrane was washed once pre-warmed (42°C) 2xSSC, 0.5% SDS for 1 min followed by two 30 minute washes at 42°C. The membrane was exposed to a phosphor screen for one hour or overnight at room temperature and then scanned with a phosphorimager (Typhoon).

Bacterial CRISPR System Construction and Evaluation

CRISPR locus elements, including tracrRNA, Cas9, and leader were PCR amplified from Streptococcus pyogenes SF370 genomic DNA with flanking homology arms for Gibson Assembly. Two Bsal type IIIS sites were introduced in between two direct repeats to facilitate easy insertion of spacers (Fig. 9). PCR products were cloned into EcoR1-digested pACYC184 downstream of the tet promoter using Gibson Assembly Master Mix (NEB). Other endogenous CRISPR system elements were omitted, with the exception of the last 50 bp of Cas2. Oligos (Integrated DNA Technology) encoding spacers with complimentary overhangs were cloned into the Bsal-digested vector pDX000 (NEB) and then ligated with T7 ligase (Enzymatics) to generate pCRISPR plasmids. Challenge plasmids containing spacers with PAM sequences (also referred to herein as “CRISPR motif sequences”) were created by ligating hybridized oligos carrying compatible overhangs (Integrated DNA Technology) into BamHI-digested pUC19. Cloning for all constructs was performed in E. coli strain JM109 (Zymo Research).

pCRISPR-carrying cells were made competent using the Z-Competent E. coli Transformation Kit and Buffer Set (Zymo Research, T3001) according to the manufacturer’s instructions. In the transformation assay, 50 ul aliquots of competent cells carrying pCRISPR were thawed on ice and transformed with 1 μg of spacer plasmid or pUC19 on ice for 50 minutes, followed by 45 second heat shock at 42°C and 2
minutes on ice. Subsequently, 250 µl SOC (Invitrogen) was added followed by shaking incubation at 37°C for 1 hr, and 100 µl of the post-SOC outgrowth was plated onto double selection plates (12.5 µg/ml chloramphenicol, 100 µg/ml ampicillin). To obtain cln-2g of DNA, total colony numbers were multiplied by 3.

To improve expression of CRISPR components in mammalian cells, two genes from the SF370 locus 1 of *Streptococcus pyogenes* (SpCas9) and RNAse II (SpRNAse III). To facilitate nuclear localization, a nuclear localization signal (NLS) was included at the amino (N)- or carboxyl (C)-termini of both SpCas9 and SpRNAse III (FIG. 2B). To facilitate visualization of protein expression, a fluorescent protein marker was also included at the N- or C-termini of both proteins (FIG. 2B). A version of SpCas9 with an NLS attached to both N- and C-termini (2xNLS-SpCas9) was also generated. Constructs containing NLS-fused SpCas9 and SpRNAse III were transfected into 293FT human embryonic kidney (HEK) cells, and the relative positioning of the NLS to SpCas9 and SpRNAse III was found to affect their nuclear localization efficiency. Whereas the C-terminal NLS was sufficient to target SpRNAse III to the nucleus, attachment of a single copy of these particular NLS's to either the N- or C-terminus of SpCas9 was unable to achieve adequate nuclear localization in this system. In this example, the C-terminal NLS was that of nuclear lamina (KPRAAFKKGGQKKKK (SEQ ID NO: 2)), and the C-terminal NLS was that of SV growth factor (PKKKKRKK (SEQ ID NO: 1)). Of the versions of SpCas9 tested, only 2xNLS- SpCas9 exhibited nuclear localization (FIG. 2B).

The tracrRNA from the CRISPR locus of *S. pyogenes* SF370 has two transcriptional start sites, giving rise to two transcripts of 89-nucleotides (nt) and 171 nt that are subsequently processed into identical 75 nt mature tracrRNAs. The shorter 89 nt tracrRNA was selected for expression in mammalian cells (expression constructs illustrated in FIG. 7A, with functionality as determined by results of the Surveyor assay shown in FIG. 7B). Transcription start sites are marked as +1, and transcription terminator and the sequence probed by northern blot are also indicated. Expression of processed tracrRNA was also confirmed by Northern blot. FIG. 7C shows results of a Northern blot analysis of total RNA extracted from 293FT cells transfected with U6 expression constructs carrying long or short tracrRNA, as well as SpCas9 and DR-EMX1(1)-DR. Left and right panels are from 293FT cells transfected without or with SpRNAse III, respectively. U6 indicate loading control blotted with a probe targeting human U6 snRNA. Transfection of the short tracrRNA expression construct led to abundant levels of the processed form of tracrRNA (~75 bp). Very low amounts of long tracrRNA are detected on the Northern blot.

To promote precise transcription initiation, the RNA polymerase III-based U6 promoter was selected to drive the expression of tracrRNA (FIG. 2C). Similarly, a U6 promoter-based construct was developed to express a pre-crRNA array consisting of a single spacer flanked by two direct repeats (DRs, also encompassed by the term “tracr-mate sequences”; FIG. 2C). The initial spacer was designed to target a 33-base pair (bp) target site (30-bp protospacer plus a 3-bp CRISPR motif (PAM)) sequence satisfying the NGG recognition motif of Cas9) in the human EMX1 locus (FIG. 2C), a key gene in the development of the cerebral cortex.

To test whether heterologous expression of the CRISPR system (SpCas9, SpRNAse III, tracrRNA, and pre-crRNA) in mammalian cells can achieve targeted cleavage of mammalian chromosomes, HEK 293FT cells were transfected with combinations of CRISPR components. Since DSBs in mammalian nuclei are partially repaired by the non-homologous end joining (NHEJ) pathway, which leads to the formation of indels, the Surveyor assay was used to detect potential cleavage at the target EMX1 locus (FIG. 8) (see e.g. Guschin et al., 2010, Methods Mol Biol 649: 247). Co-transfection of all four CRISPR components was able to induce up to 5.0% cleavage in the protospacer (see FIG. 2D). Co-transfection of all CRISPR components minus SpRNAse III also induced up to 4.7% indel in the protospacer, suggesting that there may be endogenous mammalian RNases that are capable of assisting with crRNA maturation, such as for example the related Dicer and Drosha enzymes. Removing any of the remaining three components abolished the genome cleavage activity of the CRISPR system (FIG. 2D). Sanger sequencing of amplicons containing the target locus verified the cleavage activity: in 43 sequenced clones, 5 mutated alleles (11.6%) were found. Similar experiments using a variety of guide sequences produced indel percentages as high as 29% (see FIGS. 4-9, 12, and 13). These results define a three-component system for efficient CRISPR-mediated genome modification in mammalian cells. To optimize the cleavage efficiency, Appelans also tested whether different isoforms of tracrRNA affected the cleavage efficiency and found that, in this example system, only the short (89-bp) transcript form was able to mediate cleavage of the human EMX1 genomic locus (FIG. 7B).

FIG. 14 provides an additional Northern blot analysis of crRNA processing in mammalian cells. FIG. 14A illustrates a schematic showing the expression vector for a single spacer flanked by two direct repeats (DR-EMX1(1)-DR). The 30 bp spacer targeting the human EMX1 locus protospacer 1 (see FIG. 6) and the direct repeat sequences are shown in the sequence beneath FIG. 14A. The line indicates the region whose reverse-complement sequence was used to generate Northern blot probes for EMX1(1) crRNA detection. FIG. 14B shows a Northern blot analysis of total RNA extracted from 293FT cells transfected with U6 expression constructs carrying DR-EMX1(1)-DR. Left and right panels are from 293FT cells transfected without or with SpRNAse III respectively. DR-EMX1(1)-DR was processed into mature crRNAs only in the presence of SpCas9 and short tracrRNA and was not dependent on the presence of SpRNAse III. The mature crRNA detected from transfected 293FT total RNA is ~33 bp and is shorter than the 39-42 bp mature crRNA from *S. pyogenes*. These results demonstrate that a CRISPR system can be transplanted into eukaryotic cells and reprogrammed to facilitate cleavage of endogenous mammalian target polynucleotides.

FIG. 2 illustrates the bacterial CRISPR system described in this example. FIG. 2A illustrates a schematic showing the CRISPR locus 1 from *Streptococcus pyogenes* SF370 and a proposed mechanism of CRISPR-mediated DNA cleavage by this system. Mature crRNA processed from the direct repeat-spacer array directs Cas9 to genomic targets consisting of complimentary protospacers and a protospacer-adjacent motif (PAM). Upon target-spacer base pairing, Cas9 mediates a double-strand break in the target DNA. FIG. 2B illustrates engineering of *S. pyogenes* Cas9 (SpCas9) and RNAse II (SpRNAse III) with nuclear localization signals (NLSs) to enable import into the mammalian nucleus. FIG. 2C illustrates mammalian expression of SpCas9 and SpRNAse III driven by the constitutive EF1a promoter and tracrRNA and pre-crRNA array (DR-Spacer-DR) driven by the RNA Pol III promoter U6 to promote precise transcription initiation and termination. A protospacer from the human EMX1 locus with a satisfactory PAM sequence is used as the spacer in the pre-crRNA array. FIG. 2D illustrates surveyor nucleic assay
for SpCas9-mediated minor insertions and deletions. SpCas9 was expressed with and without SprKase III, tracrRNA, and a pre-cRNA array carrying the EMX1-targeter spacer. FIG. 2E illustrates a schematic representation of base pairing between target locus and EMX1-targeting cRNA; as well as an example chromatogram showing a micro deletion adjacent to the SpCas9 cleavage site. FIG. 2F illustrates mutated alleles identified from sequencing analysis of 43 clonal amplicons showing a variety of micro insertions and deletions. Dashes indicate deleted bases, and non-aligned or mismatched bases indicate insertions or mutations. Scale bar = 10 μm.

To further simplify the three-component system, a chimeric crRNA-tracrRNA hybrid design was adapted, where a mature crRNA (comprising a guide sequence) is fused to a partial tracrRNA via a stem-loop to mimic the natural crRNA: tracrRNA duplex (FIG. 3A). To increase co-delivery efficiency, a bicistronic expression vector was created to drive co-expression of a chimeric RNA and SpCas9 in transduced cells (FIGS. 3A and 8). In parallel, the bicistronic vectors were used to express a pre-cRNA (DR-guide sequence-DR) with SpCas9, to induce processing into crRNA with a separately expressed tracrRNA (compare FIG. 1B top and bottom). FIG. 9 provides schematic illustrations of bicistronic expression vectors for pre-cRNA array (FIG. 9A) or chimeric crRNA (represented by the short line downstream of the guide sequence insertion site and upstream of the EF1α promoter in FIG. 9B) with hSpCas9, showing location of various elements and the point of guide sequence insertion. The expanded sequence around the location of the guide sequence insertion site in FIG. 9B also shows a partial DR sequence (GTGTTAGCTA (SEQ ID NO: 27)) and a partial tracrRNA sequence (TACGAAAGTAAAATAAG-GCTAAGCTCTTGTTTT (SEQ ID NO: 28)). Guide sequences can be inserted between 383 sites using annealed oligonucleotides. Sequence design for the oligonucleotides are shown below the schematic illustrations in FIG. 9, with appropriate ligation adapters indicated. WPRE represents the Woodchuck hepatitis virus post-transcriptional regulatory element. The efficiency of chimeric RNA-mediated cleavage was tested by targeting the same EMX1 locus described above. Using both Surveyor assay and Sanger sequencing of amplicons, Applicants confirmed that the chimeric RNA design facilitates cleavage of human EMX1 locus with approximately a 4.7% modification rate (FIG. 4).

Generalizability of CRISPR-mediated cleavage in eukaryotic cells was tested by targeting additional genomic loci in both human and mouse cells by designing chimeric RNA targeting multiple sites in the human EMX1 and PVALB, as well as the mouse Th loci. FIG. 15 illustrates the selection of some additional targeted protospcers in human PVALB (FIG. 15A) and mouse Th (FIG. 15B) loci. Schematics of the gene loci and the location of three protospcers within the last exon of each are provided. The underlined sequences include 30 bp of protospcer sequence and 3 bp at the 3′ end corresponding to the PAM sequences. Protospcers on the sense and anti-sense strands are indicated above and below the DNA sequences, respectively. A modification rate of 6.3% and 0.75% was achieved for the human PVALB and mouse Th loci respectively, demonstrating the broad applicability of the CRISPR system in modifying different loci across multiple organisms (FIGS. 3B and 6). While cleavage was only detected with one out of three spacers for each locus using the chimeric constructs, all target sequences were cleaved with efficiency of indel production reaching 27% when using the co-expressed pre-cRNA arrangement (FIG. 6).

FIG. 13 provides a further illustration that SpCas9 can be reprogrammed to target multiple genomic loci in mammalian cells. FIG. 13A provides a schematic of the human EMX1 locus showing the location of five protospcers, indicated by the underlined sequences. FIG. 13B provides a schematic of the pre-cRNA/tracrRNA complex showing hybridization between the direct repeat region of the pre-cRNA and tracrRNA (top), and a schematic of a chimeric RNA design comprising a 20 bp guide sequence, and tracr mate and tracr sequences consisting of partial direct repeat and tracrRNA sequences hybridized in a hairpin structure (bottom). Results of a Surveyor assay comparing the efficacy of Cas9-mediated cleavage at five protospcers in the human EMX1 locus is illustrated in FIG. 13C. Each protospcer is targeted using either processed pre-cRNA/tracrRNA complex (crRNA) or chimeric RNA (ehiRNA).

Since the secondary structure of RNA can be crucial for intermolecular interactions, a structure prediction algorithm based on minimum free energy and Boltzmann-weighted structure ensemble was used to compare the putative secondary structure of all guide sequences used in our genome targeting experiment (FIG. 3B) (see e.g. Gruber et al., 2008, Nucleic Acids Resarch, 36, W70). Analysis revealed that in most cases, the effective guide sequences in the chimeric crRNA context were substantially free of secondary structure motifs, whereas the ineffective guide sequences were more likely to form internal secondary structures that could prevent base pairing with the target protospcer DNA. It is thus possible that variability in the spacer secondary structure might impact the efficiency of CRISPR-mediated interference when using a chimeric crRNA.

FIG. 3 illustrates example expression vectors. FIG. 3A provides a schematic of a bi-cistronic vector for driving the expression of a synthetic crRNA-tracrRNA chimera (chimeric RNA) as well as SpCas9. The chimeric guide RNA contains a 20-bp guide sequence corresponding to the protospcer in the genomic target site. FIG. 3B provides a schematic showing guide sequences targeting the human EMX1, PVALB, and mouse Th loci, as well as their predicted secondary structures. The modification efficiency at each target site is indicated below the RNA secondary structure drawing (EMX1, n=216 ampiclon sequencing reads; PVALB, n=224 reads; Th, n=265 reads). The folding algorithm produced an output with each base colored according to its probability of assuming the predicted secondary structure, as indicated by a rainbow scale that is reproduced in FIG. 3B in gray scale. Further vector designs for SpCas9 are shown in FIG. 44, which illustrates single expression vectors incorporating a U6 promoter linked to an insertion site for a guide oligo, and a C6h promoter linked to SpCas9 coding sequence. The vector shown in FIG. 44b includes a tracrRNA coding sequence linked to an H1 promoter.

To test whether spacers containing secondary structures are able to function in prokaryotic cells where CRISPRs naturally operate, transformation interference of protospcer-bearing plasmids were tested in an E. coli strain heterologously expressing the S. pyogenes SF370 CRISPR locus 1 (FIG. 10). The CRISPR locus was cloned into a low-copy E. coli expression vector and the crRNA array was replaced with a single spacer flanked by a pair of DRs (pCRISPR). E. coli strains harboring different pCRISPR plasmids were transformed with challenge plasmids containing the corresponding protospcer and PAM sequences (FIG. 10C). In the bacterial assay, all spacers facilitated efficient CRISPR interference (FIG. 4C). These results suggest that there may be additional factors affecting the efficiency of CRISPR activity in mammalian cells.

To investigate the specificity of CRISPR-mediated cleavage, the effect of single-nucleotide mutations in the guide
sequence on protospacer cleavage in the mammalian genome was analyzed using a series of EMX1-targeting chimeric crRNAs with single point mutations (FIG. 4A). FIG. 4B illustrates results of a Surveyor nuclease assay comparing the cleavage efficiency of Cas9 when paired with different mutant chimeric RNAs. Single-base mismatch up to 12-bp 5' of the PAM substantially abrogated genomic cleavage by SpCas9, whereas spacers with mutations at farther upstream positions retained activity against the original protospacer target (FIG. 4B). In addition to the PAM, SpCas9 has single-base specificity within the last 12-bp of the spacer. Furthermore, CRISPR is able to mediate genomic cleavage as efficiently as a pair of TALE nucleases (TALEN) targeting the same EMX1 protospacer. FIG. 4C provides a schematic showing the design of TALENs targeting EMX1, and FIG. 4D shows a Surveyor gel comparing the efficiency of TALEN and Cas9 (n=3).

Having established a set of components for achieving CRISPR-mediated gene editing in mammalian cells through the error-prone NHEJ mechanism, the ability of CRISPR to stimulate homologous recombination (HR), a high-fidelity gene repair pathway for making precise edits in the genome, was tested. The wild type SpCas9 is able to mediate site-specific DSBs, which can be repaired through both NHEJ and HR. In addition, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of SpCas9 was engineered to convert the nuclease into a nickase (SpCas9n; illustrated in FIG. 5A) (see e.g. Supranaukas et al., 2011, Nucleic Acids Research, 39: 9275; Gasiunas et al., 2012, Proc. Natl. Acad. Sci. USA, 109:E2579), such that nicked genomic DNA undergoes the high-fidelity homology-directed repair (HDR). Surveyor assay confirmed that SpCas9n does not generate indels at the EMX1 protospacer target. As illustrated in FIG. 5B, co-expression of EMX1-targeting chimeric crRNA with SpCas9 produced indels in the target site, whereas co-expression with SpCas9n did not (n=3). Moreover, sequencing of 327 amplicons did not detect any indels induced by SpCas9n.

The same locus was selected to test CRISPR-mediated HR by co-transfecting HEK 293FT cells with the chimeric RNA targeting EMX1, hSpCas9o or hSpCas9n, as well as a HR template to introduce a pair of restriction sites (HindIII and Nhel) near the protospacer. FIG. 5C provides a schematic illustration of the HR strategy, with relative locations of recombination points and primer annealing sequences (arrows). SpCas9 and SpCas9n indeed catalyzed integration of the HR template into the EMX1 locus. PCR amplification of the target region followed by restriction digest with HindIII revealed cleavage products corresponding to expected fragment sizes (arrows in restriction fragment length polymorphism gel analysis shown in FIG. 5D), with SpCas9 and SpCas9n mediating similar levels of HR efficiencies. Applicants further verified HR using Sanger sequencing of genomic amplicons (FIG. 5E). These results demonstrate the utility of CRISPR for facilitating targeted gene insertion in the mammalian genome. Given the 14-bp (12-bp from the spacer and 2-bp from the PAM) target specificity of the wild type SpCas9, the availability of a nickase can significantly reduce the likelihood of off-target modifications, since single strand breaks are not substrates for the error-prone NHEJ pathway.

Expression constructs mimicking the natural architecture of CRISPR loci with arrayed spacers (FIG. 2A) were constructed to test the possibility of multiplexed sequence targeting. Using a single CRISPR array encoding a pair of EMX1- and PVALB-targeting spacers, efficient cleavage at both loci was detected (FIG. 4F: showing both a schematic design of the crRNA array and a Surveyor blot showing efficient mediation of cleavage). Targeted deletion of larger genomic regions through concurrent DSBs using spacers against two targets within EMX1 spaced by 119 bp was also tested, and a 1.6% deletion efficiency (3 out of 182 amplicons: FIG. 4G) was detected. This demonstrates that the CRISPR system can mediate multiplexed editing within a single genome.

**Example 2**

**CRISPR System Modifications and Alternatives**

The ability to use RNA to program sequence-specific DNA cleavage defines a new class of genome engineering tools for a variety of research and industrial applications. Several aspects of the CRISPR system can be further improved to increase the efficiency and versatility of CRISPR targeting. Optimal Cas9 activity may depend on the availability of free Mg2+ at levels higher than that present in the mammalian nucleus (see e.g. Jinek et al., 2012, Science, 337:816), and the preference for an NGG motif immediately downstream of the protospacer restricts the ability to target on average every 12-bp in the human genome (FIG. 11, evaluating both plus and minus strands of human chromosomal sequences). Some of these constraints can be overcome by exploring the diversity of CRISPR loci across the microbial metagenome (see e.g. Makarova et al., 2011, Nat Rev Microbiol, 9:467). Other CRISPR loci may be transplanted into the mammalian cellular milieu by a process similar to that described in Example 1. For example, FIG. 12 illustrates adaptation of the Type II CRISPR system from *Streptococcus thermophilus* LMD-9 for heterologous expression in mammalian cells to achieve CRISPR-mediated genome editing. FIG. 12A provides a Schematic illustration of CRISPR1 from *S. thermophilus* LMD-9. FIG. 12B illustrates the design of an expression system for the *S. thermophilus* CRISPR system. Human codon-optimized hStCas9 is expressed using a constitutive EF1a promoter. Mature versions of tracrRNA and crRNA are expressed using the U6 promoter to promote precise transcription initiation. Sequences from the mature crRNA and tracrRNA are illustrated. A single base indicated by the lower case "a" in the crRNA sequence is used to remove the polyU sequence, which serves as a RNA polIII transcriptional terminator. FIG. 12C provides a schematic showing guide sequences targeting the human EMX1 locus as well as their predicted secondary structures. The modification efficiency at each target site is indicated below the RNA secondary structures. The algorithm generating the structures colors each base according to its probability of assuming the predicted secondary structure, which is indicated by a rainbow scale reproduced in FIG. 12C in gray scale. FIG. 12D shows the results of hStCas9-mediated cleavage in the target locus using the Surveyor assay. RNA guide spacers 1 and 2 induced 14% and 6.4%, respectively. Statistical analysis of cleavage activity across biological replicate at these two protospacer sites is also provided in FIG. 6. FIG. 16 provides a schematic of additional protospacer and corresponding PM sequence targets of the *S. thermophilus* CRISPR system in the human EMX1 locus. Two protospacer sequences are highlighted and their corresponding PM sequences satisfying NNAGAAW motif are indicated by underlining 3' with respect to the corresponding highlighted sequence. Both protospacers target the anti-sense strand.

**Example 3**

**Sample Target Sequence Selection Algorithm**

A software program is designed to identify candidate CRISPR target sequences on both strands of an input DNA
sequence based on desired guide sequence length and a CRISPR motif sequence (PAM) for a specified CRISPR enzyme. For example, target sites for Cas9 from S. pyogenes, with PAM sequences NGG, may be identified by searching for 5'-NGG-3' both on the input sequence and on the reverse-complement of the input. Likewise, target sites for Cas9 of S. thermophilus CRISPR1, with PAM sequence NNNAGAAW, may be identified by searching for 5'-NN-GNNAGAAW-3' (SEQ ID NO: 29) both on the input sequence and on the reverse-complement of the input. Likewise, target sites for Cas9 of S. thermophilus CRISPR3, with PAM sequence NNNNGNG, may be identified by searching for 5'-NN-GNNNG-3' both on the input sequence and on the reverse-complement of the input. The value "X" in Nn may be fixed by the program or specified by the user, such as 20.

Since multiple occurrences in the genome of the DNA target site may lead to nonspecific genome editing, after identifying all potential sites, the program filters out sequences based on the number of times they appear in the relevant reference genome. For those CRISPR enzymes for which sequence specificity is determined by a "seed" sequence, such as the 11-12 bp 5' from the PAM sequence, including the PAM sequence itself, the filtering step may be based on the seed sequence. Thus, to avoid editing at additional genomic loci, results are filtered based on the number of occurrences of the seed:PAM sequence in the relevant genome. The user may be allowed to choose the length of the seed sequence. The user may also be allowed to specify the number of occurrences of the seed:PAM sequence in a genome for purposes of passing the filter. The default is to screen for unique sequences. Filtration level is altered by changing both the length of the seed sequence and the number of occurrences of the sequence in the genome. The program may in addition or alternatively provide the sequence of a guide sequence complementary to the reported target sequence(s) by providing the reverse complement of the identified target sequence(s).

Further details of methods and algorithms to optimize sequence selection can be found in U.S. application Ser. No. 61/836,080; incorporated herein by reference.

Example 4

Evaluation of Multiple Chimeric crRNA-tracrRNA Hybrids

This example describes results obtained for chimeric RNAs (chiRNAs; comprising a guide sequence, a tracr mate sequence, and a tracr sequence in a single transcript) having tracr sequences that incorporate different lengths of wild-type tracrRNA sequence. FIG. 18a illustrates a schematic of a bicistronic expression vector for chimeric RNA and Cas9. Cas9 is driven by the CBh promoter and the chimeric RNA is driven by a U6 promoter. The chimeric guide RNA consists of a 20 bp guide sequence (Ns) joined to the tracr sequence (running from the first "U" of the lower strand to the end of the transcript), which is truncated at various positions as indicated. The guide and tracr sequences are separated by the tracr-mate sequence GUUUUAGAGCUA (SEQ ID NO: 30) followed by the loop sequence GAAA. Results of SURVEYOR assays for Cas9-mediated indels at the human EMX1 and PVALB loci are illustrated in FIGS. 18b and 18c, respectively. Arrows indicate the expected SURVEYOR fragments. ChiRNAs are indicated by their "*n" designation, and crRNA refers to a hybrid RNA where guide and tracr sequences are expressed as separate transcripts. Quantification of these results, performed in triplicate, are illustrated by histogram in FIGS. 19a and 19b, corresponding to FIGS. 18b and 18c, respectively. (*N.D. indicates no indels detected). Protospeacer IDs and their corresponding genomic target, protospeacer sequence, PAM sequence, and strand location are provided in Table D. Guide sequences were designed to be complementary to the entire protospeacer sequence in the case of separate transcripts in the hybrid system, or only to the underlined portion in the case of chimeric RNAs.

| Table D |
|---|---|---|---|
| protospeacer genomic ID | target | protospeacer sequence (5' to 3') | PAM strand |
| 1 | EMX1 | GCAGATCATGAGCAGACCTCTACATGCATAG | TGG + |
| 2 | EMX1 | CATTGAGGAGGAGGATCTGCTGCTCCT | TGG - |
| 3 | EMX1 | GGAGGCGCTCTGAGCTGAGGAAGAA | GAG + |
| 4 | PVALB | GTCAGTGCAGAAGAGGCAGGATTTGCT | ACG - |
| 5 | PVALB | ATGGCAATATGGCTGAGGAGGCGG | GAT (SEQ ID NO: 35) |

Cell Culture and Transfection
Human embryonic kidney (HEK) cell line 293FT (Life Technologies) was maintained in Dulbecco's modified Eagle's Medium (DME) supplemented with 10% fetal bovine serum (HyClone), 2 mM GlutaMAX (Life Technologies), 100 U/mL penicillin, and 100 μg/mL streptomycin at 37° C. with 5% CO2 incubation. 293FT cells were seeded onto 24-well plates (Corning) 24 hours prior to transfection at a density of 150,000 cells per well. Cells were transfected using Lipofectamine 2000 (Life Technologies) following the manufacturer's recommended protocol. For each well of a 24-well plate, a total of 500 ng plasmid was used.

SURVEYOR assay for Genome Modification
293FT cells were transfected with plasmid DNA as described above. Cells were incubated at 37° C. for 72 hours post-transfection prior to genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer's protocol. Briefly, pelleted cells were resuspended in QuickExtract solution and incubated at 65° C. for 15 minutes and 98° C. for 10 minutes. The genomic region flanking the CRISPR target site for each gene was PCR amplified (primers listed in Table...
TABLE E

<table>
<thead>
<tr>
<th>primer name</th>
<th>genomic target</th>
<th>primer sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp-EMX1-F</td>
<td>EMX1</td>
<td>AAAACACCCTTCTTCTTCG</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 36)</td>
<td></td>
</tr>
<tr>
<td>Sp-EMX1-R</td>
<td>EMX1</td>
<td>GGAGGATTTGGGACAGCGAGG</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 37)</td>
<td></td>
</tr>
<tr>
<td>Sp-PVALB-F</td>
<td>PVALB</td>
<td>CTCGAGACGGCCAGGACTGGC</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 38)</td>
<td></td>
</tr>
<tr>
<td>Sp-PVALB-R</td>
<td>PVALB</td>
<td>GGCGACCAACTCTCCTTCTC</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 39)</td>
<td></td>
</tr>
</tbody>
</table>

Computational Identification of Unique CRISPR Target Sites

To identify unique target sites for the S. pyogenes SF370 Cas9 (SpCas9) enzyme in the human, mouse, rat, zebrafish, fruit fly, and C. elegans genome, we developed a software package to scan both strands of a DNA sequence and identify all possible SpCas9 target sites. For each example, every SpCas9 target site was operationally defined as a 20 bp sequence followed by an NNG protospacer adjacent motif (PAM) sequence, and we identified all sequences satisfying this 5'-N<sub>5</sub>NGG-3' (SEQ ID NO: 542) definition on all chromosomes. To prevent non-specific genome editing, after identifying all potential sites, all target sites were filtered based on the number of times they appear in the relevant reference genome. To take advantage of sequence specificity of Cas9 activity conferred by a ‘seed’ sequence, which can be, for example, approximately 11-12 bp sequence 5' from the PAM sequence, 5'-NNNNNNNNNNNGG-3' (SEQ ID NO: 543) sequences were selected to be unique in the relevant genome. All genomic sequences were downloaded from the UCSC Genome Browser (Human genome hg19, Mouse genome mm9, Rat genome rn5, Zebrafish genome danRer7, D. melanogaster genome dm4 and C. elegans genome ce10). The full search results are available to browse using UCSC Genome Browser information. An example visualization of some target sites in the human genome is provided in FIG. 21.

Initially, three sites within the EMX1 locus in human HEK 293FT cells were targeted. Genome modification efficiency of each chimeric RNA was assessed using the SURVEYOR nuclease assay, which detects mutations resulting from DNA double-strand breaks (DSBs) and their subsequent repair by the non-homologous end joining (NHEJ) DNA damage repair pathway. Constructs designated chimeric RNA (+4n) indicate that up to the +4 nt nucleotide of wild-type tracrRNA is included in the chimeric RNA construct, with values of 48, 54, 67, and 85 used for n. Chimeric RNAs containing longer fragments of wild-type tracrRNA (chimRNA(+67) and chimRNA(+85)) mediated DNA cleavage at all three EMX1 target sites, with chimRNA(+85) in particular demonstrating significantly higher levels of DNA cleavage than the corresponding crRNA/tracrRNA hybrids that expressed guide and tracr sequences in separate transcripts (FIGS. 18b and 19a). Two sites in the PVALB locus that yielded no detectable cleavage using the hybrid system (guide sequence and tracr sequence expressed as separate transcripts) were also targeted using chimeric RNAs. chimRNA(+67) and chimRNA(+85) were able to mediate significant cleavage at the two PVALB protospacers (FIGS. 18c and 19b).

For all five targets in the EMX1 and PVALB loci, a consistent increase in genome modification efficiency with increasing tracr sequence length was observed. Without wishing to be bound by any theory, the secondary structure formed by the 3' end of the tracrRNA may play a role in enhancing the rate of CRISPR complex formation. An illustration of predicted secondary structures for each of the chimeric RNAs used in this example is provided in FIG. 21. The secondary structure was predicted using RNAfold (http://rna.ubiduniv.ca/cgi-bin/RNAfold.cgi) using minimum free energy and partition function algorithm. Pseudocolor for each base (reproduced in grayscale) indicates the probability of pairing. Because chimeric RNAs with longer tracr sequences were able to cleave targets that were not cleaved by native CRISPR crRNA/tracrRNA hybrids, it is possible that chimeric RNA may be loaded onto Cas9 more efficiently than its native hybrid counterpart. To facilitate the application of Cas9 for target-specific genome editing in eukaryotic cells and organisms, all predicted unique target sites for the S. pyogenes Cas9 were computationally identified in the human, mouse, rat, zebrafish, C. elegans, and D. melanogaster genomes. Chimeric RNAs can be designed for Cas9 enzymes from other microbes to expand the target space of CRISPR RNA-programmable nucleases.

FIG. 22 illustrates an exemplary bimiconic expression vector for expression of chimeric RNA including up to the +85 nucleotide of wild-type tracr RNA sequence, and SpCas9 with nuclear localization sequences. SpCas9 is expressed from a C8h promoter and terminated with a bGH polyA signal (bGH pA). The expanded sequence illustrated immediately below the sequence corresponds to the region surrounding the guide sequence insertion site, and includes, from 5' to 3',3'-portion of the U6 promoter (first shaded region), BbsI cleavage sites (arrows), partial direct repeat (tracr mate sequence GTTTTAGCCTA (SEQ ID NO: 27), underlined), loop sequence GAAA, and +85 tracr sequence (underlined sequence following loop sequence). An exemplary guide sequence is inserted below the guide sequence insertion site, with nucleotides of the guide sequence for a selected target represented by an “N”. Sequences described in the above examples are as follows (polynucleotide sequences are 5' to 3'):

U6-short tracrRNA (Streptococcus pyogenes SF370):

<table>
<thead>
<tr>
<th>Sequence</th>
<th>(SEQ ID NO: 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAAGGCTCATTTTCCAGATTTCTTCAATATTGCTATTTCCACACCGAAGGCC</td>
<td></td>
</tr>
<tr>
<td>TTGGTACGCGTATTGACATATTGACGTACAATAAGAGTTATAG</td>
<td></td>
</tr>
<tr>
<td>TACAAAAATTGAGCGGCTGAGCAATTTATTTCTGGTGATTTTTGATT</td>
<td></td>
</tr>
<tr>
<td>TTAAAAATTGTTTTTATAATTTGGCAGACTCATATAGTTCTCAGTACAGTGAAN</td>
<td></td>
</tr>
<tr>
<td>GTATTTGCGATTCCTGCTTTATATATCTGTTGGAGAAGGCGAAGACCCCG</td>
<td></td>
</tr>
</tbody>
</table>
RNA-Guided Editing of Bacterial Genomes using CRISPR-Cas Systems

Applicants used the CRISPR-associated endonuclease Cas9 to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli*. The approach relied on Cas9-directed cleavage at the targeted site to kill unmutated cells and circumvented the need for selectable markers or counter-selection systems. Cas9 specificity was reprogrammed by changing the sequence of short CRISPR RNA (crRNA) to make single- and multi-nucleotide changes carried on editing templates. Simultaneous use of two crRNAs enabled multiplex mutagenesis. In *S. pneumoniae*, nearly 100% of cells that survived Cas9 cleavage contained the desired mutation, and 0% when combined with recombining in *E. coli*. Applicants exhaustively analyzed Cas9 target requirements to define the range of targetable sequences and showed strategies for editing sites that do not meet these requirements, suggesting the versatility of this technique for bacterial genome engineering.

The understanding of gene function depends on the possibility of altering DNA sequences within the cell in a controlled fashion. Site-specific mutagenesis in eukaryotes is achieved by the use of sequence-specific nucleases that promote homologous recombination of a template DNA containing the mutation of interest. Zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and homing meganucleases can be programmed to cleave genomes in specific locations, but these approaches require engineering of new enzymes for each target sequence. In prokaryotic organisms, mutagenesis methods either introduce a selection marker in the edited locus or require a two-step process that includes a counter-selection system. More recently, phage recombination proteins have been used for recombining, a technique that promotes homologous recombination of linear DNA or oligonucleotides. However, because there is no selection of mutations, recombining efficiency can be relatively low (0.1-10%) for point mutations down to $10^{-5}$-$10^{-6}$ for larger modifications, in many cases requiring the screening of a large number of colonies. Therefore new technologies that are affordable, easy to use and efficient are still in need for the genetic engineering of both eukaryotic and prokaryotic organisms.

Recent work on the CRISPR (clustered, regularly interspaced, short palindromic repeats) adaptive immune system of prokaryotes has led to the identification of nucleases whose sequence specificity is programmed by small RNAs. CRISPR loci are composed of a series of repeats separated by “spacer” sequences that match the genomes of bacteriophages and other mobile genetic elements. The repeat-spacer array is transcribed as a long precursor and processed within repeat sequences to generate small crRNA that specify the target sequences (also known as protospacers) cleaved by CRISPR systems. Essential for cleavage is the presence of a sequence motif immediately downstream of the target region, known as the protospacer-adjacent motif (PAM). CRISPR-associated (cas) genes usually flank the repeat-spacer array and encode the enzymatic machinery responsible for crRNA biogenesis and targeting. Cas9 is a dsDNA endonuclease that uses a crRNA guide to specify the site of cleavage. Loading of the crRNA guide onto Cas9 occurs during the processing of the crRNA precursor and requires a small RNA antisense to the precursor, the tracrRNA, and RNAse III. In contrast to genome editing with ZFNs or TALENs, changing Cas9 target
specificity does not require protein engineering but only the design of the short crRNA guide.

Applicants recently showed in S. pneumoniae that the introduction of a CRISPR system targeting a chromosomal locus leads to the killing of the transformed cells. It was observed that occasional survivors contained mutations in the target region, suggesting that Cas9 dsiDNA endonuclease activity against endogenous targets could be used for genome editing. Applicants showed that marker-less mutations can be introduced through the transformation of a template DNA fragment that will recombine in the genome and eliminate Cas9 target recognition. Directing the specificity of Cas9 with several different crRNAs allows for the introduction of multiple mutations at the same time. Applicants also characterized in detail the sequence requirements for Cas9 targeting and show that the approach can be combined with recombining for genome editing in E. coli.

RESULTS: Genome Editing by Cas9 Cleavage of a Chromosomal Target

S. pneumoniae strain crR6 contains a Cas9-based CRISPR system that cleaves a target sequence present in the bacteriophage $\phi$S232.5. This target was integrated into the srtA chromosomal locus of a second strain R6$^{S232.5}$. An altered target sequence containing a mutation in the PAM region was integrated into the srtA locus of a third strain R6$^{S70.1}$, rendering this strain "immune" to CRISPR cleavage (FIG. 28a). Applicants transformed R6$^{S232.5}$ and R6$^{S70.1}$ cells with genomic DNA from crR6 cells, expecting that successful transformation of R6$^{S232.5}$ cells should lead to cleavage of the target locus and cell death. Contrary to this expectation, Applicants isolated R6$^{S232.5}$ transformants, albeit with approximately 10-fold less efficiency than R6$^{S70.1}$ transformants (FIG. 28b). Genetic analysis of eight R6$^{S232.5}$ transformants (FIG. 28) revealed that the majority are the product of a double recombination event that eliminates the toxicity of Cas9 targeting by replacing the 48232.5 target with the crR6 genome’s wild-type srtA locus, which does not contain the protospacer required for Cas9 recognition. These results were proof that the concurrent introduction of a CRISPR system targeting a genomic locus (the targeting construct) together with a template for recombination into the targeted locus (the editing template) led to targeted genome editing (FIG. 23c).

To create a simplified system for genome editing, Applicants modified the CRISPR locus in strain crR6 by deleting cas1, cas2 and cas2, genes which have been shown to be dispensable for CRISPR targeting, yielding strain crRM (FIG. 28a). This strain retained the same properties of crR6 (FIG. 28b). To increase the efficiency of Cas9-based editing and demonstrate that a template DNA of choice can be used to control the mutation introduced, Applicants co-transformed R6$^{S232.5}$ cells with PCR products of the wild-type srtA gene or the mutant R6$^{S70.1}$ target, either of which should be resistant to cleavage by Cas9. This resulted in a 5- to 10-fold increase of the frequency of transformation compared with genomic crRM DNA alone (FIG. 23b). The efficiency of editing was also substantially increased, with 8/8 transformants tested containing a wild-type srtA copy and 7/8 containing the PAM mutation present in the R6$^{S70.1}$ target (FIG. 23b and FIG. 29a). Taken together, these results showed the potential of genome editing assisted by Cas9.

Analysis of Cas9 Target Requirements:

To introduce specific changes in the genome, one must use an editing template carrying mutations that abolish Cas9-mediated cleavage, thereby preventing cell death. This is easy to achieve when the deletion of the target or its replacement by another sequence (gene insertion) is sought. When the goal is to produce gene fusions or to generate single-nucleotide mutations, the abolishment of Cas9 nuclease activity will only be possible by introducing mutations in the editing template that alter either the PAM or the protospacer sequences. To determine the constraints of CRISPR-mediated editing, Applicants performed an exhaustive analysis of PAM and protospacer mutations that abrogate CRISPR targeting.

Previous studies proposed that S. pyogenes Cas9 requires an NGG PAM immediately downstream of the protospacer. However, because only a very limited number of PAM-inactivating mutations have been described so far, Applicants conducted a systematic analysis to find all 5-nucleotide sequences following the protospacer that eliminate CRISPR cleavage. Applicants used randomized oligonucleotides to generate all possible 1,024 PAM sequences in a heterogenous PCR product that was transformed into crR6 or R6 cells. Constructs carrying functional PAMs were expected to be recognized and destroyed by Cas9 in crR6 but not R6 cells (FIG. 24a). More than 2x10^5 colonies were pooled together to extract DNA for use as template for the co-amplification of all targets. PCR products were deep sequenced and found to contain all 1,024 sequences, with coverage ranging from 5 to 42,472 reads (See section “Analysis of deep sequencing data”). The functionality of each PAM was estimated by the relative proportion of its reads in the crR6 sample over the R6 sample. Analysis of the first three bases of the PAM, averaging over the two last bases, clearly showed that the NGG pattern was under-represented in crR6 transformants (FIG. 24b). Furthermore, the next two bases had no detectable effect on the NGG PAM (See section “Analysis of deep sequencing data”), demonstrating that the NGGNN sequence was sufficient to license Cas9 activity. Partial targeting was observed for NAG PAM sequences (FIG. 24b). Also the NNGN pattern partially inactivated CRISPR targeting (Table G), indicating that the NGG motif can still be recognized by Cas9 with reduced efficiency when shifted by 1 bp. These data shed light onto the molecular mechanism of Cas9 target recognition, and they revealed that NGG (or CCN on the complementary strand) sequences are sufficient for Cas9 targeting and that NGG to NAG or NN GN mutations in the editing template should be avoided. Owing to the high frequency of these tri-nucleotide sequences (once every 8 bp), this means that almost any position of the genome can be edited. Indeed, Applicants tested ten randomly chosen targets carrying various PAMs and all were found to be functional (FIG. 30).

Another way to disrupt Cas9-mediated cleavage is to introduce mutations in the protospacer region of the editing template. It is known that point mutations within the 'seed sequence' (the 8 to 10 protospacer nucleotides immediately adjacent to the PAM) can abolish cleavage by CRISPR nucleases. However, the exact length of this region is not known, and it is unclear whether mutations to any nucleotide in the seed can disrupt Cas9 target recognition. Applicants followed the same deep sequencing approach described above to randomize the entire protospacer sequence involved in base pair contacts with the crRNA and to determine all sequences that disrupt targeting. Each position of the 20 matching nucleotides (14 in the spel target present in R6$^{S232.5}$ cells (FIG. 23a) was randomized and transformed into crR6 and R6 cells (FIG. 24a). Consistent with the presence of a seed sequence, only mutations in the 12 nucleotides immediately upstream of the PAM abrogated cleavage by Cas9 (FIG. 24c). However, different mutations displayed markedly different effects. The distal (from the PAM) positions of the seed (12 to 7) tolerated most mutations and only one particular base substitution abrogated targeting. In con-
Cas9-Mediated Genome Editing in S. pneumoniae:

To develop a rapid and efficient method for targeted genome editing, Applicants engineered strain crI69k, a streaked variant of S. pneumoniae strain 603, which can be multiply transformed by PCR (FIG. 1). Applicants decided to edit the β-galactosidase (bgaA) gene of S. pneumoniae, whose activity can be easily measured. Applicants introduced alanine substitutions of amino acids in the active site of this enzyme: R481A (R→A) and N563A,E564A (NE→AA) mutations. To illustrate different editing strategies, Applicants designed mutations of both the PAM sequence and the protospacer seed. In both cases the same targeting construct with a crRNA complementary to a region of the β-galactosidase gene that is adjacent to a TGG PAM sequence (CCA in the complementary strand, FIG. 26) was used. The R→A editing template created a three-nucleotide mismatch on the protospacer seed sequence (CGG to GCA, also introducing a BglII restriction site). In the NE→AA editing template Applicants simultaneously introduced a synonymous mutation that created an inactive PAM (TGG to TTG) along with mutations that were 218 at downstream of the protospacer region (AAT GAA to GCT GCA, also generating a TseI restriction site). This last editing strategy demonstrated the possibility of using a remote PAM to make mutations in places where a proper target may be hard to choose. For example, although the S. pneumoniae R6 genome, which has a 39.7% GC content, contains on average one PAM motif every 12 bp, some PAM motifs are separated by up to 194 bp (FIG. 33). In addition Applicants designed a ΔbgaA in-frame deletion of 6,664 bp. In all three cases, co-transformation of the targeting and editing templates produced 10-times more kanamycin-resistant cells than co-transformation with a control editing template containing wild-type bgaA sequences (FIG. 25b). Applicants generated 24 transformants (8 for each editing experiment) and found that all but one incorporated the desired change. (FIG. 25c). DNA sequencing also confirmed not only the presence of the introduced mutations but also the absence of secondary mutations in the target region (FIG. 29c). Finally, Applicants measured β-galactosidase activity to confirm that all edited cells displayed the expected phenotype (FIG. 25d).

Cas9-mediated editing can also be used to generate multiple mutations for the study of biological pathways. Applicants decided to illustrate this for the sortase-dependent pathway that anchors surface proteins to the envelope of Gram-positive bacteria. Applicants introduced a sortase deletion by co-transformation of a chloramphenicol-resistant targeting construct and a ΔsrA editing template (FIG. 33a,b), followed by a ΔbgaA deletion using a kanamycin-resistant targeting construct that replaced the previous one. In S. pneumoniae, β-galactosidase is covalently linked to the cell wall by sortase. Therefore, deletion of srA results in the release of the surface protein into the supernatant, whereas the double deletion has no detectable β-galactosidase activity (FIG. 34c).

Such a sequential selection can be iterated as many times as required to generate multiple mutations. These two mutations may also be introduced at the same time. Applicants designed a targeting construct containing two spacers, one matching srA and the other matching bgaA, and co-transformed it with both editing templates at the same time (FIG. 25c). Genetic analysis of transformants showed that editing occurred in 6/8 cases (FIG. 25f). Notably, the remaining two clones each contained either a ΔsrA or a ΔbgaA deletion, suggesting the possibility of performing combinatorial mutagenesis using Cas9. Finally, to eliminate the CRISPR sequences, Applicants introduced a plasmid containing the bgaA target and a spectinomycin resistance gene along with genomic DNA from the wild-type strain R6. Spectinomycin-resistant transformants that retain the plasmid eliminated the CRISPR sequences (FIG. 34a,d).

Mechanism and Efficiency of Editing:

To understand the mechanisms underlying genome editing with Cas9, Applicants designed an experiment in which the editing efficiency was measured independently of Cas9 cleavage. Applicants integrated the ermAM erythromycin resistance gene in the srA locus, and introduced a premature stop codon using Cas9-mediated editing (FIG. 33). The resulting strain (JEN53) contains an ermAM(stop) allele and is sensitive to erythromycin. This strain may be used to assess the efficiency at which the ermAM gene is repaired by measuring the fraction of cells that restore antibiotic resistance with or without the use of Cas9 cleavage. JEN53 was transformed with an editing template that restores the wild-type allele, together with either a kanamycin-resistant CRISPR construct targeting the ermAM(stop) allele (CRISPR::ermAM(stop)) or a control construct without a spacer (CRISPR::0) (FIG. 26a,b). In the absence of kanamycin selection, the fraction of edited colonies was on the order of 10⁻³ (erythromycin-resistant cfu/total cfu) (FIG. 26c), representing the baseline frequency of recombination without Cas9-mediated selection against unedited cells. However, if kanamycin selection was applied and the control CRISPR construct was co-transformed, the fraction of edited colonies increased to about 10⁻¹ (kanamycin- and erythromycin-resistant cfu/kanamycin-resistant cfu) (FIG. 26c). This result shows that selection for the recombination of the CRISPR locus co-selected for recombination in the ermAM locus independently of Cas9 cleavage of the genome, suggesting that a subpopulation of cells is more prone to transformation and/or recombination. Transformation of the CRISPR::ermAM(stop) construct followed by kanamycin selection resulted in an increase of the fraction of erythromycin-resistant, edited cells to 99% (FIG. 26c). To determine if this increase is caused by the killing of non-edited cells, Applicants compared the kanamycin-resistant colony forming units (cfu) obtained after co-transformation of JEN53 cells with the CRISPR::ermAM(stop) or CRISPR::0 constructs.

Applicants counted 5.3 times less kanamycin-resistant colonies after transformation of the ermAM(stop) construct (2.5x10⁴/4.7x10⁵, FIG. 35a), a result that suggests that indeed targeting of a chromosomal locus by Cas9 leads to the killing of non-edited cells. Finally, because the introduction of dsDNA breaks in the bacterial chromosome is known to trigger repair mechanisms that increase the rate of recombination of the damaged DNA, Applicants investigated whether cleavage by Cas9 induces recombination of the editing template. Applicants counted 2.2 times more colonies after co-transformation with the CRISPR::erm(stop) construct than with the CRISPR::0 construct (FIG. 26d), indicating that there was a modest induction of recombination. Taken together, these results showed that co-selection of transform-
able cells, induction of recombination by Cas9-mediated cleavage and selection against non-edited cells, each contributed to the high efficiency of genome editing in S. pneumoniae.

As cleavage of the genome by Cas9 should kill non-edited cells, one would not expect to recover any cells that received the kanamycin-resistance—containing Cas9 cassette but not the editing template. However, in the absence of the editing template Applicants recovered many kanamycin-resistant colonies after transformation of the CRISPR:ermAM(stop) construct (FIG. 35a). These cells that ‘escape’ CRISPR-induced death produced a background that determined a limit of the method. This background frequency may be calculated as the ratio of CRISPR:ermAM(stop)/CRISPR::0 cfu, 6.6x10^(-7) (7.1x10^(-7)/2.7x10^(-5)) in this experiment, meaning that if the recombination frequency of the editing template is less than this value, CRISPR selection may not efficiently recover the desired mutants above the background. To understand the origin of these cells, Applicants genotyped 8 background colonies and found that 7 contained deletions of the targeting spacer (FIG. 35b) and one harbored a presumably inactivating mutation in Cas9 (FIG. 35c).

Genome Editing with Cas9 in E. coli:

The activation of Cas9 targeting through the chromosomal integration of a CRISPR-Cas system is only possible in organisms that are highly recombinogenic. To develop a more general method that is applicable to other microbes, Applicants decided to perform genome editing in E. coli using a plasmid-based CRISPR-Cas system. Two plasmids were constructed: a pCas9 plasmid carrying the tracrRNA, Cas9 and a chloramphenicol resistance cassette (FIG. 36), and a pCRISPR kanamycin-resistant plasmid carrying the array of CRISPR spacers. To measure the efficiency of editing independently of CRISPR selection, Applicants sought to introduce an A to C transversion in the rpsL gene that confers streptomycin resistance. Applicants constructed a pCRISPR::rpsL plasmid harboring a spacer that would guide Cas9 cleavage of the wild-type, but not the mutant rpsL allele (FIG. 27b).

The pCas9 plasmid was first introduced into E. coli MG1655 and the resulting strain co-transformed with the pCRISPR::rpsL plasmid and W542, an editing oligonucleotide containing the A to C mutation. streptomycin-resistant colonies after transformation of the pCRISPR::rpsL plasmid were only recovered, suggesting that Cas9 cleavage induces recombination of the oligonucleotide (FIG. 37). However, the number of streptomycin-resistant colonies was two orders of magnitude lower than the number of kanamycin-resistant colonies, which are presumably cells that escape cleavage by Cas9. Therefore, in these conditions, cleavage by Cas9 facilitated the introduction of the mutation, but with an efficiency that was not enough to select the mutant cells above the background of ‘escapers’.

To improve the efficiency of genome editing in E. coli, Applicants applied their CRISPR system with recombineering, using Cas9-induced cell death to select for the desired mutations. The pCas9 plasmid was introduced into the recombineering strain HME63 (31), which contains the Gum, Exo and Beta functions of the red phase. The resulting strain was co-transformed with the pCRISPR::rpsL plasmid (or a pCRISPR::0 control) and the W542 oligonucleotide (FIG. 27a). The recombineering efficiency was 5.3x10^(-4), calculated as the fraction of total cells that become streptomycin-resistant when the control plasmid was used (FIG. 27c). In contrast, transformation with the pCRISPR::rpsL plasmid increased the percentage of mutant cells to 65±14% (FIGS. 27c and 29). Applicants observed that the number of cfu was reduced by about three orders of magnitude after transformation of the pCRISPR::rpsL plasmid than the control plasmid (4.8x10^3/5.3x10^2, FIG. 38a), suggesting that selection results from CRISPR-induced death of non-edited cells. To measure the rate at which Cas9 cleavage was inactivated, an important parameter of Applicants’ method, Applicants transformed cells with either pCRISPR::rpsL or the control plasmid without the W542 editing oligonucleotide (FIG. 38a). This background of CRISPR ‘escapers’, measured as the ratio of pCRISPR::rpsL/pcRISPR::0 cfu, was 2.5x10^(-4) (1.2x10^(-4)/4.8x10^(-4)). Genotyping eight of these escapers revealed that in all cases there was a deletion of the targeting spacer (FIG. 38b). This background was higher than the recombineering efficiency of the rpsL mutation, 5.3x10^(-3), which suggested that to obtain 65% of edited cells, Cas9 cleavage must induce oligonucleotide recombination. To confirm this, Applicants compared the number of kanamycin- and streptomycin-resistant cfu after transformation of pCRISPR::rpsL or pCRISPR::0 (FIG. 27a). As in the case for S. pneumoniae, Applicants observed a modest induction of recombination, about 6.7 fold (2.0x10^(-3)/3.0x10^(-4)). Taken together, these results indicated that the CRISPR system provided a method for selecting mutations introduced by recombineering.

Applicants showed that CRISPR-Cas systems may be used for targeted genome editing in bacteria by the co-introduction of a targeting construct that killed wild-type cells and an editing template that both eliminated CRISPR cleavage and introduced the desired mutations. Different types of mutations (insertions, deletions or scar-less single-nucleotide substitutions) may be generated. Multiple mutations may be introduced at the same time. The specificity and versatility of editing using the CRISPR system relied on several unique properties of the Cas9 endonuclease: (i) its target specificity may be programmed with a small RNA, without the need for enzyme engineering, (ii) target specificity was very high, determined by a 20 bp RNA-DNA interaction with low probability of non-target recognition, (iii) almost any sequence may be targeted, the only requirement being the presence of an adjacent NGG sequence, (iv) almost any mutation in the NGG sequence, as well as mutations in the seed sequence of the protospacer, eliminates targeting.

Applicants showed that genome engineering using the CRISPR system worked not only in highly recombinogenic bacteria such as S. pneumoniae, but also in E. coli. Results in E. coli suggested that the method may be applicable to other microorganisms for which plasmids may be introduced. In E. coli, the approach complements recombineering of mutagenic oligonucleotides. To use this methodology in microbes where recombineering is not possible, the host homologous recombination machinery may be used by providing the editing template on a plasmid. In addition, because accumulated evidence indicates that CRISPR-mediated cleavage of the chromosome leads to cell death in many bacteria and archaea, it is possible to envision the use of endogenous CRISPR-Cas systems for editing purposes.

In both S. pneumoniae and E. coli, Applicants observed that although editing was facilitated by a co-selection of transformable cells and a small induction of recombination at the target site by Cas9 cleavage, the mechanism that contributed the most to editing was the selection against non-edited cells. Therefore the major limitation of the method was the presence of a background of cells that escape CRISPR-induced cell death and lack the desired mutation. Applicants showed that these ‘escapers’ arose primarily through the deletion of the targeting spacer, presumably after the recombination of the repeat sequences that flank the targeting spacer. Future improvements may focus on the engineering of flank-
ing sequences that can still support the biogenesis of functional crRNAs but that are sufficiently different from one another to eliminate recombination. Alternatively, the direct transformation of chimeric crRNAs may be explored. In the particular case of E. coli, the construction of the CRISPR-Cas system was not possible if this organism was also used as a cloning host. Applicants solved this issue by placing Cas9 and the tracrRNA on a different plasmid than the CRISPR array. The engineering of an inducible system may also circumvent this limitation.

Although new DNA synthesis technologies provide the ability to cost-effectively create any sequence with a high throughput, it remains a challenge to integrate synthetic DNA in living cells to create functional genomes. Recently, the co-selection MAGe strategy was shown to improve the mutation efficiency of recombining by selecting a subpopulation of cells that has an increased probability to achieve recombination at or around a given locus. In this method, the introduction of selectable mutations is used to increase the chances of generating nearby non-selectable mutations. As opposed to the indirect selection provided by this strategy, the use of the CRISPR system makes it possible to directly select for the desired mutation and to recover it with a high efficiency. These technologies add to the toolbox of genetic engineers, and together with DNA synthesis, they may substantially advance both the ability to decipher gene function and to manipulate organisms for biotechnological purposes. Two other studies also relate to CRISPR-assisted engineering of mammalian genomes. It is expected that these crRNA-directed genome editing technologies may be broadly useful in the basic and medical sciences.

Strains and Culture Conditions.
S. pneumoniae strain R6 was provided by Dr. Alexander Tomasz. Strain cr6 was generated in a previous study. Liquefied cultures of S. pneumoniae were grown in THYE medium (50 g/l Todd-Hewitt agar, 5 g/l yeast extract). Cells were plated on tryptic soy agar (TSA) supplemented with 5% defibrinated sheep blood. When appropriate, antibiotics were added as followings: kanamycin (400 μg/ml), chloramphenicol (5 μg/ml), erythromycin (1 μg/ml), streptomycin (100 μg/ml) or spectinomycin (100 μg/ml). Measurements of β-galactosidase activity were made using the Miller assay as previously described.

E. coli strains MG1655 and HME63 (derived from MG1655, A(argF-lac)U169, cl857 cropped, galk tyr 145 UAG mutSc >amp) (31) were provided by Jeff Roberts and Donald Court, respectively. Liquid cultures of E. coli were grown in LB medium (Difco). When appropriate, antibiotics were added as followings: chloramphenicol (25 μg/ml), kanamycin (25 μg/ml) and streptomycin (50 μg/ml).

S. pneumoniae Transformation.
Competent cells were prepared as described previously (25). For all genome editing transformations, cells were gently thawed on ice and resuspended in 10 volumes of M2 medium supplemented with 100 μg/ml of competence-stimulating peptide (CSP) (40), and followed by addition of editing constructs (editing constructs were added to cells at a final concentration between 0.7 ng/ml to 2.5 μg/ml). Cells were incubated 20 min at 37°C before the addition of 2 μl of targeting constructs and then incubated 40 min at 37°C. Serial dilutions of cells were plated on the appropriate medium to determine the colony forming units (cfu) count. E. coli Lambda-red recombining. Strain HME63 was used for all recombining experiments. Recombining cells were prepared and handled according to a previously published protocol (6). Briefly, a 2 ml overnight culture (LB medium) inoculated from a single colony obtained from a plate was grown at 30°C. The overnight culture was diluted 100-fold and grown at 30°C with shaking (200 rpm) until the OD₆₀₀ is 0.4-0.5 (approximately 3 hrs). For Lambda-red induction, the culture was transferred to a 42°C water bath to shake at 200 rpm for 15 min. Immediately after induction, the culture was swirled in an ice-water slurry and chilled on ice for 5-10 min. Cells were then washed and aliquoted according to the protocol. For electro-transformation, 50 μA of cells were mixed with 1 mM of salt-free oligos (IDT) or 100-150 ng of plasmid DNA (prepared by Qiagen Prep Spin MiniPrep Kit, Qiagen). Cells were electrooporated using a 1 mm Gene Pulser cuvette (Bio-rad) at 1.8 KV and were immediately resuspended in 1 ml of room temperature LB medium. Cells were recovered at 30°C for 1-2 hrs before being plated on LB agar with appropriate antibiotic resistance and incubated at 32°C overnight.
Preparation of S. pneumoniae Genomic DNA.
For transformation purposes, S. pneumoniae genomic DNA was extracted using the Wizard Genomic DNA Purification Kit, following instructions provided by the manufacturer (Promega). For genotyping purposes, 700 μl of overnight S. pneumoniae cultures were pelleted, resuspended in 60 μl of lysozyme solution (2 mg/ml) and incubated 30 min at 37°C. The genomic DNA was extracted using Qiagen Prep Spin MiniPrep Kit (Qiagen).

Strain Construction.
All primers used in this study are provided in Table 6. To generate S. pneumoniae crR6M, an intermediate strain, LAM226, was made. In this strain the apha-3 gene (providing kanamycin resistance) adjacent to the CRISPR array of S. pneumoniae crR6 strain was replaced by a cat gene (providing chloramphenicol resistance). Briefly, crR6 genomic DNA was amplified using primers L448/L444 and L447/L481, respectively. The cat gene was amplified from plasmid pC194 using primers L445/L446. Each PCR product was gel-purified and all three were fused by SOEing PCR with primers L448/L481. The resulting PCR product was transformed into competent S. pneumoniae crR6 cells and chloramphenicol-resistant transformants were selected. To generate S. pneumoniae crR6M, S. pneumoniae crR6 genomic DNA was amplified by PCR using primers L409/L488 and L448/L481, respectively. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L409/L481. The resulting PCR product was transformed into competent S. pneumoniae LAM226 cells and kanamycin-resistant transformants were selected.

To generate S. pneumoniae crR6Re, S. pneumoniae crR6M genomic DNA was amplified by PCR using primers L430/W286, and S. pneumoniae LAM226 genomic DNA was amplified by PCR using primers W288/L481. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L430/L481. The resulting PCR product was transformed into competent S. pneumoniae crR6M cells and chloramphenicol-resistant transformants were selected.

To generate S. pneumoniae crR6Rk, S. pneumoniae crR6M genomic DNA was amplified by PCR using primers L430/W286 and W287/L481, respectively. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L430/L481. The resulting PCR product was transformed into competent S. pneumoniae crR6Rc cells and kanamycin-resistant transformants were selected.

To generate JEN37, S. pneumoniae crR6Rk genomic DNA was amplified by PCR using primers L430/W356 and W357/L481, respectively. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L430/L481.
The resulting PCR product was transformed into competent *S. pneumoniae* crfR6rc cells and kanamycin-resistant transformants were selected.

To generate JEN38, R6 genomic DNA was amplified using primers L422/L461 and L459/L426, respectively. The ermAM gene (specifying erythromycin resistance) was amplified from plasmid plW154* using primers L457/L458. Each PCR product was gel-purified and all three were fused by SOEligPCR with primers L422/L421. The resulting PCR product was transformed into competent *S. pneumoniae* crf6R6rc cells and erythromycin-resistant transformants were selected.

*S. pneumoniae* JEN53 was generated in two steps. First JEN43 was constructed as illustrated in Fig. 33. JEN53 was generated by transforming genomic DNA of JEN25 into competent JEN43 cells and selecting on both chloramphenicol and erythromycin.

To generate *S. pneumoniae* JEN62, *S. pneumoniae* crf6R6rc genomic DNA was amplified by PCR using primers W256/W365 and W366/L403, respectively. Each PCR product was purified and ligated by Gibson assembly. The assembly product was transformed into competent *S. pneumoniae* crf6R6rc cells and kanamycin-resistant transformants were selected.

Plasmid Construction.

pDB97 was constructed through phosphorylation and annealing of oligonucleotides B296/B297, followed by ligation in plZ12spec digested by EcoRI/BamHI. Applicants fully sequenced plZ12spec and deposited its sequence in genebank (accession: KC112384).

pDB98 was obtained after cloning the CRISPR leader sequence was cloned together with a repeat-spacer-repeat unit into plZ12spec. This was achieved through amplification of crfR6rc DNA with primers B298/B320 and B299/B321, followed by SOEligPCR of both products and cloning in plZ12spec with restriction sites BamHI/EcoRI. In this way the spacer sequence in pDB98 was engineered to contain two BsaI restriction sites in opposite directions that allow for the scar-less cloning of new spacers.

pDB99 to pDB108 were constructed by annealing of oligonucleotides B300/B301 (pDB99), B302/B303 (pDB100), B304/B305 (pDB101), B306/B307 (pDB102), B308/B309 (pDB103), B310/B311 (pDB104), B312/B313 (pDB105), B314/B315 (pDB106), B315/B317 (pDB107), B318/B319 (pDB108), followed by ligation in pDB98 cut by BsaI.

The pCas9 plasmid was constructed as follows. Essential CRISPR elements were amplified from *Streptococcus pneumoniae* SF370 genomic DNA with flanking homology arms for Gibson Assembly. The tracrRNA and Cas9 were amplified with oligos HC008 and HC010. The leader and CRISPR sequences were amplified HC011/HC014 and HC015/HC009, so that two BsaI type IIS sites were introduced between two direct repeats to facilitate easy insertion of spacers.

pCRISPR was constructed by subcloning the pCas9 CRISPR array in pZE1-MCS1 through amplification with oligos B298/B299 and restriction with EcoRI and BamHI. The rpsL targeting spacer was cloned by annealing of oligos B354/B353 and cloning in the BsaI cut pCRISPR giving pCRISPR::rpsL.

Generation of Targeting and Editing Constructs.

Targeting constructs used for genome editing were made by Gibson assembly of Left PCRs and Right PCRs (Table 6). Editing constructs were made by SOEligPCR fusing PCR products A (PCR A), PCR products B (PCR B) and PCR products C (PCR C) when applicable (Table 7). The CRISPR::O and CRISPR::ermAMstop targeting constructs were generated by PCR amplification of JEN62 and crf6 genomic DNA respectively, with oligos L409 and L481.

Generation of Targets with Randomized PAM or Protospacer Sequences.

The 5 nucleotides following the spacer 1 target were randomized through amplification of R6*22 by genomic DNA with primers W377/L426. This PCR product was then assembled with the cat gene and the star.A upstream region that were amplified from the same template with primers L422/W376. 80 ng of the assembled DNA was used to transform strains R6 and crf6. Samples for the randomized targets were prepared using the following primers: B280-B290/L426 to randomize bases 1-10 of the target and B269-B278/L426 to randomize bases 10-20. Primers L422/B268 and L422/B279 were used to amplify the cat gene and star.A upstream region to be assembled with the first and last 10 PCR products respectively. The assembled constructs were pooled together and 30 ng was transformed in R6 and crf6. After transformation, cells were plated on chloramphenicol selection. For each sample more than 2×10⁶ cells were pooled together in 1 ml of THYE and genomic DNA was extracted with the Promega Wizard kit. Primers B250/B251 were used to amplify the target region. PCR products were tagged and run on one Illumina MiSeq paired-end lane using 300 cycles.

Analysis of Deep Sequencing Data.

Randomized PAM: For the randomized PAM experiment 3,429,406 reads were obtained for crf6 and 3,253,998 for R6. It is expected that only half of them will correspond to the PAM-target while the other half will sequence the other end of the PCR product. 1,623,008 of the crf6 reads and 1,537,131 of the R6 reads carried an error-free target sequence. The occurrence of each possible PAM among these reads is shown in supplementary file. To estimate the functionality of a PAM, its relative proportion in the crf6 sample over the R6 sample was computed and is denoted \( r_{PAM} \), where \( i, j, k, l \) are one of the 4 possible bases. The following statistical model was constructed:

\[
\log(r_{PAM}) = b_0 + b_1 \cdot x_j + b_2 \cdot x_k + b_3 \cdot x_l + b_4 \cdot x_{x_{PAM}}
\]

where \( \epsilon \) is the residual error, \( b_2 \) is the effect of the 2nd base of the PAM, \( b_3 \) of the third, \( b_4 \) of the fourth, \( b_6 \) is the interaction between the second and third bases, \( b_{34} \) between the third and fourth bases. An analysis of variance was performed:

<table>
<thead>
<tr>
<th>PAM</th>
<th>( b_0 )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( b_3 )</th>
<th>( b_4 )</th>
<th>( r_{PAM} )</th>
<th>( \epsilon )</th>
</tr>
</thead>
<tbody>
<tr>
<td>b2</td>
<td>2.156</td>
<td>2.156</td>
<td>2.156</td>
<td>2.156</td>
<td>2.156</td>
<td>2.156</td>
<td>2.156</td>
</tr>
<tr>
<td>b1</td>
<td>3.156</td>
<td>3.156</td>
<td>3.156</td>
<td>3.156</td>
<td>3.156</td>
<td>3.156</td>
<td>3.156</td>
</tr>
<tr>
<td>b0</td>
<td>4.156</td>
<td>4.156</td>
<td>4.156</td>
<td>4.156</td>
<td>4.156</td>
<td>4.156</td>
<td>4.156</td>
</tr>
<tr>
<td>b4</td>
<td>5.156</td>
<td>5.156</td>
<td>5.156</td>
<td>5.156</td>
<td>5.156</td>
<td>5.156</td>
<td>5.156</td>
</tr>
</tbody>
</table>

When added to this model, b1 or b5 do not appear to be significant and other interactions than the ones included can also be discarded. The model choice was made through successive comparisons of more or less complete models using the anova method in R. Tukey’s honest significance test was used to determine if pairwise differences between effects are significant.

NGGNN patterns are significantly different from all other patterns and carry the strongest effect (see table below).

In order to show that positions 1, 4 or 5 do not affect the NGGNN pattern Applicants looked at theses sequences only. Their effect appears to be normally distributed (see QQ plot in...
Figure 71, and model comparisons using the anova method in R shows that the null model is the best one, i.e. there is no significant role of b1, b4 and b5.

Model Comparison Using the Anova Method in R for the NQGN Sequences

| Res. | Df | RSS | Df | Sum of Sq | F  | P(F>|F|) |
|------|----|-----|----|-----------|----|--------|
| 1    | 63 | 14.579 | 9  | 3.2836    | 1.7443 | 0.1013 |
| 2    | 54 | 11.295 | 9  | 3.2836    | 1.7443 | 0.1013 |

Partial interference of NAGNN and NNGGN patterns
NAGNN patterns are significantly different from all other patterns but carry a much smaller effect than NGGN (see Tukey’s honest significance test below).

Finally, NTGGN and NGCGN patterns are similar and show significantly more CRISPR interference than NTGHN and NGGHN patterns (where H is A, T or C), as shown by a bonferroni adjusted pairwise student-test.

Pairwise Comparisons of the Effect of b4 on NYQNN Sequences Using t Tests with Pooled SD

<table>
<thead>
<tr>
<th>Data: b4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>T</td>
</tr>
</tbody>
</table>

Taken together, these results allow concluding that NNGGN patterns in general produce either a complete interference in the case of NGGN, or a partial interference in the case of NAGNN, NTGNN or NGCGN.

Tukey multiple comparisons of means: 95% family-wise confidence level

<table>
<thead>
<tr>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/G-C:A</td>
<td>-2.79911</td>
<td>-2.97511</td>
<td>-2.62311</td>
</tr>
<tr>
<td>G/G-T:A</td>
<td>-2.7809</td>
<td>-2.9569</td>
<td>-2.6049</td>
</tr>
<tr>
<td>G/G-A:C</td>
<td>-2.81643</td>
<td>-2.99244</td>
<td>-2.64043</td>
</tr>
<tr>
<td>G/G-C:C</td>
<td>-2.77903</td>
<td>-2.95304</td>
<td>-2.60303</td>
</tr>
<tr>
<td>G/G-C:T</td>
<td>-2.64867</td>
<td>-2.82468</td>
<td>-2.47267</td>
</tr>
<tr>
<td>G/G-C-T</td>
<td>-2.79718</td>
<td>-2.97319</td>
<td>-2.62118</td>
</tr>
<tr>
<td>G/G-A:T</td>
<td>-2.67068</td>
<td>-2.84668</td>
<td>-2.49468</td>
</tr>
<tr>
<td>G/G-C:G</td>
<td>-2.73525</td>
<td>-2.91125</td>
<td>-2.59255</td>
</tr>
<tr>
<td>G/G-C:G</td>
<td>-2.7976</td>
<td>-2.82399</td>
<td>-2.6736</td>
</tr>
<tr>
<td>G/G-T:G</td>
<td>-2.76727</td>
<td>-2.95127</td>
<td>-2.64327</td>
</tr>
<tr>
<td>G/G-C:T</td>
<td>-2.84114</td>
<td>-2.65513</td>
<td>-3.01714</td>
</tr>
<tr>
<td>G/G-T:T</td>
<td>-2.76409</td>
<td>-2.58809</td>
<td>-2.9409</td>
</tr>
<tr>
<td>G/G-T:T</td>
<td>-2.76781</td>
<td>-2.9518</td>
<td>-2.94381</td>
</tr>
<tr>
<td>G/G-T:G</td>
<td>-2.13964</td>
<td>-2.31565</td>
<td>-1.96364</td>
</tr>
<tr>
<td>G/A:A:A</td>
<td>-0.62511</td>
<td>-0.8011</td>
<td>-0.4491</td>
</tr>
<tr>
<td>G/A:C:A</td>
<td>-0.65947</td>
<td>-0.83547</td>
<td>-0.48346</td>
</tr>
<tr>
<td>G/A:T:A</td>
<td>-0.64126</td>
<td>-0.64525</td>
<td>-0.81726</td>
</tr>
<tr>
<td>G/A:G:A</td>
<td>-0.67679</td>
<td>-0.80507</td>
<td>-0.68279</td>
</tr>
<tr>
<td>G/A:C:C</td>
<td>-0.63939</td>
<td>-0.64339</td>
<td>-0.81539</td>
</tr>
<tr>
<td>G/A:T:T</td>
<td>-0.66574</td>
<td>-0.68814</td>
<td>-0.83554</td>
</tr>
<tr>
<td>G/A:G:G</td>
<td>-0.65104</td>
<td>-0.35503</td>
<td>-0.70704</td>
</tr>
<tr>
<td>G/A:C:G</td>
<td>-0.59561</td>
<td>-0.4196</td>
<td>-0.77161</td>
</tr>
<tr>
<td>G/A:T:G</td>
<td>-0.65795</td>
<td>-0.68195</td>
<td>-0.83196</td>
</tr>
<tr>
<td>G/A:G:T</td>
<td>-0.62763</td>
<td>-0.45453</td>
<td>-0.80363</td>
</tr>
<tr>
<td>G/A:G:T</td>
<td>-0.70149</td>
<td>-0.52549</td>
<td>-0.8775</td>
</tr>
<tr>
<td>G/A:C:T</td>
<td>-0.63445</td>
<td>-0.48844</td>
<td>-0.80045</td>
</tr>
<tr>
<td>G/A:C:T</td>
<td>-0.62817</td>
<td>-0.45216</td>
<td>-0.80417</td>
</tr>
<tr>
<td>G/G:A:A</td>
<td>-0.3333</td>
<td>-0.51133</td>
<td>-0.15392</td>
</tr>
<tr>
<td>G/G:A:C</td>
<td>-0.18118</td>
<td>-0.35719</td>
<td>-0.00518</td>
</tr>
<tr>
<td>G/G:A:T</td>
<td>-0.31626</td>
<td>-0.14026</td>
<td>-0.49226</td>
</tr>
</tbody>
</table>

Randomized Target
For the randomized target experiment 540,726 reads were obtained for crR6 and 753,570 for R6. As before, only half of the reads are expected to sequence the interesting end of the PCR product. After filtering for reads that carry a target that is error-free or with a single point mutation, 217,656 and 353, 141 reads remained for crR6 and R6 respectively. The relative proportion of each mutant in the crR6 sample over the R6 sample was computed (Figure 24c). All mutations outside of the seed sequence (13-20 bases away from the PAM) show full interference. Those sequences were used as a reference to determine if other mutations inside the seed sequence can be used to significantly disrupt interference. A normal distribution was fitted to these sequences using the fitdist function of the MASS R package. The 0.99 quantile of the fitted distribution is shown as a dotted line in Figure 24c. Figure 72 shows a histogram of the data density with fitted normal distribution (black line) and 0.99 quantile (dotted line).

<table>
<thead>
<tr>
<th>TABLE F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative abundance of PAM sequences in the crR6/R6 samples averaged over bases 1 and 5.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3rd position</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd position</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1st position</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
</tbody>
</table>
### TABLE F-continued

Relative abundance of PAM sequences in the crRd/Rd samples averaged over bases 1 and 5.

<table>
<thead>
<tr>
<th>3rd position</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>1.08</td>
<td>1.16</td>
<td>1.05</td>
<td>1.14</td>
</tr>
<tr>
<td>TAC</td>
<td>1.00</td>
<td>1.08</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>TAG</td>
<td>1.02</td>
<td>1.11</td>
<td>0.77</td>
<td>1.01</td>
</tr>
<tr>
<td>TAI</td>
<td>1.01</td>
<td>1.12</td>
<td>1.21</td>
<td>1.02</td>
</tr>
</tbody>
</table>

### TABLE G

Primers used in this study (SEQ ID NOS 60-103, respectively, in order of appearance).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>B217</td>
<td>TTCTAGACCGTATTCTGATATTACGTG- CAGTTTTGAGCTTA TGCTTTT TAA</td>
</tr>
<tr>
<td>B218</td>
<td>CTGACGATATATACGTTAAGATCTGATAG- GATATTTTGGGCAACCA TCTGAGGCAAGAC</td>
</tr>
<tr>
<td>B219</td>
<td>GGGGTGCTAGCTCTTGGATGCGGAGG</td>
</tr>
<tr>
<td>B230</td>
<td>GGCAAGGACGGAAACCCCGTATTTC</td>
</tr>
<tr>
<td>B250</td>
<td>BNNGAGCGGGAAGACATATCTGGAATTC</td>
</tr>
<tr>
<td>B251</td>
<td>BNNNTATTGGGCTCATATTGCTTG</td>
</tr>
<tr>
<td>B255</td>
<td>CTGGTACCACTGCTGGCAAGAG</td>
</tr>
<tr>
<td>B256</td>
<td>CAAAATTTCTACCTTCTCCTTGTGCTTT- CCAATAAAAACCCTCC TTA</td>
</tr>
<tr>
<td>B257</td>
<td>AGGGTGGTTTATGGGAAAGGCAAGAAGGA- GACAGTGAATAATTTTGAT ACC</td>
</tr>
<tr>
<td>B258</td>
<td>CTTACGCGGCTATATATGATCAATTTCC</td>
</tr>
<tr>
<td>B269</td>
<td>TGCTGCAATTTCTGACCCATTG</td>
</tr>
<tr>
<td>B270</td>
<td>CTGTGAAACGGGAGCTGTAATATCC- GAGCCAAAAGGGCCCCAAAG</td>
</tr>
<tr>
<td>B271</td>
<td>CTGTTGACCACTGCGAAGATCTGGAATTC- GAGCCAAAAGGGCCCCAAAG</td>
</tr>
<tr>
<td>B272</td>
<td>CTGTTGACCACTGCGAAGATCTGGAATTC- GAGCCAAAAGGGCCCCAAAG</td>
</tr>
<tr>
<td>B273</td>
<td>CTGTTGACCACTGCGAAGATCTGGAATTC- GAGCCAAAAGGGCCCCAAAG</td>
</tr>
<tr>
<td>B274</td>
<td>CTGTTGACCACTGCGAAGATCTGGAATTC- GAGCCAAAAGGGCCCCAAAG</td>
</tr>
<tr>
<td>B275</td>
<td>CTGTTGACCACTGCGAAGATCTGGAATTC- GAGCCAAAAGGGCCCCAAAG</td>
</tr>
<tr>
<td>B276</td>
<td>CTGTTGACCACTGCGAAGATCTGGAATTC- GAGCCAAAAGGGCCCCAAAG</td>
</tr>
</tbody>
</table>

### TABLE G-continued

Primers used in this study (SEQ ID NOS 60-103, respectively, in order of appearance).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>B277</td>
<td>CTTTACGCGGCAATGATGGTGGGTGGATGCAATAC- TGGCCCAAAAGCCGCA</td>
</tr>
<tr>
<td>B278</td>
<td>CTTTACGCGGCAATGATGGTGGGTGGATGCAATAC- TGGCCCAAAAGCCGCA</td>
</tr>
<tr>
<td>B279</td>
<td>CCGCTTTTTGCTCCATTTTCAG</td>
</tr>
<tr>
<td>B280</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B281</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B282</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B283</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B284</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B285</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B286</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B287</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B288</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B289</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B290</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B291</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B292</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B293</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B294</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B295</td>
<td>CAATGGCTGAATGCGCCAAATATGACAGGAGAAATCC</td>
</tr>
<tr>
<td>B296</td>
<td>gatcccccatacataacaaccagagccctgtg</td>
</tr>
<tr>
<td>B297</td>
<td>aatccacgggttgggtgtgttagag</td>
</tr>
<tr>
<td>B298</td>
<td>CATGGATCTTACCTTATACTAATCTAAAAATATG</td>
</tr>
<tr>
<td>B299</td>
<td>CATGAATTTTCTGCAAGTACTCTGATGCTG</td>
</tr>
<tr>
<td>B300</td>
<td>AAAATATTTTTTCTCATATGAAAGAAGGACTG</td>
</tr>
<tr>
<td>B301</td>
<td>AAACAAGCATCTTTTTTCTTAAATCGAAAAAAT</td>
</tr>
<tr>
<td>B302</td>
<td>AAACCAATATATTGTACGCAATGATG</td>
</tr>
<tr>
<td>B303</td>
<td>AAACCAATATATTGTACGCAATGATG</td>
</tr>
</tbody>
</table>
### Table G-continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>B304</td>
<td>AAAAACTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B305</td>
<td>AAAAAACAAAGACCTTGATGGTCTGTATATAGATGAAA</td>
</tr>
<tr>
<td>B306</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B307</td>
<td>AAAAAACTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B308</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B309</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B310</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B311</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B312</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B313</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B314</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B315</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B316</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B317</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B318</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B319</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B320</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B321</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B322</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B323</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B324</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B325</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B326</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B327</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B328</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B329</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B330</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B331</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B332</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B333</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B334</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B335</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B336</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B337</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B338</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B339</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B340</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B341</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B342</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B343</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B344</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B345</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B346</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B347</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B348</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B349</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B350</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B351</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B352</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B353</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B354</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B355</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B356</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B357</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B358</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B359</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>B360</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
</tbody>
</table>

### Table G-continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>L426</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L430</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L444</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L445</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L446</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L447</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L448</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L457</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L458</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L459</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L461</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L481</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>L488</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>W256</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>W286</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>W287</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
<tr>
<td>W288</td>
<td>AAAAAACCACTTTCTCCGTAAAGGATCTGCAAGTCTTATATTGG</td>
</tr>
</tbody>
</table>

**Notes:** The sequences represent the primers used in this study (SEQ ID Nos 68-183, respectively, in order of appearance).
### TABLE G-continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>W366</td>
<td>AATGTAAATTAAAGAACCAAGTAC-CGTTCTTTTAAACTGA TGGTGGTTGGA</td>
</tr>
<tr>
<td>W370</td>
<td>GCTTTTAACAAAGGACGTCGTCCATTATAGTTTAAAGAA</td>
</tr>
<tr>
<td>W371</td>
<td>GAAACGAGTACCGTCTTGGTTTTAAAGAAGGAA-CAGTGAAGGAGAT TTAC</td>
</tr>
<tr>
<td>W376</td>
<td>CGATTCTACCCATTGCCCTGC</td>
</tr>
<tr>
<td>W377</td>
<td>GCTTTTGACGGACGACGGAGCT- GAATGAAAGGAAAGAGGAGCC AAGAAGGAGTACCC</td>
</tr>
<tr>
<td>W391</td>
<td>TCGGTACACCCCAACACCGTACAGT-GAGGTTTTTGGGGAAGCA TGGGAGAGCTGA</td>
</tr>
<tr>
<td>W392</td>
<td>GCTCACTAGACGGGTTGTTGTTTACAG-GAGGTTTTTGGGGAAGCA TGGGAGAGCTGA</td>
</tr>
<tr>
<td>W393</td>
<td>TGTGCGACCTCCTCCTCCTCCTC</td>
</tr>
<tr>
<td>W397</td>
<td>CGATGGTGGTGTTGCTGCTCTTACGTTACCCACATCCTC</td>
</tr>
<tr>
<td>W398</td>
<td>GGGGATTAACCTCCTCGCAACACCCACAGCCCTGGCTAGTG</td>
</tr>
</tbody>
</table>

### TABLE H

Design of targeting and editing constructs used in this study (SEQ ID NOS 108, 104, 104, 105 and 106, respectively, in order of appearance).

<table>
<thead>
<tr>
<th>Targeting Constructs</th>
<th>Template DNA</th>
<th>Left PCR</th>
<th>Right PCR</th>
<th>Spacer sequence</th>
<th>PAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>bgAA R &gt; A crR6Rk</td>
<td>W256/W391</td>
<td>W392/L403</td>
<td>GCTCACTAGACGGGTTGTTGTTTACAG-GAGGTTTTTGGGGAAGCA TGGGAGAGCTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bgAA NE &gt; A crR6Rk</td>
<td>W256/W391</td>
<td>W392/L403</td>
<td>GCTCACTAGACGGGTTGTTGTTTACAG-GAGGTTTTTGGGGAAGCA TGGGAGAGCTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbgAA crR6Rk</td>
<td>W256/W391</td>
<td>W392/L403</td>
<td>GCTCACTAGACGGGTTGTTGTTTACAG-GAGGTTTTTGGGGAAGCA TGGGAGAGCTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asrta crR6Rk</td>
<td>W256/B218</td>
<td>B217/L403</td>
<td>TCTTACGACGATTTGCCAGTATCTTGACAGCAGAGTCAGA TGGGAGAGCTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ermB Stop crR6Rk</td>
<td>W256/W356</td>
<td>W357/L403</td>
<td>TTTAAGAAACCCGACGGTAGTACCGGTGTTTGACAGAAT TGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asrta AbgAA JEN51 (for Left PCR)</td>
<td>W256/W465</td>
<td>W466/W403</td>
<td>same as the ones used for TGG and JEN52 (for Right PCR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asrta and AbgAA</td>
<td>W403/W404</td>
<td>JEN56</td>
<td>W403/W404</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Editing Constructs</th>
<th>Name of resulting Primers used to verify edited genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>bgAA R &gt; A R6</td>
<td>W403/W397, W398/W404 N/A</td>
</tr>
<tr>
<td>bgAA NE &gt; A R6</td>
<td>W403/W431, W432/W333, W434/W404, W403/W404 N/A</td>
</tr>
<tr>
<td>AbgAA R6</td>
<td>B255/B256, B257/B258 N/A</td>
</tr>
<tr>
<td>Asrta R6</td>
<td>B230/W463, W464/B229 N/A</td>
</tr>
<tr>
<td>ermB Stop JEN50</td>
<td>L422/W70, L371/L426 N/A</td>
</tr>
<tr>
<td>Asrta AbgAA</td>
<td>same as the ones used for Asrta and AbgAA</td>
</tr>
</tbody>
</table>
Optimization of the Guide RNA for *Streptococcus pyogenes* Cas9 (Referred to as SpCas9)

Applicants mutated the tracrRNA and direct repeat sequences, or mutated the chimeric guide RNA to enhance the RNAs in cells.

The optimization is based on the observation that there were stretches of thymines (Ts) in the tracrRNA and guide RNA, which might lead to early transcription termination by the pol 3 promoter. Therefore, Applicants generated the following optimized sequences. Optimized tracrRNA and corresponding optimized direct repeat are presented in pairs.

Optimized tracrRNA 1 (mutation underlined):

```
GGACCCATCAAGACGATAGCAAGTATGAAATAGGCTAGTTCCGTTATC
```

Optimized direct repeat 1 (mutation underlined):

```
ACTTGAAAAGTGGGACCCAGAGTCCGTTCTTTTT
```

Optimized tracrRNA 2 (mutation underlined):

```
GGACCCATCAAGACGATAGCAAGTATGAAATAGGCTAGTTCCGTTATC
```

Optimized direct repeat 2 (mutation underlined):

```
GTTAATTAGCTATGCTTGTTGAGATGCTCCAAAAC
```

Applicants also optimized the chimeric guideRNA for optimal activity in eukaryotic cells.

Original guide RNA:

```
HNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

Optimized chimeric guide RNA sequence 1:

```
HNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

Optimized chimeric guide RNA sequence 2:

```
HNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

Optimized chimeric guide RNA sequence 3:

```
HNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

Example 7

Optimization of *Streptococcus thermophiles* 1MD-9 CRISPR1Cas9 (Referred to as St1Cas9)

Applicants designed guide chimeric RNAs as shown in FIG. 4.

The St1Cas9 guide RNAs can undergo the same type of optimization as for SpCas9 guide RNAs, by breaking the stretches of poly thymines (Ts).

Example 8

Cas9 Diversity and Mutations

The CRISPR-Cas system is an adaptive immune mechanism against invading exogenous DNA employed by diverse species across bacteria and archaea. The type II CRISPR-Cas9 system consists of a set of genes encoding proteins responsible for the “acquisition” of foreign DNA into the CRISPR locus, as well as a set of genes encoding the “execution” of the DNA cleavage mechanism; these include the DNA nuclease (Cas9), a non-coding transactivating crRNA (tracrRNA), and an array of foreign DNA-derived spacers flanked by direct repeats (crRNAs). Upon maturation by Cas9, the tracrRNA and crRNA duplex guide the Cas9 nuclease to a target DNA sequence specified by the spacer guide sequences, and mediates double-stranded breaks in the DNA near a short sequence motif in the target DNA that is required for cleavage and specific to each CRISPR-Cas system. The type II CRISPR-Cas systems are found throughout the bacterial kingdom and highly diverse in in Cas9 protein sequence and size, tracrRNA and crRNA direct repeat sequence, genome organization of these elements, and the motif requirement for target cleavage. One species may have multiple distinct CRISPR-Cas systems.

Applicants evaluated 207 putative Cas9s from bacterial species identified based on sequence homology to known Cas9s and structures orthologous to known subdomains, including the HNH endonuclease domain and the RuvC endonuclease domains [information from the Eugene Koonin and Kira Makarova]. Phylogenetic analysis based on the protein sequence conservation of this set revealed five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids) (FIGS. 39 and 40A-F).

In this example, Applicants show that the following mutations can convert SpCas9 into a nicking enzyme: D10A, E762A, H840A, N854A, N863A, D986A.
Applicants provide sequences showing where the mutation points are located within the SpCas9 gene (FIG. 41). Applicants also show that the nickases are still able to mediate homologous recombination (Assay indicated in FIG. 2). Furthermore, Applicants show that SpCas9 with these mutations (individually) do not induce double strand break (FIG. 47).

Furthermore, potential nicking mutation sites were chosen based on sequence homology between Cas9 orthologs (named original set below). The nickase mutant Cas9s were re-cloned to incorporate both N' and C'-NLS sequences as in Cong, L et al., Multiplex genome engineering using CRISPR/Cas systems, Science. 2013 Feb. 15; 339(6121):819-23. (sequences for NLS-E762A-NLS and >NLS-D986A-NLS listed below).

Nuclease and double-nicking activities for these potential nickases were tested in HEK 293FT cells as follows: co-transfection of 400 ng of nickase and 100 ng of U6-driven sgRNA (100 ng for one guide, or 50 ng each for a pair of sgRNAs) by Lipofectamine 2000 into 200,000 cells. DNAs from transfected cells were collected for SURVEYOR analysis. Nickases do not result in indel mutations when co-transfected with a single sgRNA, but do when co-transfected with a pair of appropriately offset sgRNAs. Based on data from the original D10A SpCas9 nickase, the pair of sgRNA chosen (A1/C1) for RuvC domain mutants have O'sp offset and 5' overhang for maximal cleavage.
- continued

Supplement to DNA Targeting Specificity of the RNA-Guided Cas9 Nuclease

Cell Culture and Transfection

Human embryonic kidney (HEK) cell line 293FT (Life Technologies) was maintained in Dulbecco’s modified Eagle’s Medium (DMEM) supplemented with 10% fetal bovine serum (HyClone), 2 mM glutamax (Life Technologies), 100 U/mL penicillin, and 100 μg/mL streptomycin at 37°C with 5% CO2 incubation.

293FT cells were seeded either onto 6-well plates, 24-well plates, or 96-well plates (Corning) 24 hours prior to transfection. Cells were transfected using Lipofectamine 2000 (Life Technologies) at 80–90% confluence following the manufacturer’s recommended protocol. For each well of a 6-well plate, a total of 1 μg of Cas9-sgRNA plasmid was used. For each well of a 24-well plate, a total of 500 ng Cas9-sgRNA plasmid was used unless otherwise indicated. For each well of a 96-well plate, 65 ng of Cas9 plasmid was used at a 1:1 molar ratio to the U6-sgRNA PCR product.

Human embryonic stem cell line HUES9 (Harvard Stem Cell Institute core) was maintained in feeder-free conditions on GelTrex (Life Technologies) in mTeSR medium (Stemcell Technologies) supplemented with 100 μg/mL Normocin (InvivoGen). HUES9 cells were transfected with Amaxa P3 Primary Cell 4-D Nucleofector Kit (Lonza) following the manufacturer’s protocol.

SURVEYOR Nuclease Assay for Genome Modification

293FT cells were transfected with plasmid DNA as described above. Cells were incubated at 37°C, for 72 hours post-transfection prior to genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer’s protocol.

Briefly, pelleted cells were resuspended in QuickExtract solution and incubated at 65°C for 15 minutes and 98°C for 10 minutes.

The genomic region flanking the CRISPR target site for each gene was PCR amplified (primers listed in Tables J and K), and products were purified using QiaQuick Spin Column (Qiagen) following the manufacturer’s protocol. 400 ng total of the purified PCR products were mixed with 2 μl 10× Taq DNA Polymerase PCR buffer (Enzymatics) and ultrapure water to a final volume of 20 μl, and subjected to a re-annealing process to enable heteroduplex formation: 95°C for 10 min, 95°C to 85°C ramping at ~2°C/s, 85°C to 25°C at ~0.25°C/s, and 25°C hold for 1 minute. After re-annealing, products were treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer’s recommended protocol, and analyzed on 4–20% Novex TBE poly-acrylamide gels (Life Technologies). Gels were stained with SYBR Gold DNA stain (Life Technologies) for 30 minutes and imaged with a Gel Doc gel imaging system (Bio–md). Quantification was based on relative band intensities.

Northern Blot Analysis of tracrRNA Expression in Human Cells

Northern blots were performed as previously described. Briefly, RNAs were heated to 95°C for 5 min before loading on 8% denaturing polyacrylamide gels (SequaGel, GelBiosciences). Afterweds, RNA was transferred to a pre-hybrided Hybond N+ membrane (GE Healthcare) and crosslinked with Stratagene UV Crosslinker (Stratagene). Probes were labeled with [γ-32P] ATP (Perkin Elmer) with T4 polynucleotide kinase (New England Biolabs). After washing, membrane was exposed to phosphor screen for one hour and scanned with phosphorimager (Typhoon).

Bisulfite Sequencing to Assess DNA Methylation Status

HEK 293FT cells were transfected with Cas9 as described above. Genomic DNA was isolated with the DNeasy Blood & Tissue Kit (Qiagen) and bisulfite converted with EZ DNA Methylation-Lightning Kit (Zymo Research). Bisulfite PCR was conducted using KAPA2G Robust HotStart DNA Polymerase (KAPA Biosystems) with primers designed using the Bisulfite Primer Seeker (Zymo Research, Tables J and K). Resulting PCR amplicons were gel-purified, digested with EcoRI and HindIII, and ligated into a pUC19 backbone prior
to transformation. Individual clones were then Sanger sequenced to assess DNA methylation status.

In Vitro Transcription and Cleavage Assay

HEK 293FT cells were transfected with Cas9 as described above. Whole cell lysates were then prepared with a lysis buffer (20 mM HEPES, 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 5% glycerol, 0.1% Triton X-100) supplemented with Protease Inhibitor Cocktail (Roche). T7-driven sgRNA was in vitro transcribed using custom oligos (Example 10) and HiScribe T7 In Vitro Transcription Kit (NEB), following the manufacturer’s recommended protocol. To prepare methylated target sites, pUC19 plasmid was methylated by M.SsI and then linearized by NheI. The in vitro cleavage assay was performed as follows: for a 20 μl cleavage reaction, 10 μl of cell lysate with incubated with 2 μl cleavage buffer (100 mM HEPES, 500 mM KCl, 25 mM MgCl2, 5 mM DTT, 25% glycerol), the in vitro transcribed RNA; and 300 ng pUC19 plasmid DNA.

Deep Sequencing to Assess Targeting Specificity

HEK 293FT cells plated in 96-well plates were transfected with Cas9 plasmid DNA and single guide RNA (sgRNA) PCR cassette 72 hours prior to genomic DNA extraction (FIG. 72). The genomic region flanking the CRISPR target site for each gene was amplified (FIG. 74, FIG. 80, Example 10) by a fusion PCR method to attach the Illumina P5 adapters as well as unique sample-specific barcodes to the target amplicons (schematically described in FIG. 73). PCR products were purified using Exonuclease 96-well Filter Plates (Epoch Life Sciences) following the manufacturer’s recommended protocol.

Barcoded and purified DNA samples were quantified by Quant-iT PicoGreen dsDNA Assay Kit or Qubit 2.0 Fluorometer (Life Technologies) and pooled in an equimolar ratio. Sequencing libraries were then deep sequenced with the Illumina MiSeq Personal Sequencer (Life Technologies).

Sequencing Data Analysis and Indel Detection

MiSeq reads were filtered by requiring an average Phred quality (Q score) of at least 23, as well as perfect sequence matches to barcodes and amplicon forward primers. Reads from on- and off-target loci were analyzed by first performing Smith-Waterman alignment against amplicon sequences that included 50 nucleotides upstream and downstream of the target site (a total of 120 bp). Alignments, meanwhile, were analyzed for indels from 5 nucleotides upstream to 5 nucleotides downstream of the target site (a total of 30 bp). Analyzed target regions were discarded if part of their alignment fell outside the MiSeq read itself, or if matched base-pairs comprised less than 85% of their total length.

Negative controls for each sample provided a gauge for the inclusion or exclusion of indels as putative cutting events. For each sample, an indel was counted only if its quality score exceeded μσ, where μ was the mean quality-score of the negative control corresponding to that sample and σ was the standard deviation of this mean. This yielded whole target-region indel rates for both negative controls and their corresponding samples. Using the negative control’s per-target-region-per-read error rate, q, the sample’s observed indel count n, and its read-count R, a maximum-likelihood estimate for the fraction of reads having target-regions with true indels, p, was derived by applying a binomial error model, as follows.

Letting the (unknown) number of reads in a sample having target regions incorrectly counted as having at least 1 indel be E, we can write (without making any assumptions about the number of true indels)

\[
\text{Prob}(E | p) = \binom{R(1-p)}{E} (1-q)^{R(1-p)-E}
\]

since R(1-p) is the number of reads having target-regions with no true indels. Meanwhile, because the number of reads observed to have indels is n, n=B+R, in other words the number of reads having target-regions with errors but no true indels plus the number of reads whose target-regions correctly have indels. We can then re-write the above

\[
\text{Prob}(E | p) = \text{Prob}(n = E + R | p) = \binom{R(1-p)}{n-R} (1-q)^{R(1-p)-n}
\]

Taking all values of the frequency of target-regions with true indels P is to be equally probable a priori, \text{Prob}(n/p) \times \text{Prob}(p/n). The maximum-likelihood estimate (MLE) for the frequency of target regions with true indels was therefore set as the value of p that maximized \text{Prob}(n/p). This was evaluated numerically.

In order to place error bounds on the true-indel read frequencies in the sequencing libraries themselves, Wilson score intervals (2) were calculated for each sample, given the MLE-estimate for true-indel target-regions, R, and the number of reads R. Explicitly, the lower bound and upper bound u were calculated as

\[
l = \left( \frac{R}{u} - z \cdot z - \frac{z^2}{4} \right) / (R + z^2)
\]

\[
u = \left( \frac{R}{v} + z \cdot z + \frac{z^2}{4} \right) / (R + z^2)
\]

where z, the standard score for the confidence required in normal distribution of variance 1, was set to 1.96, meaning a confidence of 95%. The maximum upper bounds and minimum lower bounds for each biological replicate are listed in FIGS. 80-83.

qRT-PCR Analysis of Relative Cas9 and sgRNA Expression

293FT cells plated in 24-well plates were transfected as described above. 72 hours post-transfection, total RNA was harvested with miRNAasy Micro Kit (Qiagen). Reverse-strand synthesis for sgRNAs was performed with qScript Flex cDNA kit (VWR) and custom first-strand synthesis primers (Tables J and K). qPCR analysis was performed with Fast SYBR Green Master Mix (Life Technologies) and custom primers (Tables J and K), using GAPDH as an endogenous control. Relative quantification was calculated by the ΔΔCT method.

Table I: Target site sequences. Tested target sites for S. pyogenes type II CRISPR system with the requisite PAM. Cells were transfected with Cas9 and either crRNA-tracrRNA or chimeric sgRNA for each target.
### TABLE 1

**Target site sequences.** Tested target sites for *S. pyogenes* type II CRISPR system with the requisite PAM. Cells were transfected with Cas9 and either crRNA-tracrRNA or chimeric sgRNA for each target.

<table>
<thead>
<tr>
<th>Target site ID</th>
<th>genomic target (5' to 3')</th>
<th>Target site sequence (5' to 3')</th>
<th>PAM strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 EMK1</td>
<td>GTCACCTTCCATGACTAGGG</td>
<td>TGG +</td>
<td></td>
</tr>
<tr>
<td>2 EMK1</td>
<td>GAGATCGATGCTCCCACTCATG</td>
<td>TGG -</td>
<td></td>
</tr>
<tr>
<td>3 EMK1</td>
<td>GAGTCCGACAGAAAGGAAGGA</td>
<td>GGG +</td>
<td></td>
</tr>
<tr>
<td>6 EMK1</td>
<td>GCACCACGCGTTGAGCTGAT</td>
<td>GGG -</td>
<td></td>
</tr>
<tr>
<td>10 EMK1</td>
<td>GGCGACACAGATGAAGAAACTG</td>
<td>AGG -</td>
<td></td>
</tr>
<tr>
<td>11 EMK1</td>
<td>GTCAAAACGCGCAAGCTGGS</td>
<td>AGG +</td>
<td></td>
</tr>
<tr>
<td>12 EMK1</td>
<td>GCAGCAAGCTCAGAGAGAAGA</td>
<td>GGG +</td>
<td></td>
</tr>
<tr>
<td>13 EMK1</td>
<td>GGACCCCTCTCTCTCTCTCTCT</td>
<td>CCG -</td>
<td></td>
</tr>
<tr>
<td>14 EMK1</td>
<td>GGCGCAACCAACCAACCCCAAGA</td>
<td>GGG +</td>
<td></td>
</tr>
<tr>
<td>15 EMK1</td>
<td>GCTCCCATCAGATCAACCCG</td>
<td>TGG +</td>
<td></td>
</tr>
<tr>
<td>16 EMK1</td>
<td>GTGCCCCCTCCGAGCAAGGCGAC</td>
<td>AGG +</td>
<td></td>
</tr>
<tr>
<td>17 EMK1</td>
<td>GCCAGACTCGTCTGTGCTG</td>
<td>TGG +</td>
<td></td>
</tr>
<tr>
<td>18 EMK1</td>
<td>GCCCCCTCGGTGCGCCCAAGG</td>
<td>TGG +</td>
<td></td>
</tr>
<tr>
<td>19 EMK1</td>
<td>GAAGCCGACACAGATCAACCTT</td>
<td>GGG -</td>
<td></td>
</tr>
<tr>
<td>20 EMK1</td>
<td>GCCCTCAGAAAACCTGCGCC</td>
<td>AGG -</td>
<td></td>
</tr>
<tr>
<td>4 PVALB</td>
<td>GGCGCCGAGATGGCTGATC</td>
<td>AGG +</td>
<td></td>
</tr>
<tr>
<td>5 PVALB</td>
<td>GTGGCGAGAGGCGGCGCAAGA</td>
<td>TGG +</td>
<td></td>
</tr>
<tr>
<td>1 SERPINB5</td>
<td>GAAGTCCCGGCGAAGCGGCGG</td>
<td>GGG +</td>
<td></td>
</tr>
<tr>
<td>2 SERPINB5</td>
<td>GAAGTCCCGGCGAAGCGGCGG</td>
<td>CCG +</td>
<td></td>
</tr>
<tr>
<td>3 SERPINB5</td>
<td>GAAGAAGCTCGCGCAGCCGAGG</td>
<td>CCG +</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE J

<table>
<thead>
<tr>
<th>Primer name</th>
<th>genomic target</th>
<th>primer sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp-EMK1-F1</td>
<td>EMK1</td>
<td>AAAACCGCCCTTCTCTCTTGCGC (SEQ ID NO: 33)</td>
</tr>
<tr>
<td>Sp-EMK1-R1</td>
<td>EMK1</td>
<td>GGGAATGGGCAACACGGAGA (SEQ ID NO: 37)</td>
</tr>
<tr>
<td>Sp-EMK1-P2</td>
<td>EMK1</td>
<td>CCATCCCCCTCTGATGTG (SEQ ID NO: 217)</td>
</tr>
<tr>
<td>Sp-EMK1-R2</td>
<td>EMK1</td>
<td>GGGATGGGCAACACGGAGA (SEQ ID NO: 218)</td>
</tr>
<tr>
<td>Sp-PVALB-F</td>
<td>PVALB</td>
<td>CTCGAAAGGCAATGCTGCAG (SEQ ID NO: 38)</td>
</tr>
<tr>
<td>Sp-PVALB-R</td>
<td>PVALB</td>
<td>GGCAGCAAGCTCCCTCTGCTC (SEQ ID NO: 39)</td>
</tr>
</tbody>
</table>

**qRT-PCR for Cas9 and sgRNA expression**

<table>
<thead>
<tr>
<th>primer name</th>
<th>primer sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>sgRNA reverse-strand synthesis</td>
<td>AAGACCGAATGCTGCGAC (SEQ ID NO: 219)</td>
</tr>
<tr>
<td>EMK1.1 sgRNA</td>
<td>TCACCTCCAATGACTAGGG (SEQ ID NO: 220)</td>
</tr>
<tr>
<td>EMK1.1 sgRNA</td>
<td>CAAGTTGAAACGGACTAGCCT (SEQ ID NO: 221)</td>
</tr>
<tr>
<td>EMK1.3 sgRNA</td>
<td>AGTCCCGAAGAAGAAGAAGT (SEQ ID NO: 222)</td>
</tr>
<tr>
<td>EMK1.3 sgRNA</td>
<td>TTCAAGTGGTAACGCACTAGACT (SEQ ID NO: 223)</td>
</tr>
<tr>
<td>Cas9 qPCR R</td>
<td>AAACACGAATGCTGCGAC (SEQ ID NO: 224)</td>
</tr>
<tr>
<td>Cas9 qPCR R</td>
<td>TCAATCCGCAGACTAGCCT (SEQ ID NO: 225)</td>
</tr>
<tr>
<td>GAPDH qPCR R</td>
<td>TCAAAAATCAATGCGGGA (SEQ ID NO: 226)</td>
</tr>
<tr>
<td>GAPDH qPCR R</td>
<td>TGATACCCCTTGGTCACTC (SEQ ID NO: 227)</td>
</tr>
</tbody>
</table>

**Bisulfite PCR and sequencing**

<table>
<thead>
<tr>
<th>primer name</th>
<th>primer sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisulfite PCR R (SERPINB locus)</td>
<td>GAGAAATTTTTTTTTTTTTTTTGAATTTTTTGAG (SEQ ID NO: 229)</td>
</tr>
<tr>
<td>Bisulfite PCR R (SERPINB locus)</td>
<td>GAGAAATTTTTTTTTTTTTTTTGAATTTTTTGAG (SEQ ID NO: 229)</td>
</tr>
<tr>
<td>pUC19 sequencing</td>
<td>CAGGAAACCCCTATGAC (SEQ ID NO: 230)</td>
</tr>
</tbody>
</table>
### Table K

Sequences for primers to test sgRNA architecture. Primers hybridize to the reverse strand of the Us promoter unless otherwise indicated. The Us priming site is in italics, the guide sequence is indicated as a stretch of Us, the direct repeat sequence is highlighted in bold, and the tracrRNA sequence underlined. The secondary structure of each sgRNA architecture is shown in Fig. 49.

<table>
<thead>
<tr>
<th>primer name</th>
<th>primer sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>06-Forward</td>
<td>GCCTCTAGAATTCACCATGATGCGAAGCCTATATTTCGCCGAT TCC (SEQ ID NO: 231)</td>
</tr>
<tr>
<td>I: sgRNA(DR +12, tracrRNA +85)</td>
<td>ACCACCTGGGGGACCCGACACTGCGTCCCTT GGACCTGACACGTGACCTGACTT GCC TCC (SEQ ID NO: 232)</td>
</tr>
<tr>
<td>II: sgRNA(DR +12, tracrRNA +85)</td>
<td>ACCACCTGGGGGACCCGACACTGCGTCCCTT GGACCTGACACGTGACCTGACTT GCC TCC (SEQ ID NO: 233)</td>
</tr>
<tr>
<td>mut2</td>
<td>ACCACCTGGGGGACCCGACACTGCGTCCCTT GGACCTGACACGTGACCTGACTT GCC TCC (SEQ ID NO: 234)</td>
</tr>
<tr>
<td>III: sgRNA(DR +22, tracrRNA +85)</td>
<td>ACCACCTGGGGGACCCGACACTGCGTCCCTT GGACCTGACACGTGACCTGACTT GCC TCC (SEQ ID NO: 235)</td>
</tr>
<tr>
<td>IV: sgRNA(DR +22, tracrRNA +85)</td>
<td>ACCACCTGGGGGACCCGACACTGCGTCCCTT GGACCTGACACGTGACCTGACTT GCC TCC (SEQ ID NO: 236)</td>
</tr>
</tbody>
</table>

### Table L

Target sites with alternate PAMs for testing PAM specificity of Cas9. All target sites for PAM specificity testing are found within the human BMG1 locus.

<table>
<thead>
<tr>
<th>Target site sequence (5' to 3')</th>
<th>PAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGGCCGCCGATTATGCTTACCAAG (SEQ ID NO: 236)</td>
<td>NAA</td>
</tr>
<tr>
<td>ACATCAAGCGGCTGGCGCAT (SEQ ID NO: 237)</td>
<td>NAT</td>
</tr>
<tr>
<td>AAGGTGGGTGTTCCAGAACAG (SEQ ID NO: 238)</td>
<td>NAC</td>
</tr>
<tr>
<td>CATGACACATACCGGCAG (SEQ ID NO: 239)</td>
<td>NAG</td>
</tr>
<tr>
<td>AACGGAGCAGAGCTGCGAG (SEQ ID NO: 240)</td>
<td>NTA</td>
</tr>
<tr>
<td>GGGAGAAGGGGAGTGAAGA (SEQ ID NO: 241)</td>
<td>NTT</td>
</tr>
<tr>
<td>GGTGGGTGTTCCAGAACAG (SEQ ID NO: 242)</td>
<td>NTC</td>
</tr>
<tr>
<td>AAGCAGGGCAAGACTACA (SEQ ID NO: 243)</td>
<td>NTG</td>
</tr>
<tr>
<td>TTTCAAAACGGGCAAGCAGAA (SEQ ID NO: 244)</td>
<td>NGA</td>
</tr>
<tr>
<td>CTGAGGAGGCTGAAAGAAG (SEQ ID NO: 245)</td>
<td>NCT</td>
</tr>
<tr>
<td>CATGCAACCGGGGCGGGAAC (SEQ ID NO: 246)</td>
<td>NCC</td>
</tr>
<tr>
<td>CAGAGGAGGCTGAAAGAAG (SEQ ID NO: 247)</td>
<td>NCG</td>
</tr>
<tr>
<td>CATAGAGGAGGCTGAAAGAAG (SEQ ID NO: 248)</td>
<td>NGA</td>
</tr>
<tr>
<td>GCAGAAGGCTGAAAGAAG (SEQ ID NO: 249)</td>
<td>NTT</td>
</tr>
<tr>
<td>CTCCTCGCTGCGGCAG (SEQ ID NO: 250)</td>
<td>NCG</td>
</tr>
</tbody>
</table>

### Supplementary Sequences

All sequences are in the 5' to 3' direction. For U6 transcription, the string of underlined Ts serve as the transcriptional terminator.

> U6-short tracrRNA (Streptococcus pyogenes SP370) (SEQ ID NO: 40)

gaggggctattttgccatatttttcattttttcatatttttcagtaagggc
tgttagagatattgtgatatttttgctgtaaccaaaagatatttag
taaattattttttattttttttattttttttattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Example 11
Oligo-Mediated Cas9-Induced Homologous Recombination

The oligo homologous recombination test is a comparison of efficiency across different Cas9 variants and different HR template (oligo vs. plasmid).
likely result is a frameshift mutation that would knockout the
gene. The targeting strategy involved finding proto-spacers in
the exons of the gene that had a PAM sequence, NGG, and
was unique in the genome. Preference was given to proto-
spacers in the first exon, which would be most deleterious to
the gene.

Each gRNA was validated in the mouse cell line, Neuro-
N2a, by liposomal transient co-transfection with hSpCas9. 72
hours post-transfection genomic DNA was purified using
QuickExtract DNA from Epicentre. PCR was performed to
amplify the locus of interest. Subsequently the SURVEYOR
Mutation Detection Kit from Transgenomics was followed.
The SURVEYOR results for each gRNA and respective con-
trols are shown in Figure A1. A positive SURVEYOR result is
one large band corresponding to the genomic PCR and two
smaller bands that are the product of the SURVEYOR nuclease
making a double-strand break at the site of a muta-
tion. The average cutting efficiency of each gRNA was also
determined for each gRNA. The gRNA that was chosen for
injection was the highest efficiency gRNA that was unique
within the genome.

RNA (hSpCas9 gRNA RNA) was injected into the pro-
nucleus of a zygote and later transplanted into a foster mother.
Mothers were allowed to go full term and pups were sampled
by tail snip 10 days postnatal. DNA was extracted and used as
a template for PCR, which was then processed by SUR-
VEYOR. Additionally, PCR products were sent for sequenc-
ing. Animals that were detected as being positive in either the
SURVEYOR assay or PCR sequencing would have their
genomic PCR products cloned into a pUC19 vector and
sequenced to determine putative mutations from each allele.

So far, mice pups from the Chd8 targeting experiment have
been fully processed up to the point of allele sequencing. The
SURVEYOR results for 38 live pups (lanes 1-38) 1 dead pup (lane
39) and 1 wild-type pup for comparison (lane 40) are shown
in Figure A2. Pups 1-19 were injected with gRNA Chd8.2 and
pups 20-38 were injected with gRNA Chd8.3. Of the 38 live
pups, 13 were positive for a mutation. The one dead pup also
had a mutation. There was no mutation detected in the wild-
type sample. Genomic PCR sequencing was consistent with the
SURVEYOR assay findings.

Example 13
CRISPR/Cas-Mediated Transcriptional Modulation

FIG. 67 depicts a design of the CRISPR-TF (Transcrip-
tional Factor) with transcriptional activation activity. The chimeric
RNA is expressed by U6 promoter, while a human-codon-
optimized, double-mutant version of the Cas9 protein
(hSpCas9m), operably linked to triple NLS and a VP64 func-
tional domain is expressed by a EF1a promoter. The double
mutations, D10A and H840A, renders the cas9 protein unable
to introduce any cleavage but maintained its capacity to bind
to target DNA when guided by the chimeric RNA.

FIG. 68 depicts transcriptional activation of the human
SOX2 gene with CRISPR-TF system (Chimeric RNA and the
Cas9-NLS-VP64 fusion protein). 293FT cells were trans-
fected with plasmids bearing two components: (1) U6-driven
different chimeric RNAs targeting 20-bp sequences within or
around the human SOX2 genomic locus, and (2) EF 1a-driven
hSpCas9m (double mutant)-NLS-VP64 fusion protein. 96
hours post transfection, 293FT cells were harvested and the
level of activation is measured by the induction of mRNA
expression using a qRT-PCR assay. All expression levels are
normalized against the control group (grey bar), which
represents results from cells transfected with the CRISPR-TF
backbone plasmid without chimeric RNA. The qRT-PCR
probes used for detecting the SOX2 mRNA is Taqman
Human Gene Expression Assay (Life Technologies). All
experiments represents data from 3 biological replicates, n=3,
error bars show s.e.m.

Example 14
NLS: Cas9-NLS

293FT cells were transfected with plasmid containing two
components: (1) EF 1a promoter driving the expression of
Cas9 (wild-type human-codon-optimized Sp Cas9) with different
NLS designs (2) U6 promoter driving the same chimeric
RNA targeting human EMX1 locus.

Cells were collect at 72h time point post transfection, and
then extracted with 50 μl of the QuickExtract genomic DNA
extraction solution following manufacturer’s protocol. Target
EMX1 genomic DNA were PCR amplified and then Gel-
purify with 1% agarose gel. Genomic PCR product were re-
anneal and subjected to the Surveyor assay following
manufacturer’s protocol. The genomic cleavage efficiency of
different constructs were measured using SDS-PAGE on a
4-12% TBE-PAGE gel (Life Technologies), analyzed and
quantified with ImagerLab (Bio-rad) software, all following
manufacturer’s protocol.

FIG. 69 depicts a design of different Cas9 NLS constructs. All
Cas9 were the human-codon-optimized version of the Sp
Cas9. NLS sequences are linked to the cas9 gene at either
N-terminus or C-terminus. All Cas9 variants with different
NLS designs were cloned into a backbone vector containing
so it is driven by EF 1a promoter. On the same vector there is
a chimeric RNA targeting human EMX1 locus driven by U6
promoter, together forming a two-component system.

Table M

| Cas9 NLS Design Test Results: Quantification of genomic cleavage | Percentage Genome Cleavage as measured by Surveyor assay |
| --- | --- | --- | --- |
| Biological Replicate 1 (%) | Biological Replicate 2 (%) | Biological Replicate 3 (%) | Average (%) | Error (S.E.M., standard error of the mean) |
| Cas9 (No NLS) | 2.50 | 3.30 | 2.73 | 2.84 | 0.24 |
| Cas9 with N-term NLS | 7.61 | 6.29 | 5.46 | 6.45 | 0.63 |
| Cas9 with C-term NLS | 5.75 | 4.86 | 4.70 | 5.10 | 0.33 |
| Cas9 with Double (N-term and C-term) NLS | 9.68 | 9.85 | 7.78 | 8.90 | 0.60 |
FIG. 70 depicts the efficiency of genomic cleavage induced by Cas9 variants bearing different NLS designs. The percentage indicate the portion of human EMX1 genomic DNA that were cleaved by each construct. All experiments are from 3 biological replicates. n=3, error indicates S.E.M.

Example 15

**Engineering of Microalgae Using Cas9**

**Methods of Delivering Cas9**

**Method 1:** Applicants deliver Cas9 and guide RNA using a vector that expresses Cas9 under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2-tubulin.

**Method 2:** Applicants deliver Cas9 and T7 polymerases using vectors that expresses Cas9 under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2-tubulin. Guide RNA will be delivered using a vector containing T7 polymerase driving the guide RNA.

**Method 3:** Applicants deliver Cas9 mRNA and in vitro transcribed guide RNA to algae cells. RNA can be in vitro transcribed. Cas9 mRNA will consist of the coding region for Cas9 as well as 3’UTR from Cop1 to ensure stabilization of the Cas9 mRNA.

For Homologous recombination, Applicants provide an additional homology directed repair template.

Sequence for a cassette driving the expression of Cas9 under the control of beta-2 tubulin promoter, followed by the 3’UTR of Cop1.
Sequence for a cassette driving the expression of T7 polynucleotide under the control of Cop1:

(SEQ ID NO: 278)

TCTTTTCTGCTAGACACTTCCGACAGACAGATGAAAGGGCGCGCTGAGA
CGGGCTCCGCGCGCTGGAGAAGACAGGAGAAAGAGCTGAGAA
CTCCCAGGGGCTGAGAAAGGCAGCGCTCCGCAAGGAGGACAGCTG
TTTAAAATAACCGGCCCCCGACTGCAACAGCTTACAGGACGCTCAG
GCAATATTCAACACCTAGTACTACACTTTTCATCAAGGACAGCTGCTGCTGAGAGACGCG
AGATACACAAACACAGCTGGAGAATAGACTGGGAGCTCGAGG
TAGAACGAGAGATCACTTCCCGGAGGAAATGAGATGATCACCCTGCTAG
GTCAGAAGCTGGGATATCCTATAGACGGGAGTTGAATGTTTCTCCGAGTGAAGTG
GGGATGATCAGTCCAGGCTGAGAGCTTATCCGAGTGAAGTG
ATGAGATGACTCTTGCCCTGCTGAGAGCTTATCCGAGTGAAGTG
AGGCGGAGTAATGTGCTTGAGGAGTGAAGTG
ATGAGATGACTCTTGCCCTGCTGAGAGCTTATCCGAGTGAAGTG
AGGCGGAGTAATGTGCTTGAGGAGTGAAGTG
Sequence of guide RNA driven by the T7 promoter (T7 promoter; Ns represent target sequence):

```
se\n
```

Gene delivery:

*Chlamydomonas reinhardtii* strain CC-124 and CC-125 from the *Chlamydomonas* Resource Center will be used for electroporation. Electroporation follows standard recommended protocol from the GeneArt *Chlamydomonas* Engineering kit.

Also, Applicants generate a line of *Chlamydomonas reinhardtii* that expresses Cas9 constitutively. This can be done by using pChlamyl (linearized using PvuI) and selecting for hygromycin resistant colonies. Sequence for pChlamyl containing Cas9 is below. In this way to achieve gene knockout one simply needs to deliver RNA for the guideRNA. For homologous recombination Applicants deliver guideRNA as well as a linearized homologous recombination template.

pChlamyl-Cas9:

```
```

```
```
\[ \text{continued} \]

\[ \text{continued} \]

For all modified \textit{Chlamydomonas} reinhardtii cells, Applicants used PCR, SURVEYOR nuclease assay, and DNA sequencing to verify successful modification.

\textbf{Example 16}

Use of Cas9 as a Transcriptional Repressor in Bacteria

The ability to artificially control transcription is essential both to the study of gene function and to the construction of synthetic gene networks with desired properties. Applicants describe here the use of the RNA-guided Cas9 protein as a programmable transcriptional repressor.

Applicants have previously demonstrated how the Cas9 protein of \textit{Streptococcus pyogenes} SF370 can be used to direct genome editing in \textit{Streptococcus pneumoniae}. In this study Applicants engineered the crRNA strain containing a minimal CRISPR system, consisting of cas9, the tracrRNA and a repeat. The D10A-H840 mutations were introduced into cas9 in this strain, giving strain crRNA**. Four spacers targeting different positions of the bagA\(\beta\)-galactosidase gene promoter were cloned in the CRISPR array carried by the previously described pDB98 plasmid. Applicants observed a X to Y fold reduction in \(\beta\)-galactosidase activity depending on the targeted position, demonstrating the potential of Cas9 as a programmable repressor (FIG. 73).

To achieve Cas9** repression in \textit{Escherichia coli} a green fluorescence protein (GFP) reporter plasmid (pDB127) was constructed to express the gfpmut2 gene from a constitutive promoter. The promoter was designed to carry several NPP PAMS on both strands, to measure the effect of Cas9** binding at various positions. Applicants introduced the D10A-H840 mutations into pCas9, a plasmid designed carrying the tracrRNA, cas9 and a minimal CRISPR array designed for the easy cloning of new spacers. Twenty-two different spacers were designed to target different regions of the gfpmut2 promoter and open reading frame. An approximately 20-fold reduction of fluorescence of was observed upon targeting regions overlapping or adjacent to the \(-35\) and \(-10\) promoter elements and to the Shine-Dalgarno sequence. Targets on both strands showed similar repression levels. These results suggest that the binding of Cas9** to any position of the promoter region prevents transcription initiation, presumably through steric inhibition of RNAP binding.

To determine whether Cas9** could prevent transcription elongation, Applicants directed it to the reading frame of gfpmut2. A reduction in fluorescence was observed both when the coding and non-coding strands where targeted, suggesting that Cas9 binding is actually strong enough to represent an obstacle to the running RNAP. However, while a 40% reduction in expression was observed when the coding strand was the target, a 20-fold reduction was observed for the non-coding strand (FIG. 21b, compare T9, T10 and T11 to B9, B10 and B11). To directly determine the effects of Cas9** binding on transcription, Applicants extracted RNA from strains carrying either the T5, T10, B10 or a control construct that does not target pDB127 and subjected it to Northern blot analysis using either a probe binding before (B477) or after (B510) the B10 and T10 target sites. Consistent with Applicants' fluorescence methods, no gfpmut2 transcription was detected when Cas9** was directed to the pro-
motor region (15 target) and a transcription was observed after the targeting of the T10 region. Interestingly, a smaller transcript was observed with the B477 probe. This band corresponds to the expected size of a transcript that would be interrupted by Cas9**, and is a direct indication of a transcriptional termination caused by dRNA::Cas9** binding to the coding strand. Surprisingly, Applicants detected no transcript when the non-coding strand was targeted (B10). Since Cas9** binding to the B10 region is unlikely to interfere with transcription initiation, this result suggests that the mRNA was degraded. DgRNA::Cas9 was shown to bind ssRNA in vitro. Applicants speculate that binding may trigger degradation of the mRNA by host nucleases. Indeed, ribosome stalling can induce cleavage on the translated mRNA in E. coli.

Some applications require a precise tuning gene expression rather than its complete repression. Applicants sought to achieve intermediate repression levels through the introduction of mismatches that will weaken the crRNA/target interactions. Applicants created a series of spacers based on the B1, T5 and B10 constructs with increasing numbers of mutations in the 5’ end of the crRNA. Up to 8 mutations in B1 and T5 did not affect the repression level, and a progressive increase in fluorescence was observed for additional mutations.

The observed repression with only an 8 nt match between the crRNA and its target raises the question of off-targeting effects of the use of Cas9** as a transcriptional regulator. Since a good PAM (NGG) is also required for Cas9 binding, the number of nucleotides to match to obtain some level of repression is 10. A 10 nt match occurs randomly once every ~1 Mbp, and such sites are thus likely to be found even in small bacterial genomes. However, to effectively repress transcription, such site needs to be in the promoter region of gene, which makes off-targeting much less likely. Applicants also showed that gene expression can be affected if the non-coding strand of a gene is targeted. For this to happen, a random target would have to be in the right orientation, but such events relatively more likely to happen. As a matter of fact, during the course of this study Applicants were unable to construct one of the designed spacer on pCas9**. Applicants later found this spacer showed a 12 bp match next to a good PAM in the essential murC gene. Such off-targeting could easily be avoided by a systematic blast of the designed spacers.

While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various analogues to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

REFERENCES


**SEQUENCE LISTING**

<160> NUMBER OF SEQ ID NOS: 531

<210> SEQ ID NO 1
<211> LENGTH: 7
<212> TYPe: PRT
<213> ORGANISM: Simian virus 40
<400> SEQUENCE: 1

Pro Lys Lys Lys Arg Lys Val 1 5

<210> SEQ ID NO 2
<211> LENGTH: 16
<212> TYPe: PRT
<213> ORGANISM: Unknown
<220> FEATURE: Other Information
<223> OTHER INFORMATION: Description of Unknown: Nucleosolamin bipartite NLS sequence

<400> SEQUENCE: 2

Lys Arg Pro Ala Ala Thr Lys Ala Gly Gln Ala Lys Lys Lys Lys 1 5 10 15

<210> SEQ ID NO 3
<211> LENGTH: 9
<212> TYPe: PRT

<210> SEQ ID NO 4
<211> LENGTH: 16
<212> TYPe: PRT
<213> ORGANISM: Unknown
<220> FEATURE: Other Information
<223> OTHER INFORMATION: Description of Unknown: Nucleosolamin bipartite NLS sequence

<400> SEQUENCE: 3

Lys Arg Pro Ala Ala Thr Lys Ala Gly Gln Ala Lys Lys Lys Lys 1 5 10 15

<210> SEQ ID NO 5
<211> LENGTH: 8
<212> TYPe: PRT

<210> SEQ ID NO 6
<211> LENGTH: 16
<212> TYPe: PRT
<213> ORGANISM: Unknown
<220> FEATURE: Other Information
<223> OTHER INFORMATION: Description of Unknown: Nucleosolamin bipartite NLS sequence

<400> SEQUENCE: 4

Lys Arg Pro Ala Ala Thr Lys Ala Gly Gln Ala Lys Lys Lys Lys 1 5 10 15

<210> SEQ ID NO 7
<211> LENGTH: 7
<212> TYPe: PRT

<210> SEQ ID NO 8
<211> LENGTH: 16
<212> TYPe: PRT
<213> ORGANISM: Unknown
<220> FEATURE: Other Information
<223> OTHER INFORMATION: Description of Unknown: Nucleosolamin bipartite NLS sequence

<400> SEQUENCE: 5

Lys Arg Pro Ala Ala Thr Lys Ala Gly Gln Ala Lys Lys Lys Lys 1 5 10 15

<210> SEQ ID NO 9
<211> LENGTH: 8
<212> TYPe: PRT

<210> SEQ ID NO 10
<211> LENGTH: 16
<212> TYPe: PRT
<213> ORGANISM: Unknown
<220> FEATURE: Other Information
<223> OTHER INFORMATION: Description of Unknown: Nucleosolamin bipartite NLS sequence

<400> SEQUENCE: 6

Lys Arg Pro Ala Ala Thr Lys Ala Gly Gln Ala Lys Lys Lys Lys 1 5 10 15

<210> SEQ ID NO 11
<211> LENGTH: 7
<212> TYPe: PRT

<210> SEQ ID NO 12
<211> LENGTH: 16
<212> TYPe: PRT
<213> ORGANISM: Unknown
<220> FEATURE: Other Information
<223> OTHER INFORMATION: Description of Unknown: Nucleosolamin bipartite NLS sequence

<400> SEQUENCE: 7

Lys Arg Pro Ala Ala Thr Lys Ala Gly Gln Ala Lys Lys Lys Lys 1 5 10 15
<213> ORGANISM: Unknown
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Unknown:
 C-myc NLS sequence

<400> SEQUENCE: 3
Pro Ala Ala Lys Arg Val Lys Leu Asp
1 5

<210> SEQ ID NO 4
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Description of Unknown:
 C-myc NLS sequence

<400> SEQUENCE: 4
Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1 5 10

<210> SEQ ID NO 5
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 5
Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15
Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gln Tyr Phe Ala Lys Pro
20 25 30
Arg Asn Gln Gly Gly Tyr
35

<210> SEQ ID NO 6
<211> LENGTH: 42
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Description of Unknown:
 IIB domain from importin-alpha sequence

<400> SEQUENCE: 6
Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15
Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Asn Lys Lys
20 25 30
Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
35 40

<210> SEQ ID NO 7
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Description of Unknown:
 Myoma T protein sequence

<400> SEQUENCE: 7
Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> SEQ ID NO 8
<211> LENGTH: 8
<212> TYPE: PRT
ORGANISM: Unknown

OTHER INFORMATION: Description of Unknown:
Nycma T protein sequence

SEQUENCE: 8
Pro Pro Lys Lys Ala Arg Glu Asp
1 5

SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: Homo sapiens
SEQUENCE: 9
Pro Gln Pro Lys Lys Lys Pro Leu
1 5

SEQ ID NO 10
LENGTH: 12
TYPE: PRT
ORGANISM: Mus musculus
SEQUENCE: 10
Ser Ala Leu Ile Lys Lys Lys Lys Met Ala Pro
1 5 10

SEQ ID NO 11
LENGTH: 5
TYPE: PRT
ORGANISM: Influenza virus
SEQUENCE: 11
Asp Arg Leu Arg Arg
1 5

SEQ ID NO 12
LENGTH: 7
TYPE: PRT
ORGANISM: Influenza virus
SEQUENCE: 12
Pro Lys Gln Lys Lys Arg Lys
1 5

SEQ ID NO 13
LENGTH: 10
TYPE: PRT
ORGANISM: Hepatitis delta virus
SEQUENCE: 13
Arg Lys Leu Lys Lys Ile Lys Lys Leu
1 5 10

SEQ ID NO 14
LENGTH: 10
TYPE: PRT
ORGANISM: Mus musculus
SEQUENCE: 14
Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1 5 10

SEQ ID NO 15
LENGTH: 20
Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
1 5 10 15
Lys Ser Lys Lys
20

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15
Lys

nnnnnnnn nnnnnnnnn nnnnnnnnnnnn 27

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15
Lys

nnnnnnnn nnnnnnnnn nnnnnnnnnnnn 19
<221> NAME/KEY: modified_base
<222> LOCATION: [21]...[22]
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 19

nnnnnnnn nnmnnnnnnn nnmgaaw

<210> SEQ ID NO 20
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)...(11)
<223> OTHER INFORMATION: a, c, t or g
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (12)...(13)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 20

nnnnnnnn nnmgaaw

<210> SEQ ID NO 21
<211> LENGTH: 137
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 21

nnnnnnnn nnmnnnnnngtttttgtac tctcaagatt tagaataaa tcttcgacaa
gtcaaaaga taaggcttca tcgcgaastc aacaccgtt gattttatgg caggggtttt
tggtttattt atttttt

<210> SEQ ID NO 22
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 22

nnnnnnnn nnmnnnnnngtttttgtac tctcagaaat gcagaatga caaagaataag
gttcatgcgc gaataccaaca cccccgttct ttatggcagg gtttttttgt ttatttaatttttt
ttt

<210> SEQ ID NO 23
<211> LENGTH: 120
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified base
<222> LOCATION: (1) . . . (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 23

nnnnnnnnnnnnnnnnnnntttgtctacctcagaaactttagagctcaagataag 60
gtttacgaccgaatcagaccttttttttttctcatccttttcttttttt 110

<210> SEQ ID NO: 24
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: modified base
<222> LOCATION: (1) . . . (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 24

nnnnnnnnnnnnnnnnnttttagacctagataactgaatttaaatagccagtgccttc 60
cgggttcgctcttgggtgctctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
```
gtttagacg ta

<210> SEQ ID NO 28
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 28
tagcaagtta aastaagct agtccgtttt t

<210> SEQ ID NO 29
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 29
tagcaagtta aastaagct agtccgtttt t

<210> SEQ ID NO 30
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 29	nnnnnnnnn nnnnnnnnn nnaaaaw

<210> SEQ ID NO 31
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 31
guuuagacg us

<210> SEQ ID NO 32
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 32
gauuagacg us

<210> SEQ ID NO 33
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 33
cattgagaggt cacatcgag ttgctcccat tgg

<210> SEQ ID NO 34
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 33
ggaagggct ggtcgcagc agaagaaga ggg

<210> SEQ ID NO 35
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 33
ggaagggct ggtcgcagc agaagaaga ggg
```

US 8,889,356 B2
ORGANISM: Homo sapiens

SEQUENCE: 34

ggtggccgaga gggcccaga tttggtggtt c agg

SEQ ID NO: 35
LENGTH: 33
TYPE: DNA
ORGANISM: Homo sapiens

SEQUENCE: 35
atgcagaggg tggcgcagag gggccgagat tgg

SEQ ID NO: 36
LENGTH: 21
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 36
aaaaaccacc c ttctctc tgg c

SEQ ID NO: 37
LENGTH: 21
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 37
ggagtttgg gacacggaga g

SEQ ID NO: 38
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 38
cggaagccc aatgcctgac

SEQ ID NO: 39
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 39
ggcagcaac tcctgtcct

SEQ ID NO: 40
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

SEQUENCE: 40
gagggccat ttcccatgat tcctcatac ttcgatatgc gatacaaggg tgtagagag 60
ataattgga aaatatttgac ttgaaacacca aagatatagg tacaataactac gtagctgtaga 120
aagaatattaa ttccttgggtat gtttgcaaggt tttaaatattgt tttaataaatg ggcgactcat 180
atgctgccgg taacctggaa gttttcctgat ttccttgccttt tatatatcttt gtagaagagga 240
cgaaaacacg gacaactttaa aacgacaataa gcaagttaaa ataagctcttag ttccttatca 300
actgtgaaa gttgcaacagc gtcgctgctt tttt 335

<210> SEQ ID NO 41
<211> LENGTH: 423
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 41

<210> SEQ ID NO 42
<211> LENGTH: 339
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 42

<210> SEQ ID NO 43
<211> LENGTH: 309
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 43
<210> SEQ ID NO: 44
<211> LENGTH: 1648
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 44

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Arg Lys Val
20 225 30

Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu
30 40 45

Asp Ile Gly Thr Arg Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
50 55 60

Lys Val Pro Ser Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
60 70 75 80

Ser Ile Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
80 90 95

Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
100 105 110

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gin Glu Ile Phe Ser Asn Glu
110 120 125

Met Ala Lys Val Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
130 135 140

Leu Val Glu Glu Asp Lys His Glu Arg His Pro Ile Phe Gly Asn
140 150 155 160

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
160 170 175

Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
180 185 190

Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
190 200 205

Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
210 215 220

Ile Glu Leu Val Glu Thr Tyr Asn Glu Leu Phe Glu Glu Asn Pro Ile
220 230 235 240

Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
240 250 255

Lys Ser Arg Asp Leu Glu Asn Leu Ile Ala Gin Leu Pro Gly Glu Lys
260 265 270

Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
270 280 285

Pro Asn Phe Lys Ser Asn Phe Leu Ala Glu Asp Ala Lys Leu Gln
285 290 295 300

Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asn Leu Leu Ala Glu
300 310 315 320

Ile Gly Asp Glu Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
320 330 335
Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr 340 345 350
Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp His His 355 360 365
Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu 370 375 380
Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly 385 390 395 400
Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys 405 410 415
Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu 420 425 430
Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser 435 440 445
Ile Pro His Gln Ile His Leu Gly Gly Leu His Ala Ile Leu Arg Arg 450 455 460
Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu 465 470 475 480
Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg 485 490 495
 Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile 500 505 510
Thr Pro Trp Asn Phe Glu Val Val Asp Lys Gly Ala Ser Ala Gln 515 520 525
Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu 530 535 540
Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr 545 550 555 560
Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro 565 570 575
Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe 580 585 590
Lys Thr Asn Arg Lys Val Thr Val Lys Glu Leu Lys Glu Asp Tyr Phe 595 600 605
Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp 610 615 620
Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile 625 630 635 640
Lys Asp Lys Asp Phe Leu Asp Asn Glu Gly Asn Glu Asp Ile Leu Glu 645 650 655
Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu 660 665 670
Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Arg Lys Val Met Lys 675 680 685
Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys 690 695 700
Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp 705 710 715 720
Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asp Phe Met Gln Leu Ile 725 730 735
His Asp Arg Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val 740 745 750
<table>
<thead>
<tr>
<th>Ser</th>
<th>Gly</th>
<th>Gln</th>
<th>Gly</th>
<th>Asp</th>
<th>Ser</th>
<th>Leu</th>
<th>His</th>
<th>Glu</th>
<th>His</th>
<th>Ile</th>
<th>Ala</th>
<th>Asn</th>
<th>Leu</th>
<th>Ala</th>
<th>Gly</th>
</tr>
</thead>
<tbody>
<tr>
<td>762</td>
<td>760</td>
<td>765</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Pro</td>
<td>Ala</td>
<td>Ile</td>
<td>Lys</td>
<td>Gly</td>
<td>Ile</td>
<td>Leu</td>
<td>Gln</td>
<td>Thr</td>
<td>Val</td>
<td>Lys</td>
<td>Val</td>
<td>Val</td>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>770</td>
<td>775</td>
<td>780</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Val</td>
<td>Lys</td>
<td>Val</td>
<td>Met</td>
<td>Gly</td>
<td>His</td>
<td>Lys</td>
<td>Pro</td>
<td>Glu</td>
<td>Asn</td>
<td>Ile</td>
<td>Val</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>785</td>
<td>790</td>
<td>795</td>
<td>800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Met</td>
<td>Ala</td>
<td>Arg</td>
<td>Glu</td>
<td>Asn</td>
<td>Gln</td>
<td>Thr</td>
<td>Thr</td>
<td>Gin</td>
<td>Lys</td>
<td>Gly</td>
<td>Gin</td>
<td>Lys</td>
<td>Asn</td>
<td>Ser</td>
</tr>
<tr>
<td>805</td>
<td>810</td>
<td>815</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Glu</td>
<td>Arg</td>
<td>Met</td>
<td>Lys</td>
<td>Arg</td>
<td>Ile</td>
<td>Glu</td>
<td>Gly</td>
<td>Ile</td>
<td>Lys</td>
<td>Glu</td>
<td>Leu</td>
<td>Gly</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>820</td>
<td>825</td>
<td>830</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Ile</td>
<td>Leu</td>
<td>Lys</td>
<td>Glu</td>
<td>His</td>
<td>Pro</td>
<td>Val</td>
<td>Glu</td>
<td>Asn</td>
<td>Thr</td>
<td>Gin</td>
<td>Leu</td>
<td>Gln</td>
<td>Asn</td>
<td>Glu</td>
</tr>
<tr>
<td>835</td>
<td>840</td>
<td>845</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Leu</td>
<td>Tyr</td>
<td>Leu</td>
<td>Tyr</td>
<td>Tyr</td>
<td>Leu</td>
<td>Gin</td>
<td>Asn</td>
<td>Gly</td>
<td>Arg</td>
<td>Asp</td>
<td>Met</td>
<td>Tyr</td>
<td>Val</td>
<td>Asp</td>
</tr>
<tr>
<td>850</td>
<td>855</td>
<td>860</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gin</td>
<td>Glu</td>
<td>Leu</td>
<td>Asp</td>
<td>Ile</td>
<td>Asn</td>
<td>Arg</td>
<td>Leu</td>
<td>Ser</td>
<td>Asp</td>
<td>Tyr</td>
<td>Asp</td>
<td>Val</td>
<td>Asp</td>
<td>His</td>
<td>Ile</td>
</tr>
<tr>
<td>865</td>
<td>870</td>
<td>875</td>
<td>880</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Pro</td>
<td>Gin</td>
<td>Ser</td>
<td>Phe</td>
<td>Leu</td>
<td>Lys</td>
<td>Asp</td>
<td>Asp</td>
<td>Ser</td>
<td>Ile</td>
<td>Asn</td>
<td>Asp</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>885</td>
<td>890</td>
<td>895</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Arg</td>
<td>Ser</td>
<td>Asp</td>
<td>Lys</td>
<td>Asn</td>
<td>Arg</td>
<td>Gly</td>
<td>Lys</td>
<td>Ser</td>
<td>Asp</td>
<td>Asn</td>
<td>Val</td>
<td>Pro</td>
<td>Ser</td>
<td>Glu</td>
</tr>
<tr>
<td>900</td>
<td>905</td>
<td>910</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Val</td>
<td>Lys</td>
<td>Met</td>
<td>Lys</td>
<td>Asn</td>
<td>Tyr</td>
<td>Trp</td>
<td>Arg</td>
<td>Gin</td>
<td>Leu</td>
<td>Leu</td>
<td>Asn</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>915</td>
<td>920</td>
<td>925</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Leu</td>
<td>Ile</td>
<td>Thr</td>
<td>Gin</td>
<td>Arg</td>
<td>Phe</td>
<td>Asp</td>
<td>Asn</td>
<td>Leu</td>
<td>Thr</td>
<td>Lys</td>
<td>Ala</td>
<td>Glu</td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>930</td>
<td>935</td>
<td>940</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Gly</td>
<td>Leu</td>
<td>Ser</td>
<td>Glu</td>
<td>Leu</td>
<td>Asp</td>
<td>Lys</td>
<td>Ala</td>
<td>Gly</td>
<td>Phe</td>
<td>Ile</td>
<td>Lys</td>
<td>Arg</td>
<td>Gin</td>
<td>Leu</td>
</tr>
<tr>
<td>945</td>
<td>950</td>
<td>955</td>
<td>960</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Glu</td>
<td>Thr</td>
<td>Arg</td>
<td>Gin</td>
<td>Ile</td>
<td>Thr</td>
<td>Lys</td>
<td>His</td>
<td>Val</td>
<td>Ala</td>
<td>Gin</td>
<td>Leu</td>
<td>Asn</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>965</td>
<td>970</td>
<td>975</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Met</td>
<td>Asn</td>
<td>Thr</td>
<td>Lys</td>
<td>Tyr</td>
<td>Asp</td>
<td>Glu</td>
<td>Asn</td>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td>Ile</td>
<td>Arg</td>
<td>Glu</td>
<td>Val</td>
</tr>
<tr>
<td>980</td>
<td>985</td>
<td>990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Val</td>
<td>Ile</td>
<td>Thr</td>
<td>Leu</td>
<td>Lys</td>
<td>Ser</td>
<td>Lys</td>
<td>Leu</td>
<td>Val</td>
<td>Ser</td>
<td>Asp</td>
<td>Phe</td>
<td>Arg</td>
<td>Lys</td>
<td>Asp</td>
</tr>
<tr>
<td>995</td>
<td>1000</td>
<td>1005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Gln</td>
<td>Phe</td>
<td>Tyr</td>
<td>Lys</td>
<td>Val</td>
<td>Arg</td>
<td>Glu</td>
<td>Ile</td>
<td>Asn</td>
<td>Asn</td>
<td>Tyr</td>
<td>His</td>
<td>His</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>1010</td>
<td>1015</td>
<td>1020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Asp</td>
<td>Ala</td>
<td>Tyr</td>
<td>Leu</td>
<td>Asn</td>
<td>Ala</td>
<td>Val</td>
<td>Val</td>
<td>Gly</td>
<td>Thr</td>
<td>Al</td>
<td>Leu</td>
<td>Ile</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>1025</td>
<td>1030</td>
<td>1035</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Tyr</td>
<td>Pro</td>
<td>Lys</td>
<td>Leu</td>
<td>Glu</td>
<td>Ser</td>
<td>Glu</td>
<td>Phe</td>
<td>Val</td>
<td>Tyr</td>
<td>Gly</td>
<td>Aep</td>
<td>Tyr</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>1040</td>
<td>1045</td>
<td>1050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Tyr</td>
<td>Asp</td>
<td>Val</td>
<td>Arg</td>
<td>Lys</td>
<td>Met</td>
<td>Ile</td>
<td>Ala</td>
<td>Lys</td>
<td>Ser</td>
<td>Glu</td>
<td>Gin</td>
<td>Glu</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>1055</td>
<td>1060</td>
<td>1065</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Lys</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Lys</td>
<td>Tyr</td>
<td>Phe</td>
<td>Phe</td>
<td>Tyr</td>
<td>Ser</td>
<td>Asn</td>
<td>Ile</td>
<td>Met</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td>1070</td>
<td>1075</td>
<td>1080</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Phe</td>
<td>Lys</td>
<td>Thr</td>
<td>Glu</td>
<td>Ile</td>
<td>Thr</td>
<td>Leu</td>
<td>Ala</td>
<td>Asn</td>
<td>Gly</td>
<td>Glu</td>
<td>Ile</td>
<td>Arg</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>1085</td>
<td>1090</td>
<td>1095</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Pro</td>
<td>Leu</td>
<td>Ile</td>
<td>Glu</td>
<td>Thr</td>
<td>Asn</td>
<td>Gly</td>
<td>Glu</td>
<td>Thr</td>
<td>Gly</td>
<td>Glu</td>
<td>Ile</td>
<td>Val</td>
<td>Trp</td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>1105</td>
<td>1110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Lys</td>
<td>Gly</td>
<td>Arg</td>
<td>Asp</td>
<td>Phe</td>
<td>Ala</td>
<td>Thr</td>
<td>Val</td>
<td>Arg</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
<td>Ser</td>
<td>Met</td>
<td></td>
</tr>
<tr>
<td>1115</td>
<td>1120</td>
<td>1125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Gln</td>
<td>Val</td>
<td>Asn</td>
<td>Ile</td>
<td>Val</td>
<td>Lys</td>
<td>Thr</td>
<td>Glu</td>
<td>Val</td>
<td>Gin</td>
<td>Thr</td>
<td>Gly</td>
<td>Oly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1130</td>
<td>1135</td>
<td>1140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Lys</td>
<td>Glu</td>
<td>Ser</td>
<td>Ile</td>
<td>Leu</td>
<td>Pro</td>
<td>Lys</td>
<td>Arg</td>
<td>Asn</td>
<td>Ser</td>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>1145</td>
<td>1150</td>
<td>1155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Ala</td>
<td>Arg</td>
<td>Lys</td>
<td>Lys</td>
<td>Asp</td>
<td>Trp</td>
<td>Asp</td>
<td>Pro</td>
<td>Lys</td>
<td>Lys</td>
<td>Tyr</td>
<td>Gly</td>
<td>Phe</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1160</td>
<td>1165</td>
<td>1170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val</td>
<td>1175</td>
<td>1180</td>
<td>1185</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu Lys Gly Lys Ser Lys Leu Lys Ser Val Lys Glu Leu Leu</td>
<td>1190</td>
<td>1195</td>
<td>1200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Ser Pro Ile</td>
<td>1205</td>
<td>1210</td>
<td>1215</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu</td>
<td>1220</td>
<td>1225</td>
<td>1230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly</td>
<td>1235</td>
<td>1240</td>
<td>1245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Glu Lys Gly Asn</td>
<td>1250</td>
<td>1255</td>
<td>1260</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala</td>
<td>1265</td>
<td>1270</td>
<td>1275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser His Tyr Glu Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln</td>
<td>1290</td>
<td>1295</td>
<td>1300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Gin Leu Phe Val Glu Gin His Lys His Tyr Leu Asp Glu Ile</td>
<td>1295</td>
<td>1300</td>
<td>1305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile Gin Gin Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp</td>
<td>1310</td>
<td>1315</td>
<td>1320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp</td>
<td>1325</td>
<td>1330</td>
<td>1335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Pro Ile Arg Glu Gin Ala Glu Asn Ile Ile His Leu Phe Thr</td>
<td>1340</td>
<td>1345</td>
<td>1350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr</td>
<td>1355</td>
<td>1360</td>
<td>1365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp</td>
<td>1370</td>
<td>1375</td>
<td>1380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Thr Leu Ile His Gin Ser Ile Thr Gly Leu Tyr Glu Thr Arg</td>
<td>1385</td>
<td>1390</td>
<td>1395</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile Asp Leu Ser Gin Leu Gly Gly Asp Ala Ala Ala Val Ser Lys</td>
<td>1400</td>
<td>1405</td>
<td>1410</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu</td>
<td>1415</td>
<td>1420</td>
<td>1425</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly</td>
<td>1430</td>
<td>1435</td>
<td>1440</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys</td>
<td>1445</td>
<td>1450</td>
<td>1455</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr</td>
<td>1460</td>
<td>1465</td>
<td>1470</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Thr Tyr Gly Val Gin Cys Phe Ser Arg Tyr Pro Asp His Met</td>
<td>1475</td>
<td>1480</td>
<td>1485</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Gin His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val</td>
<td>1490</td>
<td>1495</td>
<td>1500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gin Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr</td>
<td>1505</td>
<td>1510</td>
<td>1515</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile</td>
<td>1520</td>
<td>1525</td>
<td>1530</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly</td>
<td>1535</td>
<td>1540</td>
<td>1545</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met</td>
<td>1550</td>
<td>1555</td>
<td>1560</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Lys Gln Lys Arg Gly Ile Lys Val Asn Phe Lys Ile Arg</td>
<td>1565 1570 1575</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln</td>
<td>1580 1585 1590</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn</td>
<td>1595 1600 1605</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu</td>
<td>1610 1615 1620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly</td>
<td>1625 1630 1635</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys</td>
<td>1640 1645</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

165 -continued

<210> SEQ ID NO: 45
<211> LENGTH: 1625
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<table>
<thead>
<tr>
<th>&lt;400&gt; SEQUENCE: 45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Asp Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val</td>
</tr>
<tr>
<td>Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys</td>
</tr>
<tr>
<td>Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Asn Leu Ile</td>
</tr>
<tr>
<td>Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu</td>
</tr>
<tr>
<td>Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys</td>
</tr>
<tr>
<td>Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser</td>
</tr>
<tr>
<td>Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys</td>
</tr>
<tr>
<td>His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr</td>
</tr>
<tr>
<td>His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Leu Val Asp</td>
</tr>
<tr>
<td>Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His</td>
</tr>
<tr>
<td>Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro</td>
</tr>
<tr>
<td>Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr</td>
</tr>
<tr>
<td>Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr</td>
</tr>
<tr>
<td>Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn</td>
</tr>
<tr>
<td>Leu Ile Ala Gin Leu Pro Gly Glu Lys Asn Gin Leu Phe Gly Asn</td>
</tr>
<tr>
<td>Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe</td>
</tr>
<tr>
<td>Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp</td>
</tr>
<tr>
<td>260</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Asp</td>
</tr>
<tr>
<td>275</td>
</tr>
<tr>
<td>Leu</td>
</tr>
<tr>
<td>290</td>
</tr>
<tr>
<td>Ile</td>
</tr>
<tr>
<td>305</td>
</tr>
<tr>
<td>Met</td>
</tr>
<tr>
<td>325</td>
</tr>
<tr>
<td>Ala</td>
</tr>
<tr>
<td>340</td>
</tr>
<tr>
<td>Asp</td>
</tr>
<tr>
<td>355</td>
</tr>
<tr>
<td>Gln</td>
</tr>
<tr>
<td>370</td>
</tr>
<tr>
<td>Gly</td>
</tr>
<tr>
<td>385</td>
</tr>
<tr>
<td>Lys</td>
</tr>
<tr>
<td>400</td>
</tr>
<tr>
<td>Gly</td>
</tr>
<tr>
<td>415</td>
</tr>
<tr>
<td>Leu</td>
</tr>
<tr>
<td>430</td>
</tr>
<tr>
<td>Pro</td>
</tr>
<tr>
<td>445</td>
</tr>
<tr>
<td>Met</td>
</tr>
<tr>
<td>460</td>
</tr>
<tr>
<td>Val</td>
</tr>
<tr>
<td>475</td>
</tr>
<tr>
<td>Asn</td>
</tr>
<tr>
<td>490</td>
</tr>
<tr>
<td>Leu</td>
</tr>
<tr>
<td>505</td>
</tr>
<tr>
<td>Tyr</td>
</tr>
<tr>
<td>520</td>
</tr>
<tr>
<td>Lys</td>
</tr>
<tr>
<td>535</td>
</tr>
<tr>
<td>Val</td>
</tr>
<tr>
<td>550</td>
</tr>
<tr>
<td>Ser</td>
</tr>
<tr>
<td>565</td>
</tr>
<tr>
<td>Thr</td>
</tr>
<tr>
<td>580</td>
</tr>
<tr>
<td>Asn</td>
</tr>
<tr>
<td>595</td>
</tr>
<tr>
<td>Leu</td>
</tr>
<tr>
<td>610</td>
</tr>
<tr>
<td>His</td>
</tr>
<tr>
<td>625</td>
</tr>
<tr>
<td>Thr</td>
</tr>
<tr>
<td>640</td>
</tr>
<tr>
<td>Lys</td>
</tr>
<tr>
<td>655</td>
</tr>
</tbody>
</table>

-continued
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Glu Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Glu Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Gly Ile Gly Glu Leu Gly Ser Glu Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Val Leu Val Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Aam Val Pro Ser Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Aam Ala Lys Leu Ile Thr Glu Arg Lys
885 890 895
Phe Asp Aam Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Glu Leu Val Glu Thr Arg Glu Ile Thr
915 920 925
Lys His Val Ala Glu Ile Leu Aam Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Aam Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Tyr Tyr Tyr Tyr Tyr Val Arg
965 970 975
Glu Ile Asn Aam Tyr His His Ala His Asp Ala Tyr Leu Aam Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asp Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Aam Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Aam Ile Val Lys Lys Thr
1085 1090 1095
<table>
<thead>
<tr>
<th>Glu Val</th>
<th>Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys</th>
<th>1100</th>
<th>1105</th>
<th>1110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg Asn</td>
<td>Ser Asp Lys Leu Ile Ala Arg Lys Asp Thr Asp Pro</td>
<td>1115</td>
<td>1120</td>
<td>1125</td>
</tr>
<tr>
<td>Lys Lys</td>
<td>Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val</td>
<td>1130</td>
<td>1135</td>
<td>1140</td>
</tr>
<tr>
<td>Leu Val</td>
<td>Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys</td>
<td>1145</td>
<td>1150</td>
<td>1155</td>
</tr>
<tr>
<td>Ser Val</td>
<td>Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser</td>
<td>1160</td>
<td>1165</td>
<td>1170</td>
</tr>
<tr>
<td>Phe Glu</td>
<td>Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys</td>
<td>1175</td>
<td>1180</td>
<td>1185</td>
</tr>
<tr>
<td>Glu Val</td>
<td>Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu</td>
<td>1190</td>
<td>1195</td>
<td>1200</td>
</tr>
<tr>
<td>Phe Glu</td>
<td>Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly</td>
<td>1205</td>
<td>1210</td>
<td>1215</td>
</tr>
<tr>
<td>Glu Leu</td>
<td>Glu Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val</td>
<td>1220</td>
<td>1225</td>
<td>1230</td>
</tr>
<tr>
<td>Asn Phe</td>
<td>Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser</td>
<td>1235</td>
<td>1240</td>
<td>1245</td>
</tr>
<tr>
<td>Pro Glu</td>
<td>Asp Asn Glu Glu Lys Gln Leu Phe Val Glu Gln His Lys</td>
<td>1250</td>
<td>1255</td>
<td>1260</td>
</tr>
<tr>
<td>His Tyr</td>
<td>Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys</td>
<td>1265</td>
<td>1270</td>
<td>1275</td>
</tr>
<tr>
<td>Arg Val</td>
<td>Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala</td>
<td>1280</td>
<td>1285</td>
<td>1290</td>
</tr>
<tr>
<td>Tyr Asn</td>
<td>Lys His Arg Asp Lys Pro Ile Arg Glu Glu Ala Glu Asn</td>
<td>1295</td>
<td>1300</td>
<td>1305</td>
</tr>
<tr>
<td>Ile Ile</td>
<td>His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala</td>
<td>1310</td>
<td>1315</td>
<td>1320</td>
</tr>
<tr>
<td>Phe Lys</td>
<td>Tyr Phe Asp Thr Thr Thr Asp Arg Lys Arg Tyr Thr Ser</td>
<td>1325</td>
<td>1330</td>
<td>1335</td>
</tr>
<tr>
<td>Thr Lys</td>
<td>Glu Val Leu Asp Ala Thr Leu Ile His Glu Ser Ile Thr</td>
<td>1340</td>
<td>1345</td>
<td>1350</td>
</tr>
<tr>
<td>Gly Leu</td>
<td>Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp</td>
<td>1355</td>
<td>1360</td>
<td>1365</td>
</tr>
<tr>
<td>Ala Ala</td>
<td>Ala Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val</td>
<td>1370</td>
<td>1375</td>
<td>1380</td>
</tr>
<tr>
<td>Pro Ile</td>
<td>Leu Val Leu Leu Asp Gly Asp Val Asn Gly His Lys Phe</td>
<td>1385</td>
<td>1390</td>
<td>1395</td>
</tr>
<tr>
<td>Ser Val</td>
<td>Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu</td>
<td>1400</td>
<td>1405</td>
<td>1410</td>
</tr>
<tr>
<td>Thr Leu</td>
<td>Lys Phe Ile Cys Thr Thr Glu Lys Leu Pro Val Pro Trp</td>
<td>1415</td>
<td>1420</td>
<td>1425</td>
</tr>
<tr>
<td>Pro Thr</td>
<td>Leu Val Thr Thr Leu Thr Tyr Glu Val Gln Cys Phe Ser</td>
<td>1430</td>
<td>1435</td>
<td>1440</td>
</tr>
<tr>
<td>Arg Tyr</td>
<td>Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala</td>
<td>1445</td>
<td>1450</td>
<td>1455</td>
</tr>
<tr>
<td>Met Pro</td>
<td>Glu Gly Tyr Val Glu Glu Arg Thr Ile Phe Phe Lys Asp</td>
<td>1460</td>
<td>1465</td>
<td>1470</td>
</tr>
<tr>
<td>Asp Gly</td>
<td>Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp</td>
<td>1475</td>
<td>1480</td>
<td>1485</td>
</tr>
<tr>
<td>Thr Leu</td>
<td>Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1490</td>
<td>1495</td>
<td>1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser</td>
<td>1490</td>
<td>1495</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>1505</td>
<td>1510</td>
<td>1515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His Asn Val Tyr Ile Met Ala Asp Lys Glu Lys Asn Gly Ile Lys</td>
<td>1520</td>
<td>1525</td>
<td>1530</td>
<td></td>
</tr>
<tr>
<td>1535</td>
<td>1540</td>
<td>1545</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Glu</td>
<td>1550</td>
<td>1555</td>
<td>1560</td>
<td></td>
</tr>
<tr>
<td>1565</td>
<td>1570</td>
<td>1575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Ala Asp His Tyr Glu Glu Asn Thr Pro Ile Gly Asp Gly Pro</td>
<td>1580</td>
<td>1585</td>
<td>1590</td>
<td></td>
</tr>
<tr>
<td>1595</td>
<td>1600</td>
<td>1605</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Glu Ser Ala Leu</td>
<td>1610</td>
<td>1615</td>
<td>1620</td>
<td></td>
</tr>
<tr>
<td>1625</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SEQ ID NO 46**

**LENGTH:** 1644

**TYPE:** PRT

**ORGANISM:** Artificial Sequence

**FEATURE:**

**OTHER INFORMATION:** Description of Artificial Sequence: Synthetic polypeptide

**SEQUENCE:** 46

```
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1   5   10   15
Tyr Lys Asp Asp Asp Lys Met Ala Pro Lys Lys Arg Lys Val
20  25  30
Gly Ile His Gly Val Pro Ala Ala Asp Lys Tyr Ser Ile Gly Leu
35  40  45
Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
50  55  60
Lys Val Pro Ser Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
65  70  75  80
Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
85  90  95
Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
100 105 110
Arg Arg Lys Asn Arg Ile Cys Tyr Leu Glu Glu Ile Phe Ser Asn Glu
115 120 125
Met Ala Lys Val Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
130 135 140
Leu Val Glu Glu Asp Lys His Glu Arg His Pro Ile Phe Gly Asn
145 150 155 160
Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
165 170 175
Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
180 185 190
Ile Tyr Leu Ala Leu Ala Asp Ser Lys Phe Arg Gly His Phe Leu
195 200 205
```
<table>
<thead>
<tr>
<th>Residues</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe</td>
<td>210 215 220</td>
</tr>
<tr>
<td>Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Asn Pro Ile</td>
<td>225 230 235 240</td>
</tr>
<tr>
<td>Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser</td>
<td>245 250 255</td>
</tr>
<tr>
<td>Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys</td>
<td>260 265 270</td>
</tr>
<tr>
<td>Lys Asn Gln Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr</td>
<td>275 280 285</td>
</tr>
<tr>
<td>Pro Asn Phe Lys Ser Asn Phe Asp Ala Glu Asp Ala Lys Leu Gln</td>
<td>290 295 300</td>
</tr>
<tr>
<td>Leu Ser Lys Asp Thr Tyr Asp Asp Leu Asp Asn Leu Ala Gln</td>
<td>305 310 315 320</td>
</tr>
<tr>
<td>Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser</td>
<td>325 330 335</td>
</tr>
<tr>
<td>Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr</td>
<td>340 345 350</td>
</tr>
<tr>
<td>Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His</td>
<td>355 360 365</td>
</tr>
<tr>
<td>Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu</td>
<td>370 375 380</td>
</tr>
<tr>
<td>Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gln</td>
<td>385 390 395 400</td>
</tr>
<tr>
<td>Tyr Ile Asp Gly Gln Leu Ser Asp Gln Glu Phe Tyr Lys Phe Ile Lys</td>
<td>405 410 415</td>
</tr>
<tr>
<td>Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu</td>
<td>420 425 430</td>
</tr>
<tr>
<td>Asn Arg Glu Asp Leu Arg Lys Gln Arg Thr Phe Asp Asp Gly Ser</td>
<td>435 440 445</td>
</tr>
<tr>
<td>Ile Pro His Gln Ile His Leu Gly Leu His Ala Ile Leu Arg Arg</td>
<td>450 455 460</td>
</tr>
<tr>
<td>Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu</td>
<td>465 470 475 480</td>
</tr>
<tr>
<td>Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg</td>
<td>485 490 495</td>
</tr>
<tr>
<td>Gly Asn Ser Arg Phe Ala Thr Met Thr Arg Lys Ser Glu Glu Thr Ile</td>
<td>500 505 510</td>
</tr>
<tr>
<td>Thr Pro Trp Asn Phe Glu Val Val Asp Lys Gly Ala Ser Ala Gln</td>
<td>515 520 525</td>
</tr>
<tr>
<td>Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu</td>
<td>530 535 540</td>
</tr>
<tr>
<td>Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr</td>
<td>545 550 555 560</td>
</tr>
<tr>
<td>Asn Glu Leu Thr Lys Val Tyr Val Thr Glu Gly Met Arg Lys Pro</td>
<td>565 570 575</td>
</tr>
<tr>
<td>Ala Phe Leu Ser Gly Glu Gin Lys Lys Ala Ile Val Asp Leu Leu Phe</td>
<td>580 585 590</td>
</tr>
<tr>
<td>Lys Thr Asn Arg Lys Val Thr Val Lys Gin Leu Lys Glu Asp Tyr Phe</td>
<td>595 600 605</td>
</tr>
<tr>
<td>Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp</td>
<td>610 615 620</td>
</tr>
<tr>
<td>Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile</td>
<td></td>
</tr>
</tbody>
</table>
Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
645 650 655
Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
660 665 670
Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Arg Lys Val Met Lys
675 680 685
Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
690 695 700
Leu Ile Asn Gly Ile Arg Asp Gln Ser Gly Lys Thr Ile Leu Asp
705 710 715 720
Phe Leu Lys Ser Asp Gln Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
725 730 735
His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
740 745 750
Ser Gly Gln Gly Arg Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
755 760 765
Ser Pro Ala Ile Lys Lys Ile Leu Gin Thr Val Lys Val Val Asp
770 775 780
Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
785 790 795 800
Glu Met Ala Arg Glu Asn Gln Thr Thr Gin Lys Gly Gin Lys Asn Ser
805 810 815
Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Lys Gly Ser
820 825 830
Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gin Leu Gin Asn Glu
835 840 845
Lys Leu Tyr Leu Tyr Tyr Leu Gin Asn Gly Arg Asp Met Tyr Val Asp
850 855 860
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
865 870 875 880
Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
885 890 895
Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
900 905 910
Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gin Leu Leu Asn Ala
915 920 925
Lys Leu Ile Thr Gin Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
930 935 940
Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gin Leu
945 950 955 960
Val Glu Thr Arg Gin Ile Thr Lys His Val Ala Gin Ile Leu Asp Ser
965 970 975
Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
980 985 990
Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
995 1000 1005
Phe Gin Phe Tyr Lys Val Arg Glu Ile Asn Tyr His His Ala
1010 1015 1020
His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys
1025 1030 1035
Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys
1040 1045 1050
<table>
<thead>
<tr>
<th>Val Tyr</th>
<th>Asp Val Arg Lys Met</th>
<th>Ile Ala Lys Ser Glu</th>
<th>Gln Glu Ile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1055</td>
<td></td>
<td>1060</td>
<td>1065</td>
</tr>
<tr>
<td>Gly Lys</td>
<td>Ala Thr Ala Lys Tyr</td>
<td>Phe Phe Tyr Ser Asn</td>
<td>Ile Met Asn</td>
</tr>
<tr>
<td>1070</td>
<td></td>
<td>1075</td>
<td>1080</td>
</tr>
<tr>
<td>Phe Phe</td>
<td>Lys Thr Glu Ile Thr</td>
<td>Leu Ala Asn Gly Glu</td>
<td>Ile Arg Lys</td>
</tr>
<tr>
<td>1085</td>
<td></td>
<td>1090</td>
<td>1095</td>
</tr>
<tr>
<td>Arg Pro</td>
<td>Leu Ile Glu Thr Asn</td>
<td>Gly Glu Thr Gly Glu</td>
<td>Ile Val Trp</td>
</tr>
<tr>
<td>1100</td>
<td></td>
<td>1105</td>
<td>1110</td>
</tr>
<tr>
<td>Asp Lys</td>
<td>Gly Arg Asp Phe Ala</td>
<td>Thr Val Arg Lys Val</td>
<td>Leu Ser Met</td>
</tr>
<tr>
<td>1115</td>
<td></td>
<td>1120</td>
<td>1125</td>
</tr>
<tr>
<td>Pro Gln</td>
<td>Val Asn Ile Val Lys</td>
<td>Lys Thr Glu Val Gln</td>
<td>Thr Gly Gly</td>
</tr>
<tr>
<td>1130</td>
<td></td>
<td>1135</td>
<td>1140</td>
</tr>
<tr>
<td>Phe Ser</td>
<td>Lys Glu Ser Ile Leu</td>
<td>Pro Lys Arg Asn Ser</td>
<td>Asp Lys Leu</td>
</tr>
<tr>
<td>1145</td>
<td></td>
<td>1150</td>
<td>1155</td>
</tr>
<tr>
<td>Ile Ala</td>
<td>Arg Lys Lys Asp Trp</td>
<td>Asp Pro Lys Tyr Gly Gly Phe</td>
<td></td>
</tr>
<tr>
<td>1160</td>
<td></td>
<td>1165</td>
<td>1170</td>
</tr>
<tr>
<td>Asp Ser</td>
<td>Pro Thr Val Ala Tyr</td>
<td>Ser Val Leu Val Val</td>
<td>Ala Lys Val</td>
</tr>
<tr>
<td>1175</td>
<td></td>
<td>1180</td>
<td>1185</td>
</tr>
<tr>
<td>Glu Lys</td>
<td>Gly Lys Ser Lys Lys</td>
<td>Leu Lys Ser Val Lys</td>
<td>Glu Leu Leu</td>
</tr>
<tr>
<td>1190</td>
<td></td>
<td>1195</td>
<td>1200</td>
</tr>
<tr>
<td>Gly Ile</td>
<td>Thr Ile Met Glu Arg</td>
<td>Ser Ser Phe Glu Lys</td>
<td>Asn Pro Ile</td>
</tr>
<tr>
<td>1205</td>
<td></td>
<td>1210</td>
<td>1215</td>
</tr>
<tr>
<td>Asp Phe</td>
<td>Leu Glu Ala Lys Gly</td>
<td>Tyr Lys Glu Val Lys</td>
<td>Lys Asp Leu</td>
</tr>
<tr>
<td>1220</td>
<td></td>
<td>1225</td>
<td>1230</td>
</tr>
<tr>
<td>Ile Ile</td>
<td>Lys Leu Pro Lys Tyr</td>
<td>Ser Leu Phe Glu Leu</td>
<td>Glu Asn Gly</td>
</tr>
<tr>
<td>1235</td>
<td></td>
<td>1240</td>
<td>1245</td>
</tr>
<tr>
<td>Arg Lys</td>
<td>Arg Met Leu Ala Ser</td>
<td>Ala Gly Glu Leu Gln</td>
<td>Lys Gly Asn</td>
</tr>
<tr>
<td>1250</td>
<td></td>
<td>1255</td>
<td>1260</td>
</tr>
<tr>
<td>Glu Leu</td>
<td>Ala Leu Pro Ser Lys</td>
<td>Tyr Val Asn Phe Leu</td>
<td>Tyr Leu Ala</td>
</tr>
<tr>
<td>1265</td>
<td></td>
<td>1270</td>
<td>1275</td>
</tr>
<tr>
<td>Ser His</td>
<td>Tyr Glu Lys Leu Lys</td>
<td>Gly Ser Pro Glu Asp</td>
<td>Asn Glu Gln</td>
</tr>
<tr>
<td>1280</td>
<td></td>
<td>1285</td>
<td>1290</td>
</tr>
<tr>
<td>Lys Gln</td>
<td>Leu Phe Val Glu Gln</td>
<td>His Lys His Tyr Leu</td>
<td>Asp Glu Ile</td>
</tr>
<tr>
<td>1295</td>
<td></td>
<td>1300</td>
<td>1305</td>
</tr>
<tr>
<td>Ile Glu</td>
<td>Gln Ile Ser Glu Phe</td>
<td>Ser Lys Arg Val Ile</td>
<td>Leu Ala Asp</td>
</tr>
<tr>
<td>1310</td>
<td></td>
<td>1315</td>
<td>1320</td>
</tr>
<tr>
<td>Ala Asn</td>
<td>Leu Asp Lys Val Leu</td>
<td>Ser Ala Tyr Asn Lys</td>
<td>His Arg Asp</td>
</tr>
<tr>
<td>1325</td>
<td></td>
<td>1330</td>
<td>1335</td>
</tr>
<tr>
<td>Lys Pro</td>
<td>Ile Arg Glu Gln Ala</td>
<td>Glu Asn Ile Ile His</td>
<td>Leu Phe Thr</td>
</tr>
<tr>
<td>1340</td>
<td></td>
<td>1345</td>
<td>1350</td>
</tr>
<tr>
<td>Leu Thr</td>
<td>Asn Leu Gly Ala Pro</td>
<td>Ala Ala Phe Lys Tyr</td>
<td>Phe Asp Thr</td>
</tr>
<tr>
<td>1355</td>
<td></td>
<td>1360</td>
<td>1365</td>
</tr>
<tr>
<td>Thr Ile</td>
<td>Asp Arg Lys Arg Tyr</td>
<td>Thr Ser Thr Lys Glu</td>
<td>Val Leu Asp</td>
</tr>
<tr>
<td>1370</td>
<td></td>
<td>1375</td>
<td>1380</td>
</tr>
<tr>
<td>Ala Thr</td>
<td>Leu Ile His Gln Ser</td>
<td>Ile Thr Gly Leu Tyr</td>
<td>Glu Thr Arg</td>
</tr>
<tr>
<td>1385</td>
<td></td>
<td>1390</td>
<td>1395</td>
</tr>
<tr>
<td>Ile Asp</td>
<td>Leu Ser Gln Leu Gly</td>
<td>Gly Asp Ala Ala Ala</td>
<td>Val Ser Lys</td>
</tr>
<tr>
<td>1400</td>
<td></td>
<td>1405</td>
<td>1410</td>
</tr>
<tr>
<td>Gly Glu</td>
<td>Glu Leu Phe Thr Gly</td>
<td>Val Val Pro Ile Leu</td>
<td>Val Glu Leu</td>
</tr>
<tr>
<td>1415</td>
<td></td>
<td>1420</td>
<td>1425</td>
</tr>
<tr>
<td>Asp Gly</td>
<td>Asp Val Asn Gly His</td>
<td>Lys Phe Ser Val Ser</td>
<td>Gly Glu Gly</td>
</tr>
<tr>
<td>1430</td>
<td></td>
<td>1435</td>
<td>1440</td>
</tr>
</tbody>
</table>
Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
1445 1450 1455
Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
1460 1465 1470
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met
1475 1480 1485
Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val
1490 1495 1500
Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr
1505 1510 1515
Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile
1520 1525 1530
Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
1535 1540 1545
His Lys Leu Glu Tyr Asn Tyr Aen Ser His Asn Val Tyr Ile Met
1550 1555 1560
Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg
1565 1570 1575
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln
1580 1585 1590
Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn
1595 1600 1605
His Tyr Leu Ser Thr Glu Ser Ala Leu Ser Lys Asp Pro Asn Glu
1610 1615 1620
Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly
1625 1630 1635
Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Arg Pro Ala Ala
1640 1645 1650
Thr Lys Ala Gly Gln Ala Lys Lys Lys Lys
1655 1660

<210> SEQ ID NO 47
<211> LENGTH: 1423
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<400> SEQUENCE: 47
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1  5  10  15
Tyr Lys Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20  25  30
Gly Ile His Gly Val Pro Ala Ala Asp Lys Tyr Ser Ile Gly Leu
35  40
Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
50  55  60
Lys Val Pro Ser Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
65  70  75  80
Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
95  90  95
Thr Ala Glu Ala Thr Arg Leu Arg Thr Ala Arg Arg Arg Tyr Thr
100 105 110
Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
115 120 125
Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
130 135 140
Leu Val Glu Glu Asp Lys His Glu Arg His Pro Ile Phe Gly Asn
145 150 155 160
Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
165 170 175
Leu Arg Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
180 185 190
Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
195 200 205
Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
210 215 220
Ile Glu Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Gln Asn Pro Ile
225 230 235 240
Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
245 250 255
Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
260 265 270
Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
275 280 285
Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
290 295 300
Leu Ser Lys Asp Thr Tyr Asp Asp Leu Asn Leu Ala Gln
305 310 315 320
Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
325 330 335
Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Gln Ile Thr
340 345 350
Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His
355 360 365
Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu
370 375 380
Lys Tyr Lys Glu Ile Phe Phe Asp Glu Ser Lys Asn Gly Tyr Ala Gly
385 390 395 400
Tyr Ile Asp Gly Gln Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys
405 410 415
Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu
420 425 430
Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser
435 440 445
Ile Pro His Gln Ile His Leu Gly Leu His Ala Ile Leu Arg Arg
450 455 460
Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Arg Arg Gln Lys Ile Glu
465 470 475 480
Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg
485 490 495
Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
500 505 510
Thr Pro Trp Asn Phe Glu Val Val Asp Lys Gly Ala Ser Ala Gln
515 520 525
Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
530 535 540
Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr
545 550 555 560
Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro
565 570 575
Ala Phe Leu Ser Gly Glu Gln Lys Tyr Ala Ile Val Asp Leu Leu Phe
580 585 590
Lys Thr Asn Arg Lys Val Thr Val Lys Gin Leu Leu Gly Asp Tyr Phe
595 600 605
Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp
610 615 620
Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile
625 630 635 640
Lys Asp Lys Asp Phe Leu Asp Asn Glu Gin Asp Ile Leu Glu
645 650 655
Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
660 665 670
Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
675 680 685
Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
690 695 700
Leu Ile Asn Gly Ile Arg Asp Gin Ser Gly Lys Thr Ile Leu Asp
705 710 715 720
Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gin Leu Ile
725 730 735
His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gin Lys Ala Gin Val
740 745 750
Ser Gly Gin Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
755 760 765
Ser Pro Ala Ile Lys Gly Ile Leu Gin Thr Val Lys Val Asp
770 775 780
Glu Leu Val Lys Val Met Gin Arg His Lys Pro Gin Asn Ile Val Ile
785 790 795 800
Glu Met Ala Arg Glu Asn Gin Thr Thr Gin Lys Gly Gin Lys Asn Ser
805 810 815
Arg Glu Arg Met Lys Arg Arg Gin Arg Gin Lys Lys Leu Gly Ser
820 825 830
Gln Ile Leu Lys Gin Pro Val Gin Gin Asn Gin Gin Gin
835 840 845
Lys Leu Tyr Leu Tyr Tyr Leu Gin Asn Gly Arg Asp Met Tyr Val Asp
850 855 860
Gln Glu Leu Asp Ile Gin Arg Leu Ser Asp Tyr Val Gin His Ile
865 870 875 880
Val Pro Gin Ser Gin Gin Ser Ile Gin Arg Gin Gin Gin
885 890 895
Thr Arg Ser Gin Gin Gin Arg Gin Gin Gin Gin Gin Gin
900 905 910
Glu Val Val Lys Gin Met Gin Gin Gin Gin Gin Gin Gin
915 920 925
Lys Leu Ile Thr Gin Gin Gin Leu Gin Gin Gin Gin Gin
930 935 940
Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
945 950 955 960
Val Gin Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
975 980 985 990
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1005 1010 1015 1020
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1030 1035 1040 1045
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1050 1055 1060 1065
Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1070 1075 1080 1085
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1090 1095 1100 1105
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1110 1115 1120 1125
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1130 1135 1140 1145
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1150 1155 1160 1165
Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1170 1175 1180 1185
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1190 1195 1200 1205
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1210 1215 1220 1225
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1230 1235 1240 1245
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1250 1255 1260 1265
Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1270 1275 1280 1285
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1290 1305 1310 1315
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1320 1325 1330 1335
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1340 1345 1350 1355
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1360 1365 1370 1375
Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1380 1385 1390 1395
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1400 1405 1410 1415
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1420 1425 1430 1435
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1440 1445 1450 1455
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1460 1465 1470 1475
Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1480 1485 1490 1495
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1500 1505 1510 1515
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1520 1525 1530 1535
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1540 1545 1550 1555
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1560 1565 1570 1575
Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1580 1585 1590 1595
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1600 1605 1610 1615
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1620 1625 1630 1635
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1640 1645 1650 1655
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1660 1665 1670 1675
Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1680 1685 1690 1695
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1700 1705 1710 1715
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1720 1725 1730 1735
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1740 1745 1750 1755
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1760 1765 1770 1775
Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1780 1785 1790 1795
Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1800 1805 1810 1815
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1820 1825 1830 1835
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1840 1845 1850 1855
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1860 1865 1870 1875
Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
980 985 990
Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
995 1000 1005
Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala
1010 1015 1020
His Asp Ala Tyr Leu Asp Ala Val Val Gly Thr Ala Leu Ile Lys
1025 1030 1035
Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys
1040 1045 1050
Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile
1055 1060 1065
Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn
1070 1075 1080
Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys
1095 1090 1095
Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp
1100 1105 1110
Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met
1115 1120 1125
Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
1130 1135 1140
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu
1145 1150 1155
Ile Ala Arg Lys Lys Asp Thr Asp Pro Lys Tyr Gly Gly Phe
1160 1165 1170
Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val
1175 1180 1185
Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu
1190 1195 1200
Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile
1205 1210 1215
Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu
1220 1225 1230
Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly
1235 1240 1245
Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn
1250 1255 1260
Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala
1265 1270 1275
Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln
1280 1285 1290
Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile
1295 1300 1305
Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp
1310 1315 1320
Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
1325 1330 1335
Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr
1340 1345 1350
Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
1355 1360 1365
Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
\hspace{1em} 1370 \hspace{1em} 1375 \hspace{1em} 1380

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg
\hspace{1em} 1385 \hspace{1em} 1390 \hspace{1em} 1395

Ile Asp Leu Ser Gln Leu Gly Gly Asp Lys Arg Pro Ala Ala Thr
\hspace{1em} 1400 \hspace{1em} 1405 \hspace{1em} 1410

Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
\hspace{1em} 1415 \hspace{1em} 1420

\textless 210 \textgreater \text{ SEQ ID NO } 48
\textless 211 \textgreater \text{ LENGTH: } 483
\textless 212 \textgreater \text{ TYPE: PRT}
\textless 213 \textgreater \text{ ORGANISM: Artificial Sequence}
\textless 220 \textgreater \text{ FEATURE:}
\textless 223 \textgreater \text{ OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide}

\textless 400 \textgreater \text{ SEQUENCE: 48}

Met Phe Leu Phe Leu Ser Leu Thr Ser Phe Leu Ser Ser Ser Arg Thr
\hspace{1em} 1 \hspace{1em} 5 \hspace{1em} 10 \hspace{1em} 15

Leu Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Lys Glu Phe
\hspace{1em} 20 \hspace{25} 30

Met Arg Phe Lys Val His Met Gly Glu Ser Val Asn Gly His Glu Phe
\hspace{35} 40 \hspace{45}

Glu Ile Glu Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Glu Thr
\hspace{50} 60

Ala Lys Leu Lys Val Thr Lys Gly Glu Pro Leu Pro Phe Ala Trp Asp
\hspace{65} 70 \hspace{75} 80

Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His
\hspace{85} 90 \hspace{95}

Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe
\hspace{100} 105 \hspace{110}

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Thr Thr Val
\hspace{115} 120 \hspace{125}

Thr Glu Asp Ser Ser Leu Gln Asp Gly Phe Ile Tyr Lys Val Lys
\hspace{130} 140

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys
\hspace{145} 155 \hspace{160}

Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly
\hspace{165} 175

Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Glu Gly
\hspace{180} 190

His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val
\hspace{195} 200 \hspace{205}

Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser
\hspace{210} 215 \hspace{220}

His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly
\hspace{225} 230 \hspace{235} 240

Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Ser Lys Gln
\hspace{245} 250 \hspace{255}

Leu Glu Glu Leu Ser Thr Ser Phe Asp Ile Gln Phe Asn Asp Leu
\hspace{260} 265 \hspace{270}

Thr Leu Leu Glu Thr Ala Phe Thr His Thr Ser Tyr Ala Asn Glu His
\hspace{275} 280 \hspace{285}

Arg Leu Asn Val Ser His Asn Glu Arg Leu Glu Phe Leu Gly Asp
<210> SEQ ID NO 49
<211> LENGTH: 463
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 49

Met Lys Gin Leu Glu Leu Leu Ser Thr Ser Phe Asp Ile Gin Phe
      1      5     10     15
Asn Asp Leu Thr Leu Leu Glu Thr Ala Phe Thr His Thr Ser Tyr Ala
      20
Asn Glu His Arg Leu Leu Asn Val Ser His Asn Gin Arg Leu Gin Phe
      35
      40
Leu Gly Asp Ala Val Leu Gin Leu Ile Ile Ser Gin Tyr Leu Phe Ala
      50
      55
      60
Lys Tyr Pro Lys Lys Thr Glu Gly Asp Met Ser Lys Leu Arg Ser Met
      65
      70
      75
      80
Ile Val Arg Glu Ser Leu Ala Gly Phe Ser Arg Phe Cys Ser Phe
      95
      100
      105
      110
Asp Ala Tyr Ile Lys Leu Gly Lys Glu Gin Lys Ser Gly Gly Arg
      125
      130
      135
      140
Arg Arg Asp Thr Ile Leu Gly Asp Leu Phe Glu Ala Phe Leu Gly Ala
      145
      150
      155
      160
Leu Leu Leu Asp Lys Ile Asp Ala Val Arg Phe Leu Lys Gin
      175
      180
      185
      190
Val Met Ile Pro Gin Val Glu Lys Asn Phe Glu Arg Val Lys Asp
      205
      210
      215
      220
Tyr Lys Thr Cys Leu Gin Glu Phe Leu Gin Thr Lys Gly Asp Val Ala
<table>
<thead>
<tr>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile Asp Tyr Glu Val Ile Ser Gly Ala His Ala Lys Glu 165</td>
</tr>
<tr>
<td>180</td>
</tr>
<tr>
<td>190</td>
</tr>
<tr>
<td>Phe Glu Val Ser Ile Val Val Asn Gly Ala Val Ser Lys Gly Leu 195</td>
</tr>
<tr>
<td>200</td>
</tr>
<tr>
<td>205</td>
</tr>
<tr>
<td>Gly Lys Ser Lys Leu Ala Gly Ala Asp Ala Lys Asn Ala Leu 210</td>
</tr>
<tr>
<td>215</td>
</tr>
<tr>
<td>220</td>
</tr>
<tr>
<td>Ala Glu Leu Ser Glu Val Gly Ser Val Ser Lys Gly Glu Glu Asp Asn 225</td>
</tr>
<tr>
<td>230</td>
</tr>
<tr>
<td>235</td>
</tr>
<tr>
<td>240</td>
</tr>
<tr>
<td>Met Ala Ile Ile Lys Glu Phe Met Arg Phe Lys Val His Met Glu Gly 245</td>
</tr>
<tr>
<td>250</td>
</tr>
<tr>
<td>255</td>
</tr>
<tr>
<td>Ser Val Asn Gly His Glu Phe Glu Ile Glu Gly Glu Gly Glu Gly Gly Arg 260</td>
</tr>
<tr>
<td>265</td>
</tr>
<tr>
<td>270</td>
</tr>
<tr>
<td>Pro Tyr Glu Gly Thr Glu Thr Ala Lys Leu Lys Val Thr Lys Gly Gly 275</td>
</tr>
<tr>
<td>280</td>
</tr>
<tr>
<td>285</td>
</tr>
<tr>
<td>Pro Leu Pro Phe Ala Trp Asp Ile Leu Ser Pro Glu Phe Met Tyr Gly 290</td>
</tr>
<tr>
<td>295</td>
</tr>
<tr>
<td>300</td>
</tr>
<tr>
<td>Ser Lys Ala Tyr Val Lys His Pro Ala Asp Ile Pro Asp Tyr Leu Lys 305</td>
</tr>
<tr>
<td>310</td>
</tr>
<tr>
<td>315</td>
</tr>
<tr>
<td>320</td>
</tr>
<tr>
<td>Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val Met Asn Phe Glu 325</td>
</tr>
<tr>
<td>330</td>
</tr>
<tr>
<td>335</td>
</tr>
<tr>
<td>Asp Gly Gly Val Val Thr Val Thr Glu Ser Ser Leu Glu Asp Gly 340</td>
</tr>
<tr>
<td>345</td>
</tr>
<tr>
<td>350</td>
</tr>
<tr>
<td>Glu Phe Ile Tyr Lys Val Lys Leu Arg Gly Thr Asn Phe Pro Ser Asp 355</td>
</tr>
<tr>
<td>360</td>
</tr>
<tr>
<td>365</td>
</tr>
<tr>
<td>Gly Pro Val Met Glu Lys Thr Met Gly Trp Glu Ala Ser Ser Glu 370</td>
</tr>
<tr>
<td>375</td>
</tr>
<tr>
<td>380</td>
</tr>
<tr>
<td>Arg Met Tyr Pro Glu Asp Gly Ala Leu Lys Gly Glu Ile Lys Glu Arg 385</td>
</tr>
<tr>
<td>390</td>
</tr>
<tr>
<td>395</td>
</tr>
<tr>
<td>400</td>
</tr>
<tr>
<td>Leu Lys Leu Lys Asp Gly Gly His Tyr Asp Ala Glu Val Lys Thr Thr 405</td>
</tr>
<tr>
<td>410</td>
</tr>
<tr>
<td>415</td>
</tr>
<tr>
<td>Tyr Lys Ala Lys Lys Pro Val Glu Leu Pro Gly Ala Tyr Asn Val Asn 420</td>
</tr>
<tr>
<td>425</td>
</tr>
<tr>
<td>430</td>
</tr>
<tr>
<td>Ile Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr Thr Ile Val Glu 435</td>
</tr>
<tr>
<td>440</td>
</tr>
<tr>
<td>445</td>
</tr>
<tr>
<td>Gln Tyr Glu Arg Ala Glu Gly Arg His Ser Thr Gly Met Asp Glu 450</td>
</tr>
<tr>
<td>455</td>
</tr>
<tr>
<td>460</td>
</tr>
<tr>
<td>Leu Tyr Lys Asp Arg Pro Ala Ala Thr Lys Lys Ala Gly Glu Ala Lys 465</td>
</tr>
<tr>
<td>470</td>
</tr>
<tr>
<td>475</td>
</tr>
<tr>
<td>480</td>
</tr>
</tbody>
</table>

Lys Lys Lys

---

<210> SEQ ID NO 50
<211> LENGTH: 1423
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 50

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp 1

5

10

15

Tyr Lys Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val 20

25

30

Gly Ile His Gly Val Pro Ala Asp Lys Lys Tyr Ser Ile Gly Leu
<table>
<thead>
<tr>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Ile</td>
<td>Gly</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Val</td>
<td>Pro</td>
</tr>
<tr>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Ile</td>
<td>Lys</td>
</tr>
<tr>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Ala</td>
<td>Glu</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Arg</td>
<td>Lys</td>
</tr>
<tr>
<td>115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>Ala</td>
<td>Lys</td>
</tr>
<tr>
<td>130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td>145</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Val</td>
<td>Asp</td>
</tr>
<tr>
<td>165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Arg</td>
<td>Lys</td>
</tr>
<tr>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Tyr</td>
<td>Leu</td>
</tr>
<tr>
<td>195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Glu</td>
<td>Gly</td>
</tr>
<tr>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td>225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Ala</td>
<td>Ser</td>
</tr>
<tr>
<td>245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ser</td>
<td>Arg</td>
</tr>
<tr>
<td>260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asn</td>
<td>Gly</td>
</tr>
<tr>
<td>275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Asn</td>
<td>Phe</td>
</tr>
<tr>
<td>290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Lys</td>
</tr>
<tr>
<td>305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td>325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Ala</td>
<td>Ile</td>
</tr>
<tr>
<td>340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ala</td>
<td>Pro</td>
</tr>
<tr>
<td>355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td>370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Tyr</td>
<td>Lys</td>
</tr>
<tr>
<td>385</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Ile</td>
<td>Asp</td>
</tr>
<tr>
<td>405</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Ile</td>
<td>Leu</td>
</tr>
<tr>
<td>420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Arg</td>
<td>Glu</td>
</tr>
<tr>
<td>435</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Pro</td>
<td>His</td>
</tr>
<tr>
<td>450</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Arg Arg Glu Lys Ile Glu
465    470    475    480
Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg
485    490    495
Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
500    505
Thr Pro Trp Asn Phe Glu Val Val Asp Lys Gly Ala Ser Ala Glu
515    520    525
Ser Phe Ile Glu Arg Met Thr Arg Phe Asp Lys Asn Leu Pro Asn Glu
530    535    540
Lys Val Leu Pro Lys His Ser Leu Tyr Tyr Glu Tyr Phe Thr Val Tyr
545    550    555    560
Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro
565    570    575
Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe
580    585    590
Lys Thr Asn Arg Lys Val Thr Val Lys Glu Leu Lys Glu Asp Tyr Phe
595    600    605
Lys Lys Ile Glu Cys Phe Asp Asn Val Glu Ile Ser Gly Val Glu Asp
610    615    620
Arg Phe Asn Ala Ser Leu Gly Thr Thr His Asp Leu Leu Lys Ile Ile
625    630    635    640
Lys Asp Lys Asp Phe Leu Asp Asn Glu Asn Glu Asp Ile Leu Glu
645    650    655
Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Glu Arg Glu Met Ile Glu
660    665    670
Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Arg Lys Val Met Lys
675    680    685
Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
690    695    700
Leu Ile Asn Gly Ile Arg Asp Lys Glu Ser Gly Lys Thr Ile Leu Asp
705    710    715    720
Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Arg Phe Met Gin Leu Ile
725    730    735
His Asp Asp Ser Leu Thr Phe Gly Asp Ile Gin Lys Ala Gin Val
740    745    750
Ser Gly Glu Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
755    760    765
Ser Pro Ala Ile Lys Gly Ile Leu Gin Thr Val Val Asp
770    775    780
Glu Leu Val Lys Val Met Gly Arg His Lys Pro Gin Asn Ile Val Ile
785    790    795    800
Glu Met Ala Arg Glu Asn Gin Thr Thr Gin Lys Gly Gin Gin Lys Gin Ser
805    810    815
Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
820    825
Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gin Leu Gin Asn Glu
835    840    845
Lys Leu Tyr Leu Tyr Leu Gin Asn Gly Arg Asp Met Tyr Val Asp
850    855
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
865    870    875    880

-continued
Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu 985
990 995

Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu 900
905 910

Glu Val Val Lys Met Lys Asn Tyr Try Arg Gln Leu Leu Asn Ala 915
920 925

Lys Leu Ile Thr Gln Arg Phe Asp Asn Leu Thr Lys Ala Glu Arg 930
935 940

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu 945
950 955 960

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser 965
970 975

Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val 980
985 990

Lys Val Ile Thr Leu Ser Lys Leu Val Ser Asp Phe Arg Lys Asp 995
1000 1005

Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala 1010
1015 1020

His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys 1025
1030 1035

Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys 1040
1045 1050

Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Glu Glu Ile 1055
1060 1065

Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn 1070
1075 1080

Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys 1085
1090 1095

Arg Pro Leu Ile Glu Thr Asn Gly Thr Gly Glu Ile Val Trp 1100
1105 1110

Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met 1115
1120 1125

Pro Gln Val Asn Ile Val Lys Thr Glu Val Gln Thr Gly Gly 1130
1135 1140

Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu 1145
1150 1155

Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe 1160
1165 1170

Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Ala Lys Val 1175
1180 1185

Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu 1190
1195 1200

Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile 1205
1210 1215

Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu 1220
1225 1230

Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly 1235
1240 1245

Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gin Lys Gly Asn 1250
1255 1260

Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala 1265
1270 1275

Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gin
<table>
<thead>
<tr>
<th>1280</th>
<th>1285</th>
<th>1290</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>Gln</td>
<td>Leu</td>
</tr>
<tr>
<td>1295</td>
<td>Gln</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td>1300</td>
<td>Val</td>
</tr>
<tr>
<td>Ile</td>
<td>Gln</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td>1310</td>
<td>Ser</td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td>1320</td>
<td>Tyr</td>
</tr>
<tr>
<td>Lys</td>
<td>Pro</td>
<td>Arg</td>
</tr>
<tr>
<td>1340</td>
<td>Glu</td>
<td>Gln</td>
</tr>
<tr>
<td>Leu</td>
<td>Thr</td>
<td>Asn</td>
</tr>
<tr>
<td>1355</td>
<td>Lys</td>
<td>Leu</td>
</tr>
<tr>
<td>Thr</td>
<td>Ile</td>
<td>Asp</td>
</tr>
<tr>
<td>1370</td>
<td>Arg</td>
<td>Lys</td>
</tr>
<tr>
<td>Ala</td>
<td>Thr</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td>1390</td>
<td>His</td>
</tr>
<tr>
<td>Ile</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td>1400</td>
<td>Ser</td>
<td>Gly</td>
</tr>
<tr>
<td>Lys</td>
<td>Ala</td>
<td>Gly</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1285</th>
<th>1305</th>
<th>1315</th>
<th>1320</th>
<th>1335</th>
<th>1345</th>
<th>1350</th>
<th>1360</th>
<th>1365</th>
<th>1375</th>
<th>1380</th>
<th>1390</th>
<th>1395</th>
<th>1405</th>
<th>1410</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr</td>
<td>Leu</td>
<td>Arg</td>
<td>Tyr</td>
<td>Val</td>
<td>Leu</td>
<td>Asp</td>
<td>Arg</td>
<td>Tyr</td>
<td>Ser</td>
<td>Val</td>
<td>Leu</td>
<td>Thr</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 51
<211> LENGTH: 202
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<400> SEQUENCE: 51

gaaagtgcgc ctcagaccccg ctcccccccct gctctttgct gtcacaaggag aatgggtctct 60
cactgtggsa ttctggacta cccctgggag ctcgcaacctg agggcaaggc ccccccactc 120
gcccgacctgc aagctttctat gaggctgggg agagggctac atgaggcttg aagaagaccc 180
tccctctag gcacacacag aagtgtggag gttgggtgtt ctacagcgcg gtctctggtgc 240
ccccaggata gctgtggctg ccaggcactg ctctctgat aaaaaaccac ccctctgat 300
gaaacctatgc ccctctctgc tctctgtatg gaaagagcag tcggggctgca cgcgtggggt 360
gtctctcctt tagctcctgt gggagactat gggacccct gcaggtgggtc atagctcttc 420
tcattatct ctccatcata ctcggggaag aacggtatgt gatctctctat ctgaaactc 480
gtacagcgcg atgtctgcag gtcgctcgtc tacactcttg ctgggtctgg 540
cggggtaag acagatctag atgcacagca gttcttgag cttgtttgag aagagcagcc 600
gaaaaaaccc ccctctctctg gtcctcttgac tcctctttct gtcctcctc gcctctcttg 660
gatctggaga cccagctgg cagctgtgtca gaggggcccc ccgctttggg ccccaaaccc 720
tatagctag cctctctcct ataggactg cagctcagca tgcgtgctga ggccccagtg 780
ggtctctaggg gggcctctgtc agttctctag ctcgccccct ccctctctgg cccaggtgaa 840
gtctctagctc cagacacggc gagacagatgt cccagctggc aggctgtggag aggagggcc 900
tagctcctag cagcagcaag agggctccct ctcacatcact ccgggtgctc ccggccctga 960
gccgggcaag gggcgaagac ccagctgtca ctcgctcag ccagctgtca gggcggccaag 1020
acccacaccc ccagcggggc ggtctgtctg tctctctgtgc cggccctgag ccggctgca 1080
acgctcagct ctgcccagc ctcggtgtcgg ccggcaagcg ccggcgggaa 1140
acggctcagc cccggtgctc ccctctgctg gatggcttggc ctcgggctgg 1200
<210> SEQ ID NO 52
<211> LENGTH: 1153
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 52

M IV Lys Arg Pro Ala Ala Thr Lys Ala Gly Glu Ala Lys Lys Lys
  1 5 10 15
Lys Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
  20 25 30
Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
  35 40 45
Arg Ile Phe Pro Ala Ala Gin Ala Glu Asn Asn Leu Val Arg Arg Thr
  50 55 60
Asn Arg Gin Gly Arg Arg Leu Ala Arg Lys His Arg Arg Val
  65 70 75 80
Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
  85 90 95
Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gin Leu Arg Val Lys Gly Leu
 100 105 110
Thr Asp Glu Leu Ser Asn Glu Leu Phe Ile Ala Leu Lys Asn Met
 115 120 125
Val Lys His Arg Gly Ile Ser Tyr Leu Asp Ala Ser Asp Asp Gly
 130 135 140
Asn Ser Ser Val Gly Asp Tyr Ala Glu Ile Val Lys Gin Asn Ser Lys
 145 150 155 160
Gln Leu Thr Lys Thr Pro Gly Gin Ile Gin Leu Glu Arg Tyr Gin
 165 170 175
Thr Tyr Gin Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys
 180 190
Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu
| Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr  |
| 210 215 220 |
| Asp Glu Phe Ile Asn Arg Tyr Leu Glu Ile Leu Thr Gly Lys Arg Lys  |
| 225 230 235 240 |
| Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg  |
| 245 250 255 |
| Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile  |
| 260 265 270 |
| Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser  |
| 275 280 285 |
| Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr  |
| 290 295 300 |
| Val Pro Thr Glu Thr Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile  |
| 305 310 315 320 |
| Ile Asn Tyr Val Lys Asn Glu Ala Met Gly Pro Ala Lys Leu Phe  |
| 325 330 335 |
| Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly  |
| 340 345 350 |
| Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala  |
| 355 360 365 |
| Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp  |
| 370 375 380 |
| Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu  |
| 385 390 395 400 |
| Arg Glu Gly Ile Gln Glu Ala Glu His Glu Phe Ala Asp Gly Ser  |
| 405 410 415 |
| Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn  |
| 420 425 430 |
| Ser Ser Ile Phe Gly Lys Gly Thr His Asn Phe Ser Val Lys Leu Met  |
| 435 440 445 |
| Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Gln Met Thr  |
| 450 455 460 |
| Ile Leu Thr Arg Leu Gly Gln Lys Thr Thr Ser Ser Ser Asn Lys  |
| 465 470 475 480 |
| Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro  |
| 485 490 495 |
| Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala  |
| 500 505 510 |
| Ile Lys Glu Tyr Gly Asp Phe Asn Ala Val Ile Glu Met Ala Arg  |
| 515 520 525 |
| Glu Thr Asn Glu Asp Glu Lys Ala Ile Gln Lys Ile Gln Lys  |
| 530 535 540 |
| Ala Asn Lys Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln  |
| 545 550 555 560 |
| Tyr Asn Gly Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys  |
| 565 570 575 |
| Gln Leu Ala Thr Lys Ile Arg Leu Trp His Glu Gln Gly Glu Arg Cys  |
| 580 585 590 |
| Leu Tyr Thr Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser  |
| 595 600 605 |
| Asn Gln Phe Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp  |
| 610 615 620 |
| Asp Ser Leu Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn Gln Glu  |
Glu Glu Glu Leu Phe Arg Phe Leu Ser Arg Thr Met Pro Lys Gln
1055 1060 1065
Lys His Tyr Val Glu Leu Lys Pro Tyr Asp Lys Gln Lys Phe Glu
1070 1075
Gly Gly Glu Ala Leu Ile Lys Val Leu Gly Lys Val Ala Asn Ser
1085 1090 1095
Gly Gln Cys Lys Lys Gly Leu Gly Lys Ser Asn Ile Ser Ile Tyr
1100 1105 1110
Lys Val Arg Thr Asp Val Leu Gly Gly Gin His Ile Ile Lys Asn
1115 1120 1125
Glu Gly Asp Lys Pro Lys Leu Asp Phe Lys Arg Pro Ala Ala Thr
1130 1135 1140
Lys Lys Ala Gly Gin Ala Lys Lys Lys Lys
1145 1150

<210> SEQ ID NO 53
<211> LENGTH: 340
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<400> SEQUENCE: 53

gagggcctat tccttcctat ttgctataac gatacaaggg tgttagagag 60
ataattgaa ttaatgctac ttgtaaactc ccgctagact 120
aagtaataa ttcctggtta cttgctgct gtaaatataa ggcgtctgtag 180
agtgcttcg ccagcttgctgc ttgctctgcc tcgctgctgc 240
cgaaccattc attaatgtaa attaatgtaa attaatgtaa 300
acacacccga aacggtttttt cggattttaa 340

<210> SEQ ID NO 54
<211> LENGTH: 360
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (288)...(317)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 54

gagggcctat tccttcctat ttgctataac gatacaaggg tgttagagag 60
ataattgaa ttaatgctac ttgtaaactc ccgctagact 120
aagtaataa ttcctggtta cttgctgct gtaaatataa ggcgtctgtag 180
agtgcttcg ccagcttgctgc ttgctctgcc tcgctgctgc 240
cgaaccattc attaatgtaa attaatgtaa attaatgtaa 300
n nn nn nn nn n nnnnnnngttt ttagagctat cttggtttttc atcggtcccaaa aaccttttttt 360

<210> SEQ ID NO 55
<211> LENGTH: 219
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<220>  **FEATURE:**
<221>  **NAME/KEY:** modified_base
<222>  **LOCATION:** (250)..(269)
<223>  **OTHER INFORMATION:** a, c, t, g, unknown or other

<400>  **SEQUENCE:** 55

gagggctat ttcccatgat tcttcaatat ttcgcatatac gatacaaggg tgttagagag 60
atattggaa ttactttgcc tgtaasacca aagatatag taaacatatg tgtcagtaga 120
aagtaaatatt tcttgggta gtttgcagtt tttaaatatt gttttaaat aggatcatcat 180
atgtctacag taaacctgaa gttttgctgt tcttggcttt tatatatotat gttgaaagga 240
cgaacacn nnmmnmnnnn nnmmmmnmnnn ttttagagct agaaatagcag agttaaaata 300
aggtgtacct gttttt 318

<210>  **SEQ_ID NO:** 56
<211>  **LENGTH:** 325
<212>  **TYPE:** DNA
<213>  **ORGANISM:** Artificial Sequence

<220>  **FEATURE:**
<221>  **NAME/KEY:** modified_base
<222>  **LOCATION:** (250)..(269)
<223>  **OTHER INFORMATION:** a, c, t, g, unknown or other

<400>  **SEQUENCE:** 56

gagggctat ttcccatgat tcttcaatat ttcgcatatac gatacaaggg tgttagagag 60
atattggaa ttactttgcc tgtaasacca aagatatag taaacatatg tgtcagtaga 120
aagtaaatatt tcttgggta gtttgcagtt tttaaatatt gttttaaat aggatcatcat 180
atgtctacag taaacctgaa gttttgctgt tcttggcttt tatatatotat gttgaaagga 240
cgaacacn nnmmnmnnnn nnmmmmnmnnn ttttagagct agaaatagcag agttaaaata 300
aggtgtacct gttttcattt ttttt 325

<210>  **SEQ_ID NO:** 57
<211>  **LENGTH:** 337
<212>  **TYPE:** DNA
<213>  **ORGANISM:** Artificial Sequence

<220>  **FEATURE:**
<221>  **NAME/KEY:** modified_base
<222>  **LOCATION:** (250)..(269)
<223>  **OTHER INFORMATION:** a, c, t, g, unknown or other

<400>  **SEQUENCE:** 57

gagggctat ttcccatgat tcttcaatat ttcgcatatac gatacaaggg tgttagagag 60
atattggaa ttactttgcc tgtaasacca aagatatag taaacatatg tgtcagtaga 120
aagtaaatatt tcttgggta gtttgcagtt tttaaatatt gttttaaat aggatcatcat 180
atgtctacag taaacctgaa gttttgctgt tcttggcttt tatatatotat gttgaaagga 240
cgaacacn nnmmnmnnnn nnmmmmnmnnn ttttagagct agaaatagcag agttaaaata 300
aggtgtacct gttttacact gttaaaagtg tttttt 337

<210>  **SEQ_ID NO:** 58
<211>  **LENGTH:** 352
<212>  **TYPE:** DNA
<213>  **ORGANISM:** Artificial Sequence
gagggcctat ttcccatgt tacctcatat ttgcatatat gatacaaggcg tgtagagtga 60
atattggaata ttaatttacaa tgaatattcattacatctgcttggtt 120
aatatttattttattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>215</td>
<td>216</td>
</tr>
</tbody>
</table>

-continued

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>caacgaacag gogaactcg ggcgtagcga tcggcccttg ggcagaatga tcgaagttcgcg</td>
<td>1440</td>
<td></td>
</tr>
<tr>
<td>gggcactctc ctgatcgaag ggcagaactga cccggcaaac aagcagcttg acaagctgtt</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>cattcacaagtt gtggacacct acaacatcagt ttggctggaag aacccacatca aogccacggg</td>
<td>1560</td>
<td></td>
</tr>
<tr>
<td>ctgtgcgccgg aagggactctc tgcgtgacag aatgagcaga agcagacggc gggaaatcct</td>
<td>1620</td>
<td></td>
</tr>
<tr>
<td>gcattggcagc agagcaactcgg aagcagacct ggacaaactg tggcccaaga tgcggcagca</td>
<td>1680</td>
<td></td>
</tr>
<tr>
<td>gtcgtagctg ccgcggcacaact ctgacgacagc gccgcagttc ggagcgcact</td>
<td>1740</td>
<td></td>
</tr>
<tr>
<td>gcggctgcag gacacagca ccggacagct gagcactcag tcggcagtaa ctgtgaccgc</td>
<td>1800</td>
<td></td>
</tr>
<tr>
<td>gcggacagc tctgattgcag cccgcaacga acggcagcggc tcgggacagat</td>
<td>1860</td>
<td></td>
</tr>
<tr>
<td>gcggactgcg aacacagaga ccggcaggc gcccctgtg aacctcagta tcggagata</td>
<td>1920</td>
<td></td>
</tr>
<tr>
<td>cgagcagcgc caacgccccg tcggccctgc gggaggtccttt gcgggagcag aagcaggttt</td>
<td>1980</td>
<td></td>
</tr>
<tr>
<td>caagctacca cggattctcg tcgggacaggg cagaagggtag tggaggtccgg tgggaggg</td>
<td>2040</td>
<td></td>
</tr>
<tr>
<td>caacggaggta aagctctgta ccggagcagag aagcagaccttg gtcggagcaac aagcagacct</td>
<td>2100</td>
<td></td>
</tr>
<tr>
<td>ggatctgtaagcc atgcgagctg atggctgtcag aagcagacctg gtcggagcaac</td>
<td>2160</td>
<td></td>
</tr>
<tr>
<td>agacacacgg ctcgagctgc ctcgagctgc ctcgagctgc ctcgagctgc ctcgagctgc</td>
<td>2220</td>
<td></td>
</tr>
<tr>
<td>ggagggagt ttttttaact cctctcaggg caacaggggg gagctcttgaa ggacagagag</td>
<td>2280</td>
<td></td>
</tr>
<tr>
<td>tctcctgact cccgtactcgg gcggcgccctg gggagggaggg aacacagcat gtctctgtagt</td>
<td>2340</td>
<td></td>
</tr>
<tr>
<td>gcggaccagag cggaggggg ccggatcctg cttgagacgg ggagcagctgg tggcacaaggg</td>
<td>2400</td>
<td></td>
</tr>
<tr>
<td>gcggcctgcg cgcagatcat ctcgagcttg ggcagagcttc ggcgcaca gttacaagggc</td>
<td>2460</td>
<td></td>
</tr>
<tr>
<td>gaagggctgc cccggcacaact cggctgcctag cggctgatct ccggcagctg ataagggctgc</td>
<td>2520</td>
<td></td>
</tr>
<tr>
<td>caagatggaa ccggagcagag aagcagaccc agctctgtaac aagcagaccc</td>
<td>2580</td>
<td></td>
</tr>
<tr>
<td>aaagggcctc tggctgctgc acggagcagag aagcagaccc acggagcagag</td>
<td>2640</td>
<td></td>
</tr>
<tr>
<td>agagacatgc tcggagacgg cggagctagtt gcaggtctgtt gaagcttcttt ggagggaggt</td>
<td>2700</td>
<td></td>
</tr>
<tr>
<td>tcgggcagc gctcgttcgc gcaccacacc gcagcttctgt aacaatcttca gcacagagag</td>
<td>2760</td>
<td></td>
</tr>
<tr>
<td>ctcctgccgc ctcggagacag aagcagacac tgcggagacagt tgcggtaggg cctgagccact</td>
<td>2820</td>
<td></td>
</tr>
<tr>
<td>gttgcgaggag cggagagactg tcgggagacag gcctgagaaac ccgacccaccc tcgggagact</td>
<td>2880</td>
<td></td>
</tr>
<tr>
<td>caaatgtagg aacgacgtgg aagggcggagc atacaccggc tggggcgagg tggcagggag</td>
<td>2940</td>
<td></td>
</tr>
<tr>
<td>gttgacacgc gcggcagccag gcagacgctg acggagctggct gcggagctggct tgcgacgctg</td>
<td>3000</td>
<td></td>
</tr>
<tr>
<td>cgacagcttc gcggagcagag aatctctgtca gcggagctgc gcggagcagag</td>
<td>3060</td>
<td></td>
</tr>
<tr>
<td>aggaggagac ctcgagctgc cgggttcctgc tcgggagctgc cgggtgttcgg gcgggagctgc</td>
<td>3120</td>
<td></td>
</tr>
<tr>
<td>caattctgggc gcggagcagag cccagagcct gcagctgtag ggacagctg gcagctgtag</td>
<td>3180</td>
<td></td>
</tr>
<tr>
<td>cagacgcttgga aaggtgattgg gcgggagacc gcggagacac tgcgagctg aagggcggag</td>
<td>3240</td>
<td></td>
</tr>
<tr>
<td>agagagacac gcacccagag aaggggacag gcacccacag gcacccacag gcacccacag</td>
<td>3300</td>
<td></td>
</tr>
<tr>
<td>agaggggact aagagaggtgg gcagcgcaag ctcggagacag ggagctgggg aacacagagg</td>
<td>3360</td>
<td></td>
</tr>
<tr>
<td>gttctgagac gcgagagactg aatctctgta ctcggagact gcggagctgg</td>
<td>3420</td>
<td></td>
</tr>
<tr>
<td>cagagatcgc gcaccccagag cggagagctgc gcagctcttcgc gcagctcttcgc gcagctcttcgc</td>
<td>3480</td>
<td></td>
</tr>
<tr>
<td>ctttctgag gcagcttccag ctcggagacag gcgggagctgc gcgggagctgc gcgggagctgc</td>
<td>3540</td>
<td></td>
</tr>
<tr>
<td>cagagagacag aaggtgactgc ccggagagat gcggagagat gcggagagat gcggagagat</td>
<td>3600</td>
<td></td>
</tr>
<tr>
<td>gttgagacgc gcgaagactgc gcggagagat gcggagagat gcggagagat gcggagagat</td>
<td>3660</td>
<td></td>
</tr>
<tr>
<td>agaggggtctg agagagacag ataagggcgg gcggagagat gcggagagat gcggagagat</td>
<td>3720</td>
<td></td>
</tr>
<tr>
<td>gcagatcgc gcacagagtt gcagaggtgct gggagcccgg atgcagcttg</td>
<td>3780</td>
<td></td>
</tr>
</tbody>
</table>
gaatgacaag ctgatcgggg aagtgaaagt gatcaacctg agtteccaagc tgggtgcaga
3840
tttccggaag gattttcctg tttcaaaagt ggcgcagatc aacaaactcc accacegcaca
3900
cgagcgotac ctgaacgcccct tctggggaac ogtcctgata aaaaaagtac ccataagctga
3960
aaagccggtgg gcctgacgac atctaccaagt gtaacagcgtg cggagagtga tgcacaagag
4020
cggacagaa atoggoaaggt ctaaccgccaa gtaaccttct tacaagccaca tcattgaaacct
4080
tttcacaagc gagattaccct tggcccaacgg cggatcggg aagcgccctct tgatcggagac
4140
aaagccggaac cccgggagaaa tctgtgccgaataagccggag gatattggccca ccttgccgaa
4200
agtgtctgaacctg atgcccacaag taaaatctgt gaaagagcc caggtgcagca cagggcgctt
4260
cagccaaagc tcattccctgc ccaaggggaa cagccaaaag ctgatcgcagc gaaagagacc
4320
cctggaaccct aagagaagcg gcggcctccaa cagcccaccac ggctgcttctc cttggtgtgt
4380
gttgcccacaa gttgaaagg cgaactcctca aaaaagtag gttggtgaag gcggcctccaa
4440
gatccacac atggagagaag gacggctgca gaaatcccttc atgcagctttc tggagaacccaa
4500
gggtcataac gaaagtaaaa aagaccctgt ctaaagctctt cctgggctcc
4560
gttggagagc gcgcggagagc gaaatcccttc tggagagacag gaaagtagcttt
4620
actgcggctct cctcctcaaat atggggatatct ccgatcgaattc gocagccccact atggagagct
4680
gaagggctcc cccgggagag gatgacggttgta agtgggagac gaaagagacc
4740
cctggagagcg atcatacgagc agatcagcagc gtttctcaagc agatcagcagc gttgagagcg
4800
taatctggaa aagaggtcgag gcggcataac ccaagggcog gatgagcgagctc tggagaacccaa
4860
ggaggcagatc atccaccaagc tgggagactg gacagacggtt ggggccccag gcggcctgctc
4920
gttgttggagc gcgcggagagc gaaatcccttc tggagagacag gaaagtagcttt
4980
caccctgaca gcgggagaagc tggagagacag gaaagagccct ccgtttctaatct
5040
ggaggcagac ttttttttttt tcgccggtgc ggtccttttc aggagctcctg caagggctta
5100
a 5101

<210> SEQ ID NO 60
<211> LENGTH: 137
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 60

nnnnnnnnnn nnnnnnnnnnnn gttattgtaa tctcaagatt tagaaataaa tctttgcaaa
60
gctacaagc taagggctca tggccgaact caccctctgt catttttagg caggggatttt
120
togttattta attttttt
137

<210> SEQ ID NO 61
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 61

nnnnnnnn nnnnnnnnn gttattgtac ttcagaat gcagaagcta caagataag 60
gttcatgcc gaaatacaaca cccgtgcatt ttatggcagg gtgttttctg tattttattt 120
ttt 123

<210> SEQ ID NO 62
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide

<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 62

nnnnnnnn nnnnnnnnn gttattgtac ttcagaat gcagaagcta caagataag 60
gttcatgcc gaaatacaaca cccgtgcatt ttatggcagg gtgttttctg tattttattt 110

<210> SEQ ID NO 63
<211> LENGTH: 137
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide

<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 63

nnnnnnnn nnnnnnnnn gttattgtac ttcagagaatt tagaaatatag ttcagcagaa 60
gtacataata tgaaggctca tgcggaatcc aacacgctgt cattttatgg caggggtttt 120
tcgttattta atttttt 137

<210> SEQ ID NO 64
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide

<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 64

nnnnnnnn nnnnnnnnn gttattgtac ttcagagaatt gcagaagcta caagtataag 60
gttcatgcc gaaatacaaca cccgtgcatt ttatggcagg gtgttttctg tatatttatattt 120
ttt 123

<210> SEQ ID NO 65
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide
continuation

ggcaagacc aaccaggagga ttctctatgtg tacctgaagga agctgcgggc cgagttcgag 1200
gggccgacct aacttttcgaga aaaaaatcagc cgcagagatt tcctggcggaga gcagggcagcc 1260
ttcgacaagc gcagacccca ctctcagcagc catctcgggg aaatgctgggc cactcttggac 1320
aagccggcag agttggcaccacct atccttgccgca aagacaacaa aagcgatgcca gaagactctg 1380
acocctcggca ctcctctctca cgtgggcccc otgcccagag gcacacagcag ttttcttgg 1440
tccctccgga agccgagcgtgca cccctgccact tcggagccctg tgtgcgcccacaa 1500
gatgcacagc ccagagccttt catcaaccccg atgcacagct atgcacgctgaa ctgtccccag 1560
gaaagctgctc gcgggtggtc gacatcctgc tacagacact ctaatgctgta taacgagctg 1620
acccagaagc cgctttatcgg caggtcatctg cgggactacct aacttctggca ctcctcaacag 1680
atgccaggttc agcagccctg ctaagctttatgc aatgggcttgcg gcagcagccag ctaagagcctt 1740
cagtcaccact cccaggtgctg cacatcctac gacgctgctgc acaattatccg cacgacgacaa 1800
ccccgagagc accegccccag atcgaaga ccctgccccag cctgaccact caagcagccgctt 1860
ggctgtctga ccaagccagcag ccatcctccag cttcagcagct ctgctgggagg cagagcttttc 1920
gcttcgacgg gcagagcagttc cgcagcctgt gctggcagctg ctgagaaact gttggaatg 1980
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2040
cgagagcagc gccagcagcag gggcagcttg gtgaggtgca cgagcagcagc gctgggagcct 2100
gacatcggcct gccaagctgg aatcttgctgtg cggagccttg gtgacagcagc gctgggagcct 2160
aagctgctgcg aaagctgctgc gctgaagcctg ctgctgggagc gtcagcagctg ctgctgggagc 2220
gctgctgcc gcagcagcagc gcagcagcagc gcagcagcagc gcagcagcagc gcagcagcagc 2280
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2340
gcttcgacgg gcagagcagttc cgcagcctgt gctggcagctg ctgagaaact gttggaatg 2400
ggctgtctga ccaagccagcag ccatcctccag cttcagcagct ctgctgggagc gctgggagcct 2460
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2520
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2580
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2640
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2700
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2760
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2820
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2880
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 2940
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3000
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3060
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3120
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3180
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3240
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3300
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3360
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3420
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3480
agctggagag cccagcgggca ccagccgccc cagagcagcg caacagcagct ctgctgggagc 3540
aagaaaaaga tcacaaacgt gctggaattt caggagatct cttatctgga cccgatcac 3600
taccgggaag ataaagctgaa ctttctgtcg gaaaaaggtc acaagagcat tgaagctgatt 3660
atcagctgc ttaaagctc octgctaaga cttgagctcaag gtctcagac gatgctgcgc 3720
toctcctcgt ccacaaacaa caagcgagggc gagatccacaa agggaaaccac gttcttctcg 3780
agggagaat tttggtcaact gctgtaaccc gcccaagcgcg ttcocaaacc ctaaagatgcag 3840
aaccacccga atacagctgaa aacaccacaa aagagattttc aggaagtttt ctactacactc 3900
cggagttcag accagaatct ttggtcgcag aacagacacg gcaaatgcggc gacacccgcgc 3960
ttcagagct gcggagagaa cagcagatcgac gagggtgaca gtctcttcct cggcctacc 4020
gccagcgcg gcagagggact gtttggagctg acttcagcgag gtctcgcggc gcagctttgag 4080
ttctcggggg tgaatcacc caccgtcagc gactacaccc cttcctagct cgtgaaagac 4140
gccacctcga ctccacagcg cttgcaggc ctttggagga cccggtcagc cttgctcaag 4200
cgctggcgcg cagaggtgcgc tgggtcgtct aagagacgtg gtaagcttta gaagaaaaaa 4260
taa 4263

<210> SEQ ID NO 68
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 68
ttctagcaggg atttctgata ttactgtaac gttttagaga ttactgtgtaa gtaa 53

<210> SEQ ID NO 69
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 69
gtgcagttcag cttgcagcag gtttggagaa cccttaaaaagc agc 53

<210> SEQ ID NO 70
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 70
gggatctgcag cttgtaggta gcac 25

<210> SEQ ID NO 71
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 71
gcccatcagca gggaccttt ggtgc 24
<210> SEQ ID NO: 72
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 72
nnngacgag gcaatgctg aaatc 25

<210> SEQ ID NO: 73
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 73
nnnntattt ggtctatatt tgctg 25

<210> SEQ ID NO: 74
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 74
ctttaccca atgcctgca aagac 25

<210> SEQ ID NO: 75
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 75
casasatctct ctgttcttttt gctttttccc ataaacct cctta 45

<210> SEQ ID NO: 76
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 76
aggtttttag ggagaaagc aaagaagact agaatatttg atacc 45

<210> SEQ ID NO: 77
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 77

cctacggtgc ataagaatc aa tttcc

<210> SEQ ID NO 78
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 78

tggctcgaatt tcaagccattg c

<210> SEQ ID NO 79
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 79

ctttgaacg gcaatggtctg aatctgagcc aanaagcgc aag

<210> SEQ ID NO 80
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 80

ctttgaacg gcaatggtctg aatctgagcc aanaagcgc aag

<210> SEQ ID NO 81
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 81

ctttgaacg gcaatggtctg aatctgagcc aanaagcgc aag

<210> SEQ ID NO 82
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (36)..<(36)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 92
ctttgacgag gcaatggtgct gaatcgagcc aaaaangcgc aag 43

<210> SEQ ID NO 83
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (37)..<(37)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 93
ctttgacgag gcaatggtgct gaatcgagcc aaaaangcgc aag 43

<210> SEQ ID NO 84
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<222> LOCATION: (38)..<(38)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 94
ctttgacgag gcaatggtgct gaatcgagcc aaaaanggc aag 43

<210> SEQ ID NO 85
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<222> LOCATION: (39)..<(39)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 85
ctttgacgag gcaatggtgct gaatcgagcc aaaaaagnc aagaag 46

<210> SEQ ID NO 86
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<222> LOCATION: (40)..<(40)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 86
ctttgacgag gcaatggtgct gaatcgagcc aaaaaagcm aagaag 46

<210> SEQ ID NO 87
cttgcagcag gcatacgagcc aaaaaagcgc nagaag
46

gcgtttttt gcgtcgattt cag
23

caatgcgtga aatcgagcca aaaaaagcga ngaagaatc
40

caatgcgtga aatcgagcca aaaaaagcga naagaasatc
40

caatgcgtga aatcgagcca aaaaaagcga aagagaatc
40

caatgcgtga aatcgagcca aaaaaagcga anaagaatc
40

caatgcgtga aatcgagcca aaaaaagcga aagnagaatc
caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc
cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc

caatggtgta aatcgagcga aaaaaagcga agangaaatc

cgaatgatgta aaatcgagcga aaaaaagcga agangaaatc
OTHER INFORMATION: a, c, t, g, unknown or other

SEQ: 96
caatggtgta aatcgagcga aaaaaagca agaagaac nacc

SEQ: 97
caatggtgta aatcgagcga aaaaaagca agaagaan cacc

SEQ: 98
caatggtgta aatcgagcga aaaaaagca aagaagatn aaccagc

SEQ: 99
caatggtgta aatcgagcga aaaaaagcga aagaasatc naccagc

SEQ: 100
gacccctcgca tggacacc cacaaccctg g

SEQ: 101
caatggtgta aatcgagcga aaaaaagcga aagaasatc naccagc
<400> SEQUENCE: 101
aattccagg gttgagggtg taccgatgga g

<210> SEQ ID NO 102
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 102
catgatatct attcttaaat acataaasat atgg

<210> SEQ ID NO 103
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 103
catgaattca aactaacaag ttcgagtgtg ctg

<210> SEQ ID NO 104
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 104
aaacattttt ttccccattta ggaaaaaagga tggtg

<210> SEQ ID NO 105
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 105
aaacagcat cctttttctct aatgagaa aasat

<210> SEQ ID NO 106
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 106
aaaccttttaa tcgctcaca aatgcagcaaa aatgg

<210> SEQ ID NO 107
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 107

aaacaaatttttgctgctatttggagctgattttttaag

<210> SEQ ID NO: 108
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 108
aaactttttcaatacagcaccaatctgcatttatttg

<210> SEQ ID NO: 109
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 109
aaacaaatttaagcagataggctgtatgatgaaaa

<210> SEQ ID NO: 110
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 110
aaactctgctcaagaatttctcgttaaagaaaagttctg

<210> SEQ ID NO: 111
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 111
aaacatacgatatctacgtaaattttctgtatctagga

<210> SEQ ID NO: 112
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 112
aaacaaatctctcactaatatctctgc

<210> SEQ ID NO: 113
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 113
aaacaaatctctccttgaactaatatctctg
<210> SEQ ID NO 114
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 114

aaacagaga ttttaagga aaccttgag agatt

<210> SEQ ID NO 115
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 115

aaacgccttc gtcaggaag agctatgtc ggtgg

<210> SEQ ID NO 116
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 116

aaacactca agcatagtt gttctgagg atggc

<210> SEQ ID NO 117
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 117

aaacatotct atacctattg aatttcttt gtttg

<210> SEQ ID NO 118
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 118

aaacatgcgt gtatagtcgc aaaaaaaaaa cttcg

<210> SEQ ID NO 119
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 119

aaacgaggg ctggtttttgc ggaactatcag acgtta
<210> SEQ ID NO 120
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 120

aaacaagag agttcgagca agtaattact tttag

<210> SEQ ID NO 121
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 121

aaacaagaga ataatgacttt ggctggccctt ccgat

<210> SEQ ID NO 122
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 122

aaacaagagtg gatacgaaaa gtaagtcac ataag

<210> SEQ ID NO 123
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 123

aaacattatt gtcattacttt gttcgcatacc atcct

<210> SEQ ID NO 124
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 124

gagacctttg agttcgagag actggctctca gttttgggac cattcaaaac ag

<210> SEQ ID NO 125
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 125
tgagacctgt ctggagaagct cagaggttctc gtttttagac ttattttg

<210> SEQ ID NO 126
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 126

aaactctttt acgcagcgcg gagttggtt ttttg

<210> SEQ ID NO 127
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 127

aaacaaaaa accgaactcc gcgcgta aagta

<210> SEQ ID NO 128
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 128

atgcgggtac tgccgggct ctggcggtat tacgaatctat cctctg

<210> SEQ ID NO 129
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 129

gtgaactggcg atgtgtcgcgg aatggaagat cacactactt tcctt

<210> SEQ ID NO 130
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 130

taagaaata acottcactt aasaataact tcaagtcacct cctagctgac

<210> SEQ ID NO 131
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 131

attgatttga gtcaagtaagg aggtgactga agtatatttt aagtaaag
<210> SEQ ID NO 132
<211> LENGTH: 85
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 132
gagaccttgc agctccgag actggtctca gttttggac cattcasaac agcategtc     60
taasaactcg tagactattt tgtc                      85

<210> SEQ ID NO 133
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 133
gagaccaagtc tcgaagctc aaaggtctcg ttttagact atgtgttttt gaatgttccc     60
aaaaactctcg cacaactgaga cttg                         84

<210> SEQ ID NO 134
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 134
agtcacctca gcacaattg g                           21

<210> SEQ ID NO 135
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 135
cgtgtaaat cgataaagtc tccaagttgaa g            31

<210> SEQ ID NO 136
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 136
tgctctttct caacaaacag gg                      22

<210> SEQ ID NO 137
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 137
<210> SEQ ID NO 138
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 138

gtagctatt cagtcctaag gg

<210> SEQ ID NO 139
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 139
cgtttgtga actaatgggt gcacaatagg aaccttcac cc tgcag

<210> SEQ ID NO 140
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 140
cgtcaggaga agatctgtaa ttggcctcca ttagcctaac aacag

<210> SEQ ID NO 141
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 141
gatatttagg agcctatatg tgtggtttt tggcataaa actatag

<210> SEQ ID NO 142
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 142
catataggttt ttagcctaa aaccacaaaa aataggctcc ataatatc

<210> SEQ ID NO 143
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 143
attatcttt aataactaa aataggg
<210> SEQ ID NO 144
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 144
cgtgtaaat tgctagcgtcgcc

<210> SEQ ID NO 145
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 145
gcaccgcgtgactagtcctgag

<210> SEQ ID NO 146
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 146
cotaggaactgtagacgggtgcaaatatat gacccaaaataaatat

<210> SEQ ID NO 147
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 147
gccgtacytcgagcattgta cagctttgggt tgaactatgg ggtgcc

<210> SEQ ID NO 148
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 148	tccaatatcttcc acctttgat tctcc

<210> SEQ ID NO 149
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 149	cocaattttt cagtttaactg aastacactacc agccatcactgcctcc
agacgattca atagacaata agg

gttgggag cactcaaacc agcgatgatc taaaccaacg tagac
gctagtctgt tttgaagtct cccaaaccga ttatattaac acacgaggtg
gctagtctgt tttgaagtct cccaaaccgc acccattagt tcaaccaacg
aattcttttc ttcatatcg gtc

aagaagaat gaagatgtt catg
> SEQ ID NO: 156
> LENGTH: 25
> TYPE: DNA
> ORGANISM: Artificial Sequence

FEATURES:
- OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 156

```
ggtactaatc aaaaatgtga ggagg
```

> SEQ ID NO: 157
> LENGTH: 23
> TYPE: DNA
> ORGANISM: Artificial Sequence

FEATURES:
- OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 157

```
gtttttcaaa atctgcggtt gcg
```

> SEQ ID NO: 158
> LENGTH: 26
> TYPE: DNA
> ORGANISM: Artificial Sequence

FEATURES:
- OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 158

```
aaaatgtgae aaaaatggtgg aasccac
```

> SEQ ID NO: 159
> LENGTH: 53
> TYPE: DNA
> ORGANISM: Artificial Sequence

FEATURES:
- OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 159

```
atttcgatca cgtataggt tcttttttac aatttgggac cattcaaaac agc
```

> SEQ ID NO: 160
> LENGTH: 53
> TYPE: DNA
> ORGANISM: Artificial Sequence

FEATURES:
- OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 160

```
AAAAAAagaa accgataccg tttacgaat gtttagagc tattctgttt tga
```

> SEQ ID NO: 161
> LENGTH: 53
> TYPE: DNA
> ORGANISM: Artificial Sequence

FEATURES:
- OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 161

```
aaacggtatcc gttttctttt aaatcaatt gtttgggac cattcaaaac agc
```

> SEQ ID NO: 162
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 162
aattgaattt aaaaaagac gataccgttt gtttagagc tattgcttt tga 53

<210> SEQ ID NO 163
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 163
gttctttaaa ccaaagctg atcggtttct ttaaattc 39

<210> SEQ ID NO 164
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 164
gaaccgata cggtttggct ttaaggaaca ggttaagggc atttac 47

<210> SEQ ID NO 165
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 165
cgatttcagc catggcctcg tc 22

<210> SEQ ID NO 166
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<220> FEATURE:
<221> NAME/KEY: modified base
<222> LOCATION: (39) (33)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 166
gcttttgac aggcaatgcg tgaatcgyun nnnaaaaagc gcaagaagat atcaac 56

<210> SEQ ID NO 167
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 167
tcggtaaacc ccacaacoct gctagcgcag gttttggac cattcaaaac agc 53
<210> SEQ ID NO 168
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 168

gtcatagc agggtgtgg gttgtagga gtttttagc tattgttgg tga

53

<210> SEQ ID NO 169
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 169

tttgccttc ttc
ttc

23

<210> SEQ ID NO 170
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 170

cagggtgtg ggtggttgcg atggagttaa ctcccatctc c

41

<210> SEQ ID NO 171
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 171

gggagttac tccatgcga caacccgaa cctgtctag t g

41

<210> SEQ ID NO 172
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 172

gtggtagctg tctgtagtg ggc

22

<210> SEQ ID NO 173
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 173

ttaccgaaac ggaatttata tgc

23
<210> SEQ ID NO 174
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 174

aaagctaggttccgcaattgg

<210> SEQ ID NO 175
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 175

gtggggtgtaaggattgagttacctcctactcttc

<210> SEQ ID NO 176
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 176

getgggaggtgacctacccgtacaccccaacccctg

<210> SEQ ID NO 177
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 177

gtttcaccatttgacgacctgataatgaccacatgaagatag

<210> SEQ ID NO 178
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 178

gtggtcactagtgcagcattttgagtattaggatgtatggtggtgatg

<210> SEQ ID NO 179
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 179

ctgattttgatgacatttgctgacacccgctccttc
<210> SEQ ID NO: 180
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 180
 gcataatgct tcaaaatgaa tcaaatcag tgsaactcg  40

<210> SEQ ID NO: 181
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 181
 gttttgggac cattcaaaaac agcatagct ctaaaagtga cagtaatato ag  52

<210> SEQ ID NO: 182
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 182
 gttttagac tatgctgttt tgaatggctc caaaaagctc actagcaggg ttg  53

<210> SEQ ID NO: 183
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 183
 atactttagc cagcgcgag ttccggttttg taggaaggtg aagtataca cagatcat  59

<210> SEQ ID NO: 184
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 184
 gtcacagc atgggttcgg ttggtagcag  33

<210> SEQ ID NO: 185
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 185
 ctctgacag atttctgaaa ttacttgtatg  33

<210> SEQ ID NO: 186
tttaaagaa aaccgataccg tttacgaaat tgg

ggaaaccttc ataacagccg agcaagttat aataaggctc gtcctttatc aaccttgaaaa

agtggcacccg agtcggtgtc tttt

gttatagacg tatctgtgta tgaatggttcc caaaac

ggaaaccttc ataacagccg agcaagttaa tataaggctc gtcctttatc aaccttgaaaa

agtggcacccg agtcggtgtc tttt

gttatagacg tatctgtgat tgaatggttcc caaaac

ggaaaccttc ataacagccg agcaagttat aataaggctc gtcctttatc aaccttgaaaa

agtggcacccg agtcggtgtc tttt

<210> SEQ ID NO 191
<211> LENGTH: 103
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 191

nnnnnnnn nnmnnnnnn gtttagagc tagaaatagc aagttaaat aaggctagtc 60
cgttaroac tggaaaaagt gcacccagttt gcgggttttt ttt 103

<210> SEQ ID NO 192
<211> LENGTH: 103
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) . . . (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 192

nnnnnnnn nnmnnnnnn gtttagagc tagaaatagc aagttaatat aaggctagtc 60
cgttaroac tggaaaaagt gcacccagttt gcgggttttt ttt 103

<210> SEQ ID NO 193
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) . . . (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 193

nnnnnnnn nnmnnnnnn gtttagagc tatgtgttt tggaacaaa acagcatagc 60
aagttaaat aaggctagtc cgttatcaac tggaaaaagt gcacccagttt gcgggttttt ttt 123

<210> SEQ ID NO 194
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) . . . (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 194

nnnnnnnn nnmnnnnnn gtttagagc tatgtgttat tggaacaaat acagcatagc 60
aagttaatat aaggctagtc cgttatcaac tggaaaaagt gcacccagttt gcgggttttt ttt 123

<210> SEQ ID NO 195
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 195

gtaacctca atgactaggg 20
<210> SEQ ID NO 196
<211> LENGTH: 4203
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 196

atggcccaagaagcaagcgc gaacgtcgggt atccacaggc acaccaacgc gcacacaag  60
tacctgcct ggccacact cgagcccaac tcctggtggt gggcgctgat caccgaacgc 120
tacaaggtgc ccaagcaaga atccacaggt cgtacaccaac cccacggcaca cagcatcaag 180
aacgacctga tcggagccct gctgtgagcc ageggygga aaagcgagac gcacccgctg 240
aagcggtgc ccagacaccc atacaccgca cggagacacc ggtacctgta tgtcagagag 300
atcctcagca acagagatgcg caaggggagc gacaacgttc tctcagacag ggaagacttc 360
ttcctgtgcgg agaggtgatgc gaaacagccca cctccagccca ccagacatgg caaagccagc 420
gaggtggcct acacgggaag ggaccacacc atctacacag tggcaagaag aacctgtggcc 480
agcaccagca aggcgcaccc gcctgctgatct atcttcgccc tcgcccacat gacaaagctc 540
cgggccactc tacctgacga ggccagccct acacccgacac agacagacgt gcacacagtg 600
ttcactacag gtcgctgcga ctacacacag gctcggagcc aaaccccaac ccacacacgc 660
ggcttgcaag ccaagcggtc cctgctggtc gcacacgacc ggaccagcaag gcagctgaac 720
cggggtgcc gcacagcgcg gagaagagac aagggcgtgc tcggaaaccc gattccagcc 780
agcgtggccg tcaccccaac cttcacaagc acctccgacg ggcggcagga tggccacatct 840
cagctgcagc agacacctcc agacccagac gccacagcaac gctgctgccc gatgctgccc 900
cagttcgcag acctttctgc ggccagccag acacgctccg ggcctcctag gctcttctat gatcagcg 960
atcctgagag tgaacacagg gcacacacag gcccctcgtga gcgccttctac gtcagcagc 1020
tacgagcag accacccagc cttcagcttc gtaaagcttc tcggctgcgg caagctgcctg 1080
gaseactac aagccatgtt cttccagccag agcaagaccc gtaaccgctgc ctacatgtcc 1140
cccggacagc gcaacaggaag gttctcagaag ttcatacag gcacccgcttg gaaagctccc 1200
ggcaccacag aatcagctgc gcagacagac gccacagcgg tctgctgccc gcacagagcc 1260
ttcaccagc gcagcatcc ccaaccagtc cccctgagag gatcgcagct ccatccggag 1320
cggcagcag atttttccac attctcagag gacacgccaag aacagctgga gaaagactct 1380
acctccgca tctccactct cgtccggcct cttgcccaggg gaaacagccag atggcctctg 1440
atcaccagca agacacggag aaccccaccc cccctgactc tcggaggaag gttggacacg 1500
gggcctcgc cagcctggtt gacgacgccg gtaaaccact tcgataagaa gttgcctac 1560
gagaaggtgc tgcccaaccg cagccctgtcg taccggcttc acatcgtctt taacagcgctg 1620
acccactgta atacggtgac ccagggagat ggcaagcccg cttcctcggag cggcgagcag 1680
aaaaagccga tcgggtcact gcctttcacg accacccgaa aagtccagct gcacacgctg 1740
aaagagctg actccacgaag aatcagtgct gcagccctgc tggaaatctc cgccgctgga 1800
gaccgtcaca cccgttgcct ggccacaccg cgcacatctg tggaaatatt caagccagca 1860
agctctcgg acaatagggag aacaagagatct cctgctgaat gatcgcagct gccctcggag 1920
cgtggtgagg cagacagcag tagcagggaa ggcgcgtgaa ccttgtggca cctgtggcag 1980
gacacagttc tgaagcgagc gaagccgagct cggcggccag gctgctgagc gctgtggcag 2040
aagctgatcc cgagcctccg ggacagcagc ttcgacagg cacatcggga tgtcttgcaag 2100
...
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 197

atgccctaa agaagaagcg gaagctccgt atccacgaag tccacacago cgaacaagag 60
taacagctcg gcggtgcacat ccagcaccaac atcgtgtggtc ggccagcgat caccaaagag 120
tacaaagctgc cagcagaagaa aatccagcgtt ctggccacaca cagaccgca gcaagtca 180
aagaacctgta tggagcgcct gcctgtccag acgcgctgaag caggccgccc caccggcctg 240
aagaagacg cagcagagaa atacacccga cgggaacac ccgatactgt ctgctcagag 300
atcttcacga acagatgggc caaaggagc gcacagtctt tcacagacat ggaagagtc 360
ttcctgtgag aaagaagataa gaagacccag cgccacccaa tcttgccgaa catgtgtgcag 420
aggtggctgt acaccgagaa gtaacacccac accttcacca ctggaagaaat actgtctgac 480
agccaggaca aagcgcacact ggcgctgtagc tattcgtggcc tggccacac acacaggttc 540
cggggcact ttcgctatgca gggagacgct aaccccgaca acacgagctg ggaacagtgc 600
ttccttcgac tggctgcacag ctcaacccag ctgctgtcagaaaaccccat cacgccggagc 660
ggctgtggac ccacagccga cctgtgctgac ccagctatga cgacagacagct gcgtggaat 720
cggatcgccc agcgtacctg cggaaagcaag aactggtcgt cgtaaaacct gattgccctg 780
agctggtgcgc taccgctcaca ctccagacgc aatccgacga aatacggcag ggccgcctacg 840
cagctgacga caagcaccaata cgcgcgcagcg tccggacacgc tgcctggcaca agatgcggac 900
caggctgacg cagctctttcc gcgcgaccaag aacgctgtcc acacgcacact gcgtgcggag 960
atctcttgagc taacacccaga gatcccaacgc gccccctgta gcgcctctac gatccagaga 1020
tacgacggg cgcggccggc gtctctcagc ctgatagcct cgacagcgtcg ctgctgctgtg 1080
gagagattca aagagatatttt cttcgccagcg caacgagaaag gtaacgcggct ctcacagctt 1140
ggccgagcc gcggccagaa gtttctcaagt tctccgagc cccatctgga aagatggcag 1200
gcgcagcccc aatgtagcgg cagcgggaaac ccgtctgcag cgtgcgcaag ccggcctggac 1260
tgcgagctgg gcgcctcag cacccgacgc ccaactctgca ctcctgggag aagtagcgtgc ctctcgctgg 1320
cggccagca atttttacgcc atctctgtgac gccacgcttg aaaaatgagca gaagattcgg 1380
acctctcgcg ccacccactg cgtggggcct cgctcccggc gcacagcagcg tctggctgctg 1440
atgccacga agacagaggg aacacccagag cccctggcagct ctcaggaagt ggtgcagcag 1500
ggcgtctccgc ccagacgcttt ctcagagcag atgcagacagt ctctagggaa cctgcggcatac 1560
gagaaggtgct ggcacccagca cagccgtgtgt taagctatcc caagcgtgta taacagcgttg 1620
accacagtga aaatcgtgac cgagggatag aagagcgcgg cctctcctgag ggggagccag 1680
aagggcgcgc ggcgcgctcg gctgtcctgg gcaccggagaa aattgacggc gaaggtcgtct 1740
aagagacgtct ccctccagag aatcgtggtgc tccgctctcag gcgtcaccct gccgctgaaa 1800
gatcggcotta acgcgtccct gcgcacccatc gcacagctgtgc tcgaaattta caagccgcaag 1860
gactctctgg aacaaagggc aacacggagaccttccagagt aatctgtgag ctgcttgacca 1920
cgtgtttagg aagagagatag gtcgcgagaa cgcgtcggaa cctctgcoca ctgttctgcac 1980
gacaaaggtct gtagccgtgt ggaagcgcgg aatacagccg gctgctcgcc acgtgcggccg 2040
aagctgacga ccggcctcag gcgcgagcag ctcgctgagg cctccttgaga aatacctgttc 2100
tgcgaggtc tcgcaccaag aaacctctag ctgcctgtcc gazgcacagag ccgctcctttt 2160
aaagagacca tccsagaagc ccaggtcctc gcgcagcggcg atagcgtcga cgacgacattt 2220
-continued

gcctatcgg cggcgcgcc cgccattaaag aagggcatac tgcagacagt gaaggtggtg
2280
gacagcctg tgaagaggtg ggggcggccac aagcggacag cagatgtctg gaaatatgccc
2340
agagagaacc agagccccc agagggcag aagacacgcc gcgcgagaaat gaagggcggac
2400
gagaagggcca tcaagagctgt ggcagcggcag atctgtaaag acacccgcgt ggaaacaccc
2460
cagctgcaag agagagatgt gtaacctgtaa tacctgcaaa agtggcggtga taaggttgct
2520
gaccagggcacttgaaccaag ccctggctgc gctctgctgat tcctagaacgt cgtgctgctg
2580
agctttctga agagcagctt cactgaaacag aaggtgctga ccagagaagc caagaaaccc
2640
gccaagagcga aacaacgctg ctcagcaagaag ctgctgaaaga gataagaaaa ctaagggccg
2700
cagctgtgga agcgcaagct gattaaccaag agaagtttgg acaaatctgc acaagcccgag
2760
agagggcgcct gtcagcaact gataagggcgc gcctttataca agagacagcct ggtggaaccc
2820
gccagatca aacagagctt gcgcagatcct ttggaccccc gattgaacca acaagtaacag
2880
gagatgaaaa agttgatctct ggagctgctg cagaagattc gatgtggttcc
2940
gttttcgcgc aggattttaca gttttacacaag ggagcgcgaca tcaaaccacat ccacacacgg
3000
cacgcccgcct ctcgtaggta cgtcgctgga aacccggctg tcaaaaaagta cccaaacctg
3060
gaaagagagt tgtagttcagcg gcaactacaag gtgtaacagcg ttcgctggag atcgcccaag
3120
agcgagagag aatccggcga gcctagcggc gaactcttcct ttcatacgaaag catcataggc
3180
tttttcaaga cgagattacc cttggcccaac ggcgcggatca gsgaagcgccc ttcgatcgaag
3240
aacaacgagcg aacacgggga gatggtgtaa gataagggcgc ggatttttggc caccggtggc
3300
aagatggctga gcctggccca aagtgaataat cggaaaaagga ccggagtgcga gcaagccggc
3360
ttcagcacag agttctactc gcgcagaggg aagcagacata agctgtgtagcg cagaaaaagaagn
3420
gatcggcgcac ctaagagatga cggcgcgttcc gcagccsccaa cgtgcgctctta tctgtgcgct
3480
gtggggcacc aggtgggaaa gcggcaactc agaaaaaga agaagatggtg gggatggacaag
gccaacctca aagtgaagagaaaaggcctag actacaagcg tccatttcctctccccctgt
3540
ggacggatat ggcagaagctg gcaagatgtct gcctgtgctgcc ggcagatgcga aaaggggggc
3600
aagggctcac aagtaaggttgg aaagacgctcg cagtcagcct cttggaagacc gcagggagac
3660
agaatggctttt gcgccgccccat ctaagagctgt gcctgtgctgcc gcagatggcg aagggggacc
3720
gaacgcggcgc ctcgctgccc attatgcggcg ttctgctgtaacc gcgcagccac ctatgaaagaag
3780
tcagaggtct ccccccagaga aagtggccatt gcctgtgctgct gcctgtggccca gcaagacaag
3840
tcagttgagcg agatctcagca gcctcttcga cggtttgcgta aagagatgtct cttgggcggac
3900
tcagataggg agaagactcc tgcagctcag tcaaacacgt cagaagagctg
gcagcgagagt gacccgacag tcggtgggtgc gcggagacgcgc ccgataggtg
tagcagtggtg ctggctgttg catctgctggt
gcgcccctg tctcagccag cttcagcggc ctgctggagcd ccgagatcag
tgggggggca aaaaaagcgg ccggggtgcc aaaaaagcgg gcggagccaa aaaaaaaaaa
4140
ctggggagcg aacaaaaagcc ggccccgcgcct aaaaaagcgg gcggagccaa aaaaaagcgg
4200
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 199

gagtcgcagc agagaagsg ggg  
23

<210> SEQ ID NO 200
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 200

gcgccccagc tcgatgtgt ggg  
23

<210> SEQ ID NO 201
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 201

ggggcacaga tgaagaactc agg  
23

<210> SEQ ID NO 202
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 202

gtacaaacgg cagaagctgg agg  
23

<210> SEQ ID NO 203
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 203

ggcsgagct ggagggggs ggg  
23

<210> SEQ ID NO 204
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 204

gagcccccttc tcctotgct cgg  
23

<210> SEQ ID NO 205
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 205

gggcaacca cacccacca ggg  
23

<210> SEQ ID NO 206
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 206

gctcccatca catcaacgg tgt  
23

<210> SEQ ID NO 207
gtggcgccttt gccacgaagc agg

<210> SEQ ID NO 209
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 208

ggcagagtgc tgttgggtgc tgg

<210> SEQ ID NO 209
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 209

gccctctggt ggcgcccaagc tgg

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

<400> SEQUENCE: 210

gagtggcccg agtccagcgtt ggg

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

<400> SEQUENCE: 211

ggcctccccca aagctggtgc agg

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

<400> SEQUENCE: 212

ggggaggaga tgtgggtttc agg

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

<400> SEQUENCE: 213

gtgccgagag ggcgcgagat tgg

<210> SEQ ID NO 214
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 214

gagtggccgc gaggcggggc ggg
<210> SEQ ID NO: 215
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 215

ggagtgccgc cgaggcgggg cgg

<210> SEQ ID NO: 216
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 216

ggataagag gcgcggcgaag cgg

<210> SEQ ID NO: 217
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 217

cocccocct cgtgtaatgt

<210> SEQ ID NO: 218
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 218

ggataatggc gcacgcgaga

<210> SEQ ID NO: 219
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 219

aacacgcac tcggtgccac

<210> SEQ ID NO: 220
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 220

tacctccac tgaactagggg
<400> SEQUENCE: 221
cagttgata acggactagc ct 22

<210> SEQ ID NO 222
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 222
agtccgacga gaagaagaag ttt 23

<210> SEQ ID NO 223
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 223
tttcaagtga ataaccgacct agcct 25

<210> SEQ ID NO 224
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 224
aaacagcag ttcgcctgga 20

<210> SEQ ID NO 225
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 225
tcaacctgctc gatgaagctc 20

<210> SEQ ID NO 226
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 226
tccaaactca aagtgyggcga 20

<210> SEQ ID NO 227
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 227

tgatgaccttttggtccc

<210> SEQ ID NO 228
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 228
gaggaatctttttgtgtaaatggttgaggttttttggag

<210> SEQ ID NO 229
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 229
gaggaaccttaaataaaaaacoacactactcacoacacccacoacacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoacoac
<211> LENGTH: 133
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (92)...(111)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 233

acctctagaa aaaaagcccc gacctggtgc caacttttaca aagttgataac ggacagct 60
tatatatta cgtatattctt cagctctaaa ccccccccccccccccccccccccccccccnnnnnnnn nngtggttcg 120
tcctttccac aag 133

<210> SEQ ID NO: 234
<211> LENGTH: 153
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (112)...(131)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 234

acctctagaa aaaaagcccc gacctggtgc caacttttca aagttgataac ggacagct 60
tatatatta cgtatattctt tttttttttt ttacacgcat aacagctctaaa cccccccccccccccccccccccccccnnnnnnnn nngtggttcg 120
nngtggttcg tcctttccac aag 163

<210> SEQ ID NO: 235
<211> LENGTH: 153
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (112)...(131)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 235

acctctagaa aaaaagcccc gacctggtgc caacttttca aagttgataac ggacagct 60
tatatatta cgtatattctt tttttttttt ttacacgcat aacagctctaaa cccccccccccccccccccccccccccnnnnnnnn nngtggttcg 120
nngtggttcg tcctttccac aag 163

<210> SEQ ID NO: 236
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 236

agcccccaagt ggtgtcttn aa 22

<210> SEQ ID NO: 237
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 237
acatcaaccgtgggcatnt

<210> SEQ ID NO: 238
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 238
aaggtgtggttccagaaacacnac

<210> SEQ ID NO: 239
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 239
ccatcacatcaacggttgagnag

<210> SEQ ID NO: 240
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 240
aaacggcagaagttgaggnnta

<210> SEQ ID NO: 241
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 241
gggcagagctggaggaggnntt

<210> SEQ ID NO: 242
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 242
ggtgtgttcagacacggtntc
<210> SEQ ID NO 243
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 243
aaccggagga caagtagctg

<210> SEQ ID NO 244
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 244
tttcagaacc gggagcaaaac

<210> SEQ ID NO 245
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 245
gtgtggctcc agaaccggan ct

<210> SEQ ID NO 246
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 246
tccagaacgc ggagcaaaac cc

<210> SEQ ID NO 247
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 247
cagaagcttg agagggaagn cg

<210> SEQ ID NO 248
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 248

catcaacgg tggcgcattn ga

<210> SEQ ID NO 249
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 249

gcagaagctg gagaggagaan gt

<210> SEQ ID NO 250
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 250

cctccctccc tgccccccaggn gc

<210> SEQ ID NO 251
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 251

tcactcgtgc cctccctcctn aa

<210> SEQ ID NO 252
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 252

gggaggacat cgatgtcacaat

<210> SEQ ID NO 253
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 253

canacgcssc aagcgtgacgn ac

<210> SEQ ID NO 254
<211> LENGTH: 22
<210> SEQ ID NO 255
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
  <221> NAME/KEY: modified_base
  <222> LOCATION: (20) . (20)
  <223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 255
  ggtgggcac ccacaacnn ag

<210> SEQ ID NO 256
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
  <221> NAME/KEY: modified_base
  <222> LOCATION: (20) . (20)
  <223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 256
  ggtgggcaac ccacaacccn ta

<210> SEQ ID NO 257
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
  <221> NAME/KEY: modified_base
  <222> LOCATION: (20) . (20)
  <223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 257
  gaagggcctg agtccgagcn tc

<210> SEQ ID NO 258
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
  <221> NAME/KEY: modified_base
  <222> LOCATION: (20) . (20)
  <223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 258
  caacgggtgg cgcattgacncn tg

<210> SEQ ID NO 259
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
  <221> NAME/KEY: modified_base
  <222> LOCATION: (20) . (20)
  <223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 259
aggaggaagg gcgtgagtcn ca

<210> SEQ ID NO 260
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 260

agctggagga ggaagggccn cc

<210> SEQ ID NO 261
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 261
gcattgccag ggaagggccn cc

<210> SEQ ID NO 262
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 262
attgcaagca agaagggccan cg

<210> SEQ ID NO 263
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 263
agaaccggag gacaaagtan ga

<210> SEQ ID NO 264
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 264
tcaaccgtg gcgtgagtcn gt

<210> SEQ ID NO 265
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: modified_base
gaagctgagg gaggsaggg gc

coaatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac

caataatgggga gacatatgtg tcacctcca atgactaggg tgggcaacca caaaccacgcg 60
agggcagagt gttcgttgtg gctggcagg cccccgcttg ggcccaagc ctgacttgcc 120
cac
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 270
ctcagcttt cccatcaggc ttcagcctca gcctgagtgt tgaggccccca gttggtgctc 60
tgagggcttc ctgagttttcatcgtgccc ctcctccccc tgtgccaggga ggaaggtgct 120
ttcaca 125

<210> SEQ ID NO 271
<211> LENGTH: 129
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 271
tcactgtgccc cctcctcccc tgtgaggtgtg gttccagagac cggagaca 60
agtcacaaac gcgagactgc gaggaggaag ggcctcagtc cgagcagag aagaggtc 120
cctcctcca 129

<210> SEQ ID NO 272
<211> LENGTH: 129
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 272
tcctcaagc taggggtggcc aaccaacaac ccaagagggc aagaggtctg ttcgcgcttg 60
cggagccct gctggggccc aagcgtcagc ctcggccacc ccttcggggc ctctctctcttg 120
gcctgag 129

<210> SEQ ID NO 273
<211> LENGTH: 127
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 273
tcgtctgctg ctcggccctg gcttcgcctg gcccaagcttg gactctgtgc ctcctctctc 60
cagcgttcttg gcggcgcttg agtcagggcc ccaagggcct tgaagccgg ggccccatt 120
gagag 127

<210> SEQ ID NO 274
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 274
gaaatta cgaactcactt taggg 25

<210> SEQ ID NO 275
<211> LENGTH: 126
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 275

aaaaaagcc gacttgtg gcacatcccc aagtgtagaa cggcattagc tttattttac 60
ttgctatccc tagcttcaaa acaacagca gcgtgacacc acctatatgt ggtgctatcc 120
aatttc 126

<211> SEQ ID NO: 276
<212> LENGTH: 126
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 276

aaaaaagcc gacttgtg gcacatcccc aagtgtagaa cggcattagc tttattttac 60
ttgctatccc tagcttcaaa acaacacat taatagcattg acctatatgt ggtgctatcc 120
aatttc 126

<211> SEQ ID NO: 277
<212> LENGTH: 4677
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 277

ttttttgc gcatacgac ttcagccaaa agttaggccc gctgtaggga cggctcccg 60
gcgtgctag caacacagat gatgtctgaa gcccccaag ctctcttggg gctgtcaggg 120
cgctctagt gcgtctcaag gggacgctg tttaatatgc cagggccccc ggtgcgaaga 180
cattatagc agctacccca gcatatccaa aacccattgc tcccattccgg gatctcag 240
gcgcctagc ctgcttgtac gatcctcga attgggggcc ttttctctcg ctttacctcc 300
acaacccgca aacatgtaac catagatgt tccagattac cttgcgcaga aaaaaaagcg 360
caacacgca gcgtgcacca agaagctcag cattggcctg gacatcgccca caaactctctg 420
ggctgtggccc gtgatacagc agaactacaa gttgccagag aaaaatcca agttgtgggg 490
caacacagac ggccacacca tccaagagaa cctgtacgga gggctgcttg gtaacagcgg 540
cgacacagcc ggcaccaccc ggcagagagc aacccagaga agaagataca ccaagcggas 600
gacccgagct gcctatcgtg aagagacttc cagcaacagag atggccacgg tggagcagc 660
cctttctcc agaacttgaac atcctctctg gttgcgaagg gataagagc agagcgggca 720
catatttc gcgtacagct ggagcagagt ggccttaccc gagacagctc cccctctcta 780
cacccgtgaa aagaasaggcc tgtgacagcccg gacacagcat gccttggccc tgttttactt 840
ggctcttgccc cacatcagct ggctggggcc cacttcctcttg atcgagggcc acctcagacc 900
cgacacagcc gcagtcctcg cccctctctct ccccttctctgg cagacactaca accagctgt 960
cggagaaac ccacacactc ccacgcccgct gggacgcaag gcacgtcccttg ctgcacact 1020
gagcagagcc aacgcggcct gcctttctgt cgcctcagct cccccagcaga aagaagattg 1080
cctgttcggc aacctgatgg cctgtgacct ggcccctgcacc ccaaccttca agacacactt
1140
cgacgctgcc gcggcagcga aacggaaggt gacgcaagcct acctcgacg aacgctgcctg
1200
cacgctctgc gcggacgacg cccaggacag ccacagcctg tctccggggc ccaacagcgtc
1260
gttgacgagc ctctgtggatcc cctcatcgtt ggtgctggag cgaacacgag agacactcctt
1320
cctggagacgc aagacagcag gaaggacagt cctaaatatg ccattacgagag cagagcctcag
1380
gcgtgctgct cgccagcgcg gccaccggttt gcctgcttctg ggtggtgactgtg ggtggacgtt
1440
ggagcggacg gcgaacgacg cagacagcag cagaacgatgc ttcagacacgt caagacacacg
1500
cagcctgatg aagagctggg cagctgtcgcg caacggtgag cggccacagc aaccgcttgtc
1560
ggagcacttc ctggaagga tagacgattcgagt ggtgacgcgtt ctagactcgtc ggcagcagcag
1620
gggagcggctc ccacgctgcc cagcagcag cagcagcag cgcagcagcag cggccacagcag
1680
cgggaaatcg cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg
1740
cggggagtcg cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg
1800
ggctgtgagt aagcagcag cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg
1860
gccgaggtgta atcggagtcg ctcgagtcg ctcgagtcg ctcgagtcg ctcgagtcg
1920
gtggcgtggag cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg
1980
gcccggccgg cgccagcagc agacacacgcc acctttgt cgctgtgccttt ggctgtgagtta
2040
cgagcctgcc cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg cagcgaaggg
2100
cgcgagcgttc ctcgagtcg ctcgagtcg ctcgagtcg ctcgagtcg ctcgagtcg
2160
tctgagacg aacgtgagc ctttgctctg gcctgtgagt aagcagcag cagcgaaggg
2220
gccatagtc tctgagacg aagcagcag cagcgaaggg cagcgaaggg cagcgaaggg
2280
gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc
2340
cgccgagcag cgcagcagc agacacacgcc acctttgc cgcagcagcag cgcagcagcag
2400
cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
2460
gcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
2520
gcggcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
2580
gacctgccct gctctgcctt gcctgctctt gcctgctctt gcctgctctt gcctgctctt
2640
cgagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
2700
caccgcgcgg gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc
2760
gagacacacgcc acctttgc cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
2820
gagacacacgcc acctttgc cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
2880
gcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
2940
gctctgctgt gcctgtgagc ctcgagtcg ctcgagtcg ctcgagtcg ctcgagtcg
3000
gctctgctgt gcctgtgagc ctcgagtcg ctcgagtcg ctcgagtcg ctcgagtcg
3060
gcggcgtggc gctgagtcg ctcgagtcg ctcgagtcg ctcgagtcg ctcgagtcg
3120
cattcgaagg cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
3180
tctggcgtag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
3240
ctctggggtgt ctacagcttg ctctgcctgt cctctgcctgt cctctgcctgt cctctgcctgt
3300
cgagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
3360
cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
3420
gcgagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag cgcagcagcag
3480
ctcttcctac agcaacatca tgaacttttt caagacccg ag tataaccttg ccacaggcga
3540
gatccggaag ggctcctgtg tcgagcaaa cggcgaacc ggggagatcg tgtggataaa
3600
ggccccggt actggtcgag ag ggtcagaggt ggtcagaggt gtcgtggtgtag
3660
aadagaacag gtagcaagag gcggcttccag caaagttct ctcgtgcocca aagagcaacag
3720
cggacagctg atgcggagaa agaagagatct atcgtgcocca aagagcaacag
3780
cccccatcggt gcctatcctg tgcctgggtg ggcacaattg gaaagggcga agtcacagaa
3840
actgcagag tggcaagaagc ctgctgggtg gccccatcag gaaagaacgta ggtcggcga
3900

gacctcact acgtctcgag gacgcaagag ctcaaacagaa gttgaaagg gacctgca
3960
caagctgctt aagatccctcc tgtgtcgact gggaaacggc cgggaagagaa tgtgcgtcctc
4020
tgcgcggggaa gtcgcagagg gaaagcaact gcgcctgcctc ctcctaatg ctcagactctc
4080
gtacctgacct gacgccatag agaagtgaag ggcctccccc cgggataatg agcagacaac
4140
gctgttttgtg gaacagcaca agcactactc ggcagcagc atcgcgactgt
4200
cctcaacagag gtgatccggt cccgatcagaa tgtggcacaag tgtgtcggct cccctcaca
4260
ggagccgggt gacccactcag gacagcaggg cagcactagt ttcagccgtc ggtactgac
4320
caactcggtgg gcccgttgca ccccttcagta cccttgacacc ccacagccacc cggagaggta
4380
cacccgacace aaagaggtggt gtcgcagccag ctcggctaccc aacccagaca ggccttctga
4440
cggagacagg atgcgcctgtgt ctcgcgctgg aggccagcag cccagagaag agagaaaaag
4500
gggagggacg aaagatctgg gcagcactgtg cgcccttgg gacgcaacaac atgcagccct
4560
ccctgtcctg tgtggggtgtag tcgcctctagtt gtcgttgtggt ggcacaaggg tcaaccgcgaa
4620
ccgatttgat cccgctttcgag cttctctcag cagatagtaa cgctgcttgta tgggtact
4677

<210> SEQ ID NO 270
<211> LENGTH: 3150
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<400> SEQUENCE: 278

tctctcctgc gctatgacac ttccagccaa agttagggcg gcgcggcaga cggtctcccg
60
gcggtcgctc caacacgctg gatgttcgca cccccgaagg cttctctggg gcggtcatggg
120
gcggctagct gcgcttccag gcgcggcctg tttataatacg cgagccccgg attgcaaaga
180
cattattgag aagctaccag gcatcctaca aacacatgaa tcacactaca cttcatacg
240
gcagctgctc tgtgtcgcag ccagcagctg aagagcgcac ccccttccttc cgttcagc
300
acaaccgaaga aacagcctcag aqagaagag aagcgctcta acagacaaaa aagtagccga
360
cagctttctc gcacagcaac tggctgtctat cccgttccac actctgctgtg accattgcgg
420
tgagccgtta gctatgacac aagttgcccct tgtgtcgtcag tgtctcagaa cgggtgtagc
480
acgcctcgc aagttttgctg aagatcgcact caaagctggt caggttgccgc atacgccgtc
540
cgcggcgccg ccctctctccca ctcagctttgtc cagttgcttcg atacgccgtc
600
tgagagttag aaagttcagcg gcggccagcg ccagacacggt ccctctttccat cggagaaaaa
660
cagctggcga acgctgtggt tggcgcaccat aaacacagaa tggccacagt ctcggtgtagc
720
tgacaacca aaggtttggc gcgcggcggc gcctctttccat ctcggtgtagc ggcggggggc
780
tcgcttcggt gtagcttggt ttcgctgctc gcggcgtttg ctcggtgtagc gcggcttcctc
840
aactcaacaag cgcctaggcc acgtctacaa gaaacacttt atgcaagttg tcagggctga 900
catgctctct aagggcttac gccgtggcga ggccgtggct ctgctggcata agaaacaactt 960
ttacacgta ggaagtaqeg gcatacgagat gcatacttgag tcaacacggag tggatagttt 1020
acacgcgacca aactctgcag tattaggctga agactgtgag acatcgcgac acgatagactga 1080
atacgctgg acgtacgcga cccgctggcag tcgctgtgct gcagtctcto ecatcggcga 1140
acagggtcga tggctcttca acgcctggac tgcgtatcact ggtgctgtgg attggctaat 1200
cgcttcgctg ctcgttgccgc tcgctcttgct tcaacagtaag aaacgactga tcgctacgga 1260
agaccttacg atgcctggag tgtcacaagac gattaacatt gcgcacacaa ccgcatggas 1320
aactcaacaag aaagctctag qgctcgccaa cgtactaccc aagtggaaga aagttctcggt 1380
cagagagctc cctgctagtg acgcgtgaaga acctcccgatt aacccggag aactagcasat 1440
geactctgg acgtcctcagc gcgtgagagcg tctgtgcgct gcagtctaccc gcaaggacca 1500
ggtcgaogaa ttcgctgga tcaacccgag tcgctatctg ggcgaaagcc ataaagtggg 1560
taaccctag gcattcctgt gctccattaca acgtgagctt ggcgctgctg tttacgtcgt 1620
gtcaagtgtc aacccggaag gtaacgcgt tatgacagat gcggccaggag cggcttcgggt 1680
taaccctagc gcagctacgt gcgtcagaac gcgcgcgcga aactctgcggg 1740
tgcggaagat gcctcgcttc ctcgactgact ccattcattg gggagaaacc aagagacatac 1800
caatgggtgc gcctatacog cccctgagcaa aaatagggtg gcgtcagcaag atcccgtcgtt 1860
tgcggtctct gggtctttgc ttcgcttcac gtcggctgct gcattgctaca ggtcttcagta 1920
tgcgctctc cggctgtttcg gtcctggcgc atcgcgactc tccgctgatt 1980
gotcggcgat gcagcttagtg cgcgcgcggct ttcacgtgct ctatagtcgg ggtgccaagg 2040
catctacggg attgtgtcgta aagagctgca caagcttaac gcagccagac ctcacatgga 2100
acacgactca acacagctgta cccgctgagg tgcctgcagct gttggaaaat ctggagaaga 2160
caagctggcc aactacacgaac ggtggcttga atgggtgtgtg taoxgtgtta ctgcaaggtt 2220
gactagagct ctaagatgta acgcggtgta cggctggcata ggtgctgccc gcgcgcgcgat 2280
agctgactgga gctacactac acggcgcgtag ggtgctgctg tggcttcggc ctcacaact 2340
gccagctcag gctgttggag acctgtggtaa ggtgctgttgaga cttgctggta ggtgactgg 2400
gggtagctgag gttgcagaca tggaaatggct taaagctgtt gcaattggag ggtgctgtgg 2460
gggcaagaag aagagacttg gagagacctct tgcgagacgt tgcgttcgag atttggtaac 2520
ttcagcttcg ttctggcttg ggcaggaatta caagaaacct attccagagc gttggatccc 2580
gctgtctctc gccgcgttcgg ctcctcagac cacaacaacc aacagcagat gtcgatgcgtg 2640
ttagctagg acacagagtgt ctggcaatgc tctcaacatt gttgagcagcc aagagcttcg 2700
ccacccctgg aagacattcg tagtggccaca gcagagttgt gcacccgaatt ctttcgact 2760
gttggacgcc ctcctcggga cagctggcagc aacgcgtgcag ggtgctgtgg atttggacat 2820
cgcaccaaat gttgctgggc ctcacatggct ctaaagtgtaa ctggatgcagc 2880
tggtgcagcg ctggagcaagtg ttcagcttcg ctccttgtggt ggtggtgctgg atgggtgaac 2940
tctggacctgc gttggatgcct ttagtttgga ggtgcacgact gcgtgctgtaa cgggacgaac 3000
tgggctgctg gttggacgcc aacagctggc ctcctctagt gtttgggatgt gcgtgctgtaa 3060
gatcttgtt gggtgccgac ggtgcacccag ccagagatgt atccagcttg tggctttcct 3120
tgtcagaaacg taacgtcagct tgcaggtact 3150
<210> SEQ ID NO 279
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)...(42)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 279

gaaattacta cgaacctca taannnnnnn nnnnnnnnnn ngtttttaga gctagaaata 60
gcaagtttaa ataaggtcag tcggttatca acttagaaaaa tgtgccaccca gtcggtgctt 120
ttttt 125

<210> SEQ ID NO 280
<211> LENGTH: 8452
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 280

tcgggtatct cacacccgaa cagggtgacaa tttctgagga aatgtgctgcg gaacccctat 60
tgttatttt ttctaaattc atcctaaatg atgctgctcg atggattatct caaaaaggt 120
tccctctcg atcttctttat aaaaaatg aagttttaaa tcaatctaa aatattatgg 180
gttaaatgtt tcgtcagcta ctgcaagcttg gtcggtcacta atacgttga caaatctctat cagcagcttg 240
tctatttctgt tccttctcact tcgcatctgtg cccgctgttg ctagataact cgtacacggga 300
ggcctaccata tgtgctcttt cttggtctgc gatagcggga gacccacgcc cacagcgttc 360
agacttctat ccataaaaccc gacccagccag aagggccgag cgacagattg gctctgcaac 420
tttctggcgc tcataactctg ctaatattgg tcaggggga gcctagttta gtaagtccgc 480
agttatagtct ttcggcaagg tttgttcgat tgcataaggc atcgtgggctc aacgccgtcg 540
gtggtaatgt tcctcattcct gtcgcgcttg ccacagatca aggcagatgt atttgatccc 600
cattgtgtgac aaaaaacgcc tgttccccgt tcgcttccc gatgctgctc aagtaagtt 660
gcggcgaaagt ttattctcctg agctgttgctg ctagataatt ggatctgggg 720
atctagagaa tggttcttcg tgaactgttga gtaactcacc cacaatattt taaatattg 780
tgcgggggc ccaagcgtgc ttttcgctttg gatgttaacc gataatagg gactttcgtc 840
cggacacatca ctcatttgaa aactcctctc agggccaaac ctggcataatgc 900
ctcctgtcc tcgtgactca cttggtcgtg cccactgaat ctagtttttt gacacccaact gatcagtag 960
atcttttact tcgacagccgg ttcttggggt gcacgggaaa aacaaaagc gaaagcctaa aagccgccaa 1020
aaagggaaat gggggccgaea ggaatacgtg aataactata cttctctcttct tcaaatatta 1080
tttgagcttt cttactctct gacccaaact cttactcgtg agtttcttgt 1140
ccattaagcg ccagccctgct taaaaagag cctagatctct tctggagata cttttttttttt 1200
gccctacatct tgtgtttgg gacaacaccc caacggtctgc ccaggttggtg tttggatgatc 1260
gatggatcg ttaccacacat tttttcgcgg gtaaatggc aacggcacag ccggatattc 1320
aatagtttt cttcattgtg agcagtttgc gggccacac ccgaagaact ctggatcagc 1380
gatcataatcctgcgctgtcgctagctggcttctggaagcttcagcggcttcagc 1440
gaagagttct acaagttcct caagccctac ctggaaangga ttgacggcac cgagaactg 3840
tctgtgaagc tgtaagccag gacgctgtcg ggaagcagcg gagaacctta caaagcagcc 3900
atcccccaac agatatcact ggaggacgct acagccctac ttgaggccccca ggaagatattt 3960	
taccctttggc tcaagccgaa attccgaccc tctgagccct cctaaccctcc 4020
tactacgtgg gtcctttgcgg cagggggaaagt aagagatctgc ctggatgac cagaaagag 4080
gagaaagacca tccacccttcg gaactccgag gaaatttggt ggacgggggc ttcgggcccc 4140
agcttcatcg agagctgagc caaacctgcg aagaacctgc ccacagagggag attcgctgcc 4200
aacacacggc tgcctgacga gttaaccctc ggtaataaag agtgagaaac 4260
gttgacgaggg gaaagagaaa gcccgccttc ctgacgagcg aagcagaaaaa ggcccattgctg 4320
gacgctgctgt tcaagccgaa cccgaagtttc acgctgaagc atcagtaagag aagactacct 4380
aagaaaaacg agacatcttc ggaagatact gttcgctgccc ttgactcgttg tagggcaga 4440
tccgctggcc cattacacagc tctggtgcga attatcaagg aacaaggttt cttgccacct 4500
gagaaagacgg gagacatttgg ctaagactgtttt ttgctgaagc gggcagcagc 4560
gagatgc tgtgagagctg gaaacaatttc cggcagatgt ttgacggaac aagtgaag 4620
cagctgacgc gggcagagttc caccgctgag ggacggctga cggcagagt gacacgcagc 4680
aatccggcaga agcagctgct tatcgcagccc tatttaagaa gggacatccag 4740
aaccagacat tacagcagctc gataacagcag gacagctgta ccatttaaaa gggacatccag 4800
aaagaacgac cttcctgcgt cttcctgcgt cttcctgcgt cttcctgcgt 4860
agcccggcctt aataggggct cttcctgcgt cttcctgcgt cttcctgcgt 4920
gttgatgggg ggcacagccag gagaacatct gtttagcagaa tggcagcaca gaaacaaggg 4980
accagaggg gacgagacag caccgctgag ggaagcagag ggaacagcacaa 5040
gactgcgctg gcccagatctc ctaagacgctg ccggagatgt aggccaagcagc 5100
aatccagcgg gggcagctgct atcgcacagc gactagcagc ggaacagcacaa 5160
atatcagcgg ggcagctgct attcctgcgt cttcctgcgt cttcctgcgt 5220
gactccatcg aacacagcagc gctgacagaaa cgggccgaca gggagcagcgc 5280
gttgccctcg aagaaacctc ggaagactact ggcagcagcagc gctgacagaaa 5340
aagactgatca cggcagaggg gttgacagca ctcacccaggg cggagagaggg cggctgagc 5400
gactggatag gacgccggtt catcagaga gactgctgtagc aacccgcccc gacacagaca 5460
cactgccgctc gtagcttgag ctcgctgagc ctcgctgagc ctcgctgagc 5520
atctgggagc tgaagagcatg caccctcagc ttgacgagttc ctcgctgagc ctcgctgagc 5580
tccagttttt aacagactgct cctgagcacag aacagactgct cctgagcacag 5640
ccagggctgg ctggagctgc cctgcattca aagctcagctt aggtgtgggc 5700
tacgctgacc gactaaggtgact gacagctgagc ctcgctgagc ctcgctgagc 5760
gggagcaggct gcgcagctgct cttgctgctctt gcgcagctgct cttgctgctctt 5820
atcctccctg cccaggccgag ctcagctgct cttgctgctctt gcgcagctgct cttgctgctctt 5880
gggagcaggct gcgcagctgct cttgctgctctt gcgcagctgct cttgctgctctt 5940
ccgccaggtg atacgctgaa gtagactgag gcgtcagagc gcgtcagagc gcgtcagagc 6000
atctgcccag gtagactgag gcgtcagagc gcgtcagagc gcgtcagagc gcgtcagagc 6060
aagtaagggactgcagagc gcgtcagagc gcgtcagagc gcgtcagagc gcgtcagagc gcgtcagagc 6120
gaaaagccccg ctcaccgctac actgagactg ctaagagagct gtcgctgggct cacacaagctg 6180
gaagaagaca gotccagaa gaatcccaact gacattctgg aagcccaaggg ctacaaagaa 6240
gtgaagaaag acctgatcct caaactgcct aagtactccc tgtcatggct ggaasaacggc 6300
ggaagaagaga tgtgtgctct tcggccggagc atggccaggg gaagaaccagc gcggctggcc 6360
tccaaatagct tgaatacctct gcattctggcc agacactatg aagaactgtaa gggtcctccc 6420
gaggtataact agccagaacac gctttggttg gaaagccaca agaactaacct gcagaagtact 6480
atgcacgacg ctcagcaggtct tccaaagaggtgtacocct cctgacgctct ggcgatctgac 6540
gtgcgcgtcct cccacaccaac gcacccggagt aagccacact gagaagacgac cggatctatc 6600
atccaaactct ttcctctgga aaactgtggga gccttcgtcgc ccttcattctc tctttcctc 6660
acacgcagcc ggaagaggtta caacacgaccc caaatggtgc tcgacaccc accgcactcc 6720
cagacgatca cccgcgctgta cggacacaggt atgcacccgtct cggccttggg cggacagagc 6780
ccccaaaggag aagaaaggttg gcggagccggc tgcctattgta ttgcgaattgc gttttgctgc 6840	
tcgacacttc ccaagccaaag gtagggcgcc tgggaaagctgctg ccctggctctg cggagatgca 6900
acacgctatc tgtagcagcc cccctacagct cctgggctgg cgctgagccgctctgcgag 6960
gtgcagggcg gacgcgctgct tataactggc ggcggcgggcag tgcacagaca tttactgag 7020
cctacaaacgcatctcagatc caactactcc actaacactct cttacagccgc cccgagctcgt 7080
tgtgagcctc ctcggcattg cgggctgttct ttcctctctctct gcgagcgtgaa 7140
cttgacacaa gaatccctgtct tcattctccgatcggttagatt ctcgaagctctggacgag 7200
cggggcgaag cggacagagct tgtgggagctg gatgcagaca gcagccggcg cccgacagcg 7260
agcgcgttctgc caagattccct gcggagccgctcg ggtcggcttg gcggagccgcttgccgctc 7320
tcgccggtct tcaacgagc tgcggtcgcgg cttccatccgt cagccggtgg 7380
cgcgcgctac cccgcgctgta cgggctgtgac ggtcggctgc gggccgggctg cgcggtggtcgtg 7440
agcgcgacac gcggcgccac gcggctgctgt cctgccggcg ccagacctgc gcggcgccgtgc 7500
aagttgtgccg cggagcggagcg ctggagccgctgctc gtaagctgctgctgagcgccttg 7560
cggggcgcct ggtgccgctcg ctcggcctcgc ggggaagcttc gcggccggcgc aaggagccgt 7620
gggtgccgctg ctcgcggagt caagccgagct gcggccggcgc tcggccggcagc 7680
agcgccgacgc gcagggtcggt cctgccggcg cggcgccgagt cggcgccgagt cggcgccgagt 7740
gcggtgcgtca cgggctgtgctgctc gagccgctgtc cccgcgctgtc cggccgctgtc 7800
tccggacagt cggacggagagcg cggccggattg aaggagttctc aagccgctgtc ctgggtgggct 7860
aacctgctgc gcggccgtgctc gcggccgctgtc gagccgctgtc cccgcgctgtc cggccgctgtc 7920
tggccggcgc gcggccgctcg ctcggcctcgc ggggaagcttc gcggccggcgc aaggagccgt 7980
acacgcggtc gacggctgtgc acggcggatgc tgggctactac gcgggctactac gcgggctgttc 8040
ccagctacag cccgctgcgag ctgcacactcga acgcctgcgtc gcggcgccgc gcagtcctggc 8100
cgcggctgtcg gcggccgctgtc cccgcgctgtc cggccgctgtc aagccgctgtc cggccgctgtc 8160
cctgcctctg ctcgagctgc gcgggctgttc gcggccgctgtc gcggccgctgtc cggccgctgtc 8220
tccgatcgc gcgggagcccgtgccttgctc cccgcgctgtc gcggccgctgtc cggccgctgtc 8280
cctgtaggtgc gtgggctgttc gcggccgctgtc gcggccgctgtc cggccgctgtc 8340
gtggtgctgtc gcggccgctgtc gcggccgctgtc cggccgctgtc cggccgctgtc 8400
tgatcgcctgc cttgcagccgca gcgtcgcggtc gcggccgctgtc 8452
ttttagac agtttagttgg tgaatggcc caaaccgaa gggctgagt cggacgacag
  60
gagaagttt tagagcagtg tcttgtggaa tggcccccac ac
  102
gagagacaa agtcaacag gcagaagctg gaggaggaag ggctgagtc cggcgaag
  60
agaagggct ccattcatc caaccgggtg cgcattgcca
  100
agctggagga ggaaggggct ggtcggagc agaagaagaa gggctccccc
  50
gaguccgagc agaagaagaa guuuagagc
  30
agctggagga ggaaggggct ggtcggagc agaagaagaa ggtcccccat
  49
agctggagga ggaaggggct ggtcggagc agaagaagaa ggtcccccat
  49
ctgagggaggaagggctgagctgccgacgaaagagaagggctceccatcacca

<210> SEQ ID NO 288
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 288
c tgagggaggaagggctgagctgccgacgaaagagaagggctceccatcacca  54

<210> SEQ ID NO 289
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 289
c tgagggaggaagggctgagctgccgacgaaagagaagggctceccatcacca  50

<210> SEQ ID NO 290
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 290
c tgagggaggaagggctgagctgccgacgaaagagaagggctceccatcacca  47

<210> SEQ ID NO 291
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<220> FEATURE:
<223> OTHER INFORMATION: Description of Combined DNA/RNA Molecule: Synthetic oligonucleotide

<221> NAME/KEY: modified_base
<222> LOCATION: (1) . . . (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 291
nnnnnnnnnnnnnnnnnnnnnnnnnnnnguauagc uagaaauagc aaguuaaau aagggtagtc  60
cguuuu  66

<210> SEQ ID NO 292
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 292
gaguccgaagc aagagaaga  20

<210> SEQ ID NO 293
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 293
gacaucaug uccuccocau  20
<210> SEQ ID NO 300
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 300

gaaagccucu ggccaggaa

<210> SEQ ID NO 301
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 301

ctggaagagg aagggcctga gtccgagcag aagaagaagg gtcoccat

<210> SEQ ID NO 302
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 302

gaguccgac agaagaagau

<210> SEQ ID NO 303
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 303

gaguccgac agaagaagaua

<210> SEQ ID NO 304
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 304

gaguccgac agaagaacaa

<210> SEQ ID NO 305
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 305

gaguccgac agaagaugaa

<210> SEQ ID NO 306
<211> LENGTH: 20
<212> TYPE: RNA
<210> SEQ ID NO: 307
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 307

gguccgagc agaagauagaa

20

<210> SEQ ID NO: 308
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 308

gguccgagc agaagauagaa

20

<210> SEQ ID NO: 309
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 309

gguccgagc agaagauagaa

20

<210> SEQ ID NO: 310
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 310

gguccgagc agaagauagaa

20

<210> SEQ ID NO: 311
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 311

gguccgagc agaagauagaa

20

<210> SEQ ID NO: 312
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 312

gguccgagc agaagauagaa

20
<220>  FEATURE:
<221>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400>  SEQUENCE: 312

gagacccgagc agagagagaa

20

<210>  SEQ ID NO 313
<211>  LENGTH: 24
<212>  TYPE: DNA
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400>  SEQUENCE: 313

aatgacacgc ttgctagcgg tggg

24

<210>  SEQ ID NO 314
<211>  LENGTH: 39
<212>  TYPE: DNA
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400>  SEQUENCE: 314

aaaaacggagc gcacagtc cgacgagag aagaagtttt

39

<210>  SEQ ID NO 315
<211>  LENGTH: 39
<212>  TYPE: DNA
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400>  SEQUENCE: 315

aaaaacggggc cgagattcgg tgtctcagggc aagagttttt

39

<210>  SEQ ID NO 316
<211>  LENGTH: 38
<212>  TYPE: DNA
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400>  SEQUENCE: 316

aaaaacggagc gcacagtc cgagcagag aagaagtttt

38

<210>  SEQ ID NO 317
<211>  LENGTH: 40
<212>  TYPE: DNA
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<223>  OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400>  SEQUENCE: 317

aacggagga gggtgagcag tcgagaactc agaggttttag

40

<210>  SEQ ID NO 318
<211>  LENGTH: 38
<212>  TYPE: DNA
<213>  ORGANISM: Homo sapiens
agcctctttcttcttgctg gactcaggcc cttcctcc

<210> SEQ ID NO 319
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 319
cagggagggga gggcagacaga tgagaacctc aggggcccc

<210> SEQ ID NO 320
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 320
ggcaatgcgc caccgyttga tggatggga gccccttctag gaggccccca ggcagcgc

tggggcctca aacctcaggc

<210> SEQ ID NO 321
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 321
gtcaacctcc aatgcataggg tgg

<210> SEQ ID NO 322
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6) (25)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 322
caccgnnnn nnnmnmnmn nnnn

<210> SEQ ID NO 323
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5) (24)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 323
aadnnnnnn nnnnnnnnn nnnnc

<210> SEQ ID NO 324
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 324

catcgatgctctcoccattggcgtgttctgg

<210> SEQ ID NO 325
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 325

ttcgtggcactgcgccacggtgtgtgatgg

<210> SEQ ID NO 326
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 326

tgtggcactgcgccacgttgatgtggtgg

<210> SEQ ID NO 327
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 327
tccagcctcgccttgatattaaccccttcgg

<210> SEQ ID NO 328
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 328

ggagggaggggacagatgtgaaactcagagg

<210> SEQ ID NO 329
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 329
aggggccagatgtgtgccaagccagacaggg

<210> SEQ ID NO 330
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 330
aacacccggtctctgagagaaactgttatagcagaaaagtaagtttaaat

<210> SEQ ID NO 331
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
caacgccagt ctctgagaag acgttatga gctatgctgt ttgaatggt ccca

<210> SEQ ID NO 332
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<400> SEQUENCE: 332
caacgccagt ctctgatag caaatgcat agg

<210> SEQ ID NO 333
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<400> SEQUENCE: 333
aatgcatagg ttaccaccca caggtgccag ggg

<210> SEQ ID NO 334
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<400> SEQUENCE: 334
acacacatgg gaaagctctt gggccaggaag agg

<210> SEQ ID NO 335
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 335
ggaggggtas gtatacagaa acacagagaa gtagaat

<210> SEQ ID NO 336
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 336
agagaatgga gaggctcagc aaactcagca ctagaaa

<210> SEQ ID NO 337
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 337
ggacgaaca ccggaacct tcaaaacgc atagcagtt aaataaaggg tagttcgtta

toaacctgaa aaaaaggcaac gaggcttgac cttttttttttt

<210> SEQ ID NO 338
<211> LENGTH: 166
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide
<400> SEQUENCE: 338
ggacgaaaca ccggtagtat taagtattgt ttatggtcg ataaatattc tttgaatttc
cccttgattat tgttgataa aaagttaaaaa taatctttgtt ggacaccttc aaacccgcat
agcaagttaa aataaggata gtccggtttc acattgaaaa aagtggaacc gcaggggtgtc

<210> SEQ ID NO 339
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 339

gggtttttaga gctatgctgt tttgaatgtt cccaaacgg gtcctcgaga agacgtttta

<210> SEQ ID NO 340
<211> LENGTH: 95
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 340

aggctatgct gttttgaaat gcctcaasac ttttt

<210> SEQ ID NO 341
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 341

aaacnnnnnn nnnnnnnnnnn nnnnnnn nnnntg

<210> SEQ ID NO 342
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 342

gtggaaagga caaaccaccc gcggttcggag aagacggttt tttagagctag aataagcaag

<210> SEQ ID NO 343
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

<400> SEQUENCE: 343

tttaaatag gcctagtcggt tttt
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6) .. (24)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 343

acccgnnnn nnnnnnnnnn nnn

24

<210> SEQ ID NO 344
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5) .. (23)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 344

aacccnnnn nnnnnnnnnnn nnc

24

<210> SEQ ID NO 345
<211> LENGTH: 88
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 345

gttttagac tgtgtgttt tgtatgttcc caaaaactag accaaaagtc cgggttttga 60
gatgtgtgt tttgaatgtt cccaaac

88

<210> SEQ ID NO 346
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 346

aaacggsgag gcgtagttcc gagcagaaga agaag

35

<210> SEQ ID NO 347
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 347

aaaacctttt ccctgcgtg ccgagcagac cttcc

35

<210> SEQ ID NO 348
<211> LENGTH: 46
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (19)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 348

nnnnnnnnnnnnnnnnnnuuaauuguauccucaagauuuuuuuuu46
<210> SEQ ID NO 349
<211> LENGTH: 91
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 349

guaauuuaua ucuuacgaa gcauaaaga uaaacgguuca uggcggaauc aacacccugu60

cauuuauag cagggguuu ucuauuuua a91
<210> SEQ ID NO 350
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 350

ttttctagtct ctaggttctc tgactcctc ttcattctac ttctctgtgt ttctgtatac60
taacctcctc70
<210> SEQ ID NO 351
<211> LENGTH: 122
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 351

ggacggaggg ctagctog cagcagaa gaagggctcc catcaca caaaccgtggcg60
cattgccac aagcaggcca atcgatcgtc aaccttcaatg actagggtg120
gc122
<210> SEQ ID NO 352
<211> LENGTH: 48
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<222> LOCATION: (3) .. (32)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 352

a<mismatch>nnnnnnnnnnnnnnnnnnnnnnnnnnnnnuguuuauagcuaaguuu48
<210> SEQ ID NO 353
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<223> OTHER INFORMATION: Description of Combined DNA/RNA Molecule: Synthetic oligonucleotide

<400> SEQUENCE: 353
agcauagcag guuaaaaua ggctagucgc uuuaaucauuu gasaaagugc cacggcaguc 60

uguacuu 67

<210> SEQ ID NO 354
<211> LENGTH: 62
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<221> NAME/KEY: modified_base
<222> LOCATION: (1)...(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 354

nnnnnnnn nnnnnnnnn guuuuagagc uagaaauagc aaguuuaauu aagggcaguc 60
gq 62

cgg

<210> SEQ ID NO 355
<211> LENGTH: 73
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 355
tgatgttcc caaaacggaa gggcctgtag cggacgaa gaagaagtat tagagctatg 60
cgttttgcg aag 73

cctgcttcc acacctctgc cctgaacacc caatctctgc cctctctgcc acacctctgc 60
attctctgt 69

<210> SEQ ID NO 356
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 356
cctgcttcc acacctctgc cctgaacacc caatctctgc cctctctgcc acacctctgc 60
attctctgt 69

<210> SEQ ID NO 357
<211> LENGTH: 138
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 357
acccaagcag tgaagcctag tagctaaatg tgtagggct accccacagg tgccaggggc 60
cctccccaa gttccccacg cccccctccaa cccccctcg ccacaggct ttccccatgtg 120
tgcgctgg accccctg 138

cctgcttcc acacctctgc 60

<210> SEQ ID NO 358
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 358
gtgcctgtgc gaagcctgcc 20

<210> SEQ ID NO 359
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 359

tctggagcgc agtgcagtagt

<210> SEQ ID NO 360
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 360

accttcgctg ttctcaccat tc

<210> SEQ ID NO 361
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 361

ttgggagtg cacagacctc

<210> SEQ ID NO 362
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 362

ggctcccttg gttcaagta

<210> SEQ ID NO 363
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 363

agagaggtct ggtggtgcga a

<210> SEQ ID NO 364
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic probe

<400> SEQUENCE: 364

tagctcaas aacctctttt ctagctgac

<210> SEQ ID NO 365
<211> LENGTH: 30
<212> TYPE: DNA
<211> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic probe

<400> SEQUENCE: 365
cagcctttatatctttgc ttagcgtt
  30  

<210> SEQ ID NO 366
<211> LENGTH: 99
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (99)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 366
nunnununn nnnnnnnnnn uuuuuaguc uugaaaaugc aaguuauuuu aagccacug
  60
cguauaacc uugaaaaagc ggcaacggagc cggugcuu
  99  

<210> SEQ ID NO 367
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 367
tagcggggt.a gc
  12  

<210> SEQ ID NO 368
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 368
togtgacat gt
  12  

<210> SEQ ID NO 369
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 369
actcccegta gg
  12  

<210> SEQ ID NO 370
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 370
actgcgtgtt aa
  12  

<210> SEQ ID NO 371
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 371
aagtgcgcctg at
  12
<210> SEQ ID NO 372
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 372

tagtcgacc ag

<210> SEQ ID NO 373
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 373

ggcgtaatg at

<210> SEQ ID NO 374
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 374

tgtcgcatgt ta

<210> SEQ ID NO 375
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 375

atggaaagcg at

<210> SEQ ID NO 376
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 376

gccgaattcc tc

<210> SEQ ID NO 377
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 377

gcatggtagc ga

<210> SEQ ID NO 378
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 378

cggtactttt ac

<210> SEQ ID NO 379
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 379

gcctggtgco gta
<210> SEQ ID NO 380
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 380
tacgtaagt cg 12

<210> SEQ ID NO 381
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 381
cacgaaatta cc 12

<210> SEQ ID NO 382
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 382
aaccaagata cg 12

<210> SEQ ID NO 383
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 383
sgatgtcata gc 12

<210> SEQ ID NO 384
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 384
gtcctcagat cg 12

<210> SEQ ID NO 385
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 385
tcgtcggtgc ca 12

<210> SEQ ID NO 386
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 386
actcgtagt ga 12

<210> SEQ ID NO 387
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 387
caggagtcactg 12
<210> SEQ ID NO 388
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 388

tgattatccct ac 12
<210> SEQ ID NO 389
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 389

tttcaagggc gg 12
<210> SEQ ID NO 390
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 390

cgcctgtgga at 12
<210> SEQ ID NO 391
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 391

acacgcgtcc ta 12
<210> SEQ ID NO 392
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 392

gattcatcag cg 12
<210> SEQ ID NO 393
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 393

acacgcgtct tc 12
<210> SEQ ID NO 394
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 394

atcgtgcccct aas 12
<210> SEQ ID NO 395
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 395
`<210> SEQ ID NO 396
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 396
gcgtagtatct cg
<210> SEQ ID NO 397
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 397
cgattctcct cg
<210> SEQ ID NO 398
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 398
tgcgctccca gt
<210> SEQ ID NO 399
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 399
atasgtaggc gc
<210> SEQ ID NO 400
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 400
aaggtcgyccc at
<210> SEQ ID NO 401
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 401
gtccgggaact at
<210> SEQ ID NO 402
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 402
ttgcgagcat tt
<210> SEQ ID NO 403
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

...
<400> SEQUENCE: 403

tgagtcgctg ag
12

<210> SEQ ID NO: 404
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 404

tttacgcaga gg
12

<210> SEQ ID NO: 405
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 405

aggsagttatc gc
12

<210> SEQ ID NO: 406
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 406

actcgatacc at
12

<210> SEQ ID NO: 407
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 407

cgctacatcg ca
12

<210> SEQ ID NO: 408
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 408

ttcgtaaccgc gc
12

<210> SEQ ID NO: 409
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 409

ccaaacgggtt aa
12

<210> SEQ ID NO: 410
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 410

cgattccttgc
12

<210> SEQ ID NO: 411
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 411
cgtcatgast aa 12

<210> SEQ ID NO 412
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 412
agtgccgatg ac 12

<210> SEQ ID NO 413
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 413
ccctacggc ac 12

<210> SEQ ID NO 414
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 414
gccacccgc ac 12

<210> SEQ ID NO 415
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 415
tggacaccg gt 12

<210> SEQ ID NO 416
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 416	ttgactcgag cg 12

<210> SEQ ID NO 417
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 417
actatgcgyta gg 12

<210> SEQ ID NO 418
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 418
taccccaag cg 12

<210> SEQ ID NO 419
<211> LENGTH: 12
<212> TYPE: DNA
<400> SEQUENCE: 419
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 419

gcaggacgtc cg

12

<210> SEQ ID NO 420
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 420

acaccgaaaa cg

12

<210> SEQ ID NO 421
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 421

caggtgattg ag

12

<210> SEQ ID NO 422
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 422

cacgaggtat gc

12

<210> SEQ ID NO 423
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 423
	taaagcgacc cg

12

<210> SEQ ID NO 424
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 424

cctagtccgc ca

12

<210> SEQ ID NO 425
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 425

cgaaaaacgtg gc

12

<210> SEQ ID NO 426
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 426

cgtgcctgac ac

12

<210> SEQ ID NO 427
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 427

tttacctag aa

<210> SEQ ID NO 428
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 428
cgtagcctag tt

<210> SEQ ID NO 429
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 429
cccaacggt ta

<210> SEQ ID NO 430
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 430
gcgttatcag aa

<210> SEQ ID NO 431
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 431
tcgtgtgtaa ac

<210> SEQ ID NO 432
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 432
cgactttttg ca

<210> SEQ ID NO 433
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 433
tcgacgactc ac

<210> SEQ ID NO 434
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 434
acgctcgaga ta

<210> SEQ ID NO 435
cgtacgcaac  

12

catatgctgga  

12

cggttcatagat  

12

gsctcgtgta  

12

ttcgcacagg  

12

gtcgtagtac  

12

gtgacgcttc  

12

gttaacagcg
tagctaaccg tt 12

agtaaagcg ct 12

ggtattaatcg tg 12

guacaccuca augcuauaggg guuuuagagc uagaaauagc aaguuaaaau aagggcaguc 60
cguuuuuuu 69

gacaucuauag cuuucccua guuuuagagc uagaaauagc aaguuaaaau aagggcaguc 60
cguuuuuuu 69

gaguccgagc agaagaagaa guuuuagagc uagaaauagc aaguuaaaau aagggcaguc 60
cguuuuuuu 69

<210> SEQ ID NO 448
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artifical Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 448

cguuuuuuu 69
<220> FEATURE:
<221> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 449

ggggccqsga uuggguguuc guuuuagsgc uagaaauagc aaguuuaaau aagggcuaugc 60
cguuuuuuu 69

<210> SEQ ID NO 450
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 450

guggcgqsgag ggggqcgagaau guuuuagsgc uagaaauagc aaguuuaaau aagggcuaugc 60
cguuuuuuu 69

<210> SEQ ID NO 451
<211> LENGTH: 76
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 451

gucaccuca aagacuaqgg guuuuagsgc uagaaauagc aaguuuaaau aagggcuaugc 60
cguuuuuuuuuuu 76

<210> SEQ ID NO 452
<211> LENGTH: 76
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 452

gacaucqga uccucoccau guuuuagsgc uagaaauagc aaguuuaaau aagggcuaugc 60
cguuuuuuuuuuu 76

<210> SEQ ID NO 453
<211> LENGTH: 76
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 453

gaguccqsga aqagaqga guuuuagsgc uagaaauagc aaguuuaaau aagggcuaugc 60
cguuuuuuuuuuu 76

<210> SEQ ID NO 454
<211> LENGTH: 76
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 454

gggccc gag uuggguguc guuuaagac uagaaauagc aaguuuaaau aagcguaguc  60
cguuacau uuuuuuu  76

<210> SEQ ID NO 455
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 455

ugagccgagag uggcagau guuuaagac uagaaauagc aaguuuaaau aagcguaguc  60
cguuacau uuuuuuu  76

<210> SEQ ID NO 456
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 456

uguaccauca augucuaggg guuuaagac uagaaauagc aaguuuaaau aagcguaguc  60
cguuacac cuugaaaggu guuuuuuu  88

<210> SEQ ID NO 457
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 457

uguaccauca uccucoccoau guuuaagac uagaaauagc aaguuuaaau aagcguaguc  60
cguuacac cuugaaaggu guuuuuuu  88

<210> SEQ ID NO 458
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 458

gaguccgagc agaaagaagaa guuuaagac uagaaauagc aaguuuaaau aagcguaguc  60
cguuacac cuugaaaggu guuuuuuu  88

<210> SEQ ID NO 459
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 459

ggggccgaga uuggguguc guuuaagac uagaaauagc aaguuuaaau aagcguaguc  60
<210> SEQ ID NO 460
<211> LENGTH: 88
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 460

cguuauacac uugaaagau guuuuuuu

<210> SEQ ID NO 461
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 461

gucaccuuca uagcuaggg guuuuaagac uagaaauagc aaaguuuaau aaggcuaguc 60
cguuauaac uugaaagau guuuuuuuuu

<210> SEQ ID NO 462
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 462

gcaaucuag uccucccaau guuuuaagac uagaaauagc aaaguuuaau aaggcuaguc 60
cguuauaac uugaaagau gugcaccagu cggugccuuu uu

<210> SEQ ID NO 463
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 463

gagucccgag cagaaagas guuuuaagac uagaaauagc aaaguuuaau aaggcuaguc 60
cguuauaac uugaaagau gugcaccagu cggugccuuu uu

<210> SEQ ID NO 464
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 464

ggggccgaga uuggguguuc guuuuaagac uagaaauagc aaaguuuaau aaggcuaguc 60
cguuauaac uugaaagau gugcaccagu cggugccuuu uu

<210> SEQ ID NO 465
<211> LENGTH: 103
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 465

guggcgaag ggccgagau guuuuagagc uagaaauagc aaguuuaauu aaggcuaagc 60
cguuauaac uugaaaaagu ggcacacagc cgggcuuuu uuu 103

<210> SEQ ID NO 466
<211> LENGTH: 120
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide

<400> SEQUENCE: 466

gtggaagaag acaaaccog ggtcttcagag aagacttgttt ttagagctag aaatagcag 60
ttttaaaaa gctagttcgt tataacctgtt aaaagttgac acgagtctgg ttgctttt 120

<210> SEQ ID NO 467
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 467

tgtgtgcgct ggttgtatttc ttttgctgt tttttgctt 40

<210> SEQ ID NO 468
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of ArtificialSequence: Synthetic oligonucleotide

<400> SEQUENCE: 468

gauuucucu uggccuunuu guuuuu 26

<210> SEQ ID NO 469
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified base
<222> LOCATION: (26) (26)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 469

tgaatcttct tgtgcttttt ttttttttt 26

<210> SEQ ID NO 470
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
tgatttctc ttgcgttttt tggc
<210> SEQ ID NO 471
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 470

tnatttctc ttgcgttttt tggc
<210> SEQ ID NO 472
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 472
gatttcttct ttgcgttttt tgg
<210> SEQ ID NO 473
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<223> OTHER INFORMATION: CDS
<222> LOCATION: (1)..(33)
<400> SEQUENCE: 473
toc atc gct aca acc cac aac cct got agt gag c
Ser Ile Arg Thr Thr His Asn Pro Ala Ser Glu
1 5 10
<210> SEQ ID NO 474
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<400> SEQUENCE: 474
Ser Ile Arg Thr Thr His Asn Pro Ala Ser Glu
1 5 10
<210> SEQ ID NO 475
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (33)

<400> SEQUENCE: 475

tcc atc gca acc acc cac aac cct gct agt gag c
1 5 10

<210> SEQ ID NO 476
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 476

Ser Ile Ala Thr Thr His Asn Pro Ala Ser Glu
1 5 10

<210> SEQ ID NO 477
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 477

tca atc gct aca acc cac aac cct gct agt gag c
1 5 10

<210> SEQ ID NO 478
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (36)

<400> SEQUENCE: 478

caa ttg aat tta aaa gaa acc gat acc gtt tgt gtt taagga
1 5 10

<210> SEQ ID NO 479
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 479

Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Leu Val
1 5 10

<210> SEQ ID NO 480
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (42)

<400> SEQUENCE: 480

caa ttg aat tta aaa gaa acc gat acc gtt tgt gtt taagga
1 5 10
cct ttg sat tta aca gag acc gat acc gtt tsc gas att gga
Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> SEQ ID NO 481
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 481

Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> SEQ ID NO 482
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (2) .. (34)

<400> SEQUENCE: 482

t cct aca acc gac tcc gcg ctc cgt aca gta
Pro Lys Leu Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> SEQ ID NO 483
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 483

Pro Lys Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> SEQ ID NO 484
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (2) .. (34)

<400> SEQUENCE: 484

t cct aca acc gac tcc gcg ctc cgt aca gta
Pro Thr Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> SEQ ID NO 485
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 485

Pro Thr Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> SEQ ID NO 486
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 486

tgccctggtt gatctctctt tgcgcctttt tgg

<210> SEQ ID NO 487
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 487

tacgtgtttt gattttccct tgccgcttttt ttg

<210> SEQ ID NO 489
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 488

ggaggttttt atgggaag gccattg

<210> SEQ ID NO 489
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 489

gtaaaaaa gaagactagaa ttttgatac

<210> SEQ ID NO 490
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 490

ggaggttttt atgggaag gccaaagaag aaggtgatta ttgatac

<210> SEQ ID NO 491
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 491

aggtgaagca ttatgtctca aaaaata

<210> SEQ ID NO 492
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 492

attttattaa tacaatcaag tgaaatcat

<210> SEQ ID NO 493
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 493

aggtgaagca ttatgtctca aatataac aatacgtga aatcat

<210> SEQ ID NO 494
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (36)
<400> SEQUENCE: 494
aaat tta aaaa gaa acc gat acc gtt tac gaa att gga 36
Ann Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> SEQ ID NO 495
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 495
Ann Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> SEQ ID NO 496
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (30)
<400> SEQUENCE: 496
aaat tta aaaa gaa acc gat acc gtt ttg gtt taagga 36
Ann Leu Lys Glu Thr Asp Thr Val Leu Val
1 5 10

<210> SEQ ID NO 497
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 497
Ann Leu Lys Glu Thr Asp Thr Val Leu Val
1 5 10

<210> SEQ ID NO 498
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (36)
<400> SEQUENCE: 498
tggt gat cca aaaa tat gtt gtt ttt gat agt cca 36
Trp Asp Pro Lys Tyr Gly Gly Phe Asp Ser Pro
1 5 10

<210> SEQ ID NO 499
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 499
Trp Asp Pro Lys Tyr Gly Gly Phe Asp Ser Pro
1 5 10

<210> SEQ ID NO 500
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (36)
<400> SEQUENCE: 500

```
tgg gat cca aaa aaa tat tgt ggt ttt gat agt cca
Trp Asp Pro Lys Lys Tyr Cys Gly Phe Asp Ser Pro
  1  5  10
```

<210> SEQ ID NO 501
<211> LENGTH: 12
<212> TYPE: PPT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 501

```
Trp Asp Pro Lys Lys Tyr Cys Gly Phe Asp Ser Pro
  1  5  10
```

<210> SEQ ID NO 502
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(4104)

<400> SEQUENCE: 502

```
aaactatcattt aacgcagcogc gatgtaggggt ttttg
```

<210> SEQ ID NO 503
<211> LENGTH: 4104
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(4104)

<400> SEQUENCE: 503

```
atrg acc aag aag tac agc att gcc ctc gcc gac acc acc aac tct gtg
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asp Ser Val
  1  5  10  15
```

```
ggc tgg gcc gtc gac gcc gcc ggc gag tac aag gtg gcc acc aac aac ttc
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
  20 25 30
```

```
aag gtc ctc gcc aac acc gcc ggc cag cac gac aca aag aac ctc gtc
Lys Val Leu Gly Aam Thr Asp Arg His Ser Ile Lys Arg Leu
  35 40
```

```
gpa gcc ctc ctc gcc gcc ggc gaa acc gac gag gcc acc cag gcc
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
  50 55 60
```

```
aag agg acc gcg gaa aga aca acc aca cag cgc asp tgc
Lys Arg Thr Ala Arg Arg Tyr Thr Arg Lys Aam Arg Ile Cys
  65 70 75 80
```

```
ta ccc cca gag ctc gcc gac aag gtc ggc gag gag gag cag cag
Tyr Leu Gln Glu Ile Phe Ser Arg Glu Met Ala Lys Val Asp Ser
  85 90 95
```

```
ngcc ctc ccc acc gcc gcc gcc ggc gag ctc gcc gtc gac gag gag cag aag
Phe Phe His Arg Leu Gly Ser Phe Leu Val Glu Glu Asp Lys Lys
  100 105 110
```

```
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcd
His Glu Arg His Pro Ile Phe Gly Aam Ile Val Arg Val Ala Tyr
  115 120 125
```

```
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
cac gag cgc cac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gc
<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>145</td>
<td>agc acc gac aag gcc gac ctg cgg ctg atc tat ctg gcc ctg gcc cac Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His</td>
</tr>
<tr>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>155</td>
<td>160</td>
</tr>
<tr>
<td>160</td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>atg atc aag ttc cgg ggc cac ttc atc gag ggc gac ctg aac ccc Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Aen Pro</td>
</tr>
<tr>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>175</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>gac aac agc gac gtc gac aag ctg ttc atc cag ctg gtc cag acc tac Amp Asn Ser Arg Val Asp Leu Phe Ile Glu Val Glu Thr Tyr</td>
</tr>
<tr>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>190</td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>aac cag ctg ttc gag gaa aac ccc atc aac gcc agc ggc gtc gac gcc Asn Glu Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala</td>
</tr>
<tr>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>205</td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>aag gcc atc ctg tct gcc aga ctg aag agc aag aag cgg ctg gaa aat Lys Ala Ile Leu Ser Ser Leu Arg Leu Ser Arg Arg Leu Glu Asn</td>
</tr>
<tr>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td>220</td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>ctc atc ggc cag ctg ccc ggc gag aag aag aat gag ctt gcc aac Leu Ile Ala Gln Leu Pro Gly Lys Asn Gly Leu Phe Gly Aac</td>
</tr>
<tr>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>235</td>
<td>240</td>
</tr>
<tr>
<td>240</td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>ctc att ggc ctg aag ctc ggc ctg acc ccc aac ttc aag agc aac ctc Leu Ile Ala Leu Ser Gly Thr Asp Phe Leu Ser Aen Phe</td>
</tr>
<tr>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>255</td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>gac ctg gcc gag gat gcc aaa cag ctg aag aag acc acc tac gac Amp Leu Ala Gln Asp Ala Lys Leu Gln Ser Leu Asp Tyr Thr Tyr Asp</td>
</tr>
<tr>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>270</td>
<td></td>
</tr>
<tr>
<td>275</td>
<td>gac gac ctg aac ctg ctg gcc cag atc ggc gac cag tac gcc gac Aep Asp Leu Asp Aam Leu Ala Gln Ile Gly Gln Tyr Ala Aep</td>
</tr>
<tr>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>285</td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>ctc ttc ctg gcc aag aac ctg tcc gac gcc atc ctg ctc gtc aac Leu Phe Leu Ala Ala Lys Leu Ser Asp Ala Ile Leu Leu Ser Asp</td>
</tr>
<tr>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>300</td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>aat ctc aga gta agc gac atc acc aag ggc ccc ctg aag gcc ctc Ile Leu Arg Val Aem Thr Glu Ile Thr Lye Ser Pro Leu Ser Aam Ser</td>
</tr>
<tr>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>315</td>
<td>320</td>
</tr>
<tr>
<td>320</td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>gct ctc gtc gga cag cag ctc cct gag aag tac aag gag att ttc ttc Ala Leu Val Arg Gln Glu Leu Pro Glu Lys Tyr Lye Ile Phe Phe</td>
</tr>
<tr>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>335</td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>gac cag aca aag aac gcc tac gcc gac tac att gcc gga gcc gac Asp Glu Ser Lys Aam Gly Tyr Ala Gly Tyr Ile Asp Gln Gly Ala Ser</td>
</tr>
<tr>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>350</td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>cag gaa gag tgc tac atg gaa cag cct gta aag ccc ace atc ctg gaa aag atg gag Gln Glu Glu Phe Tyr Lys Phe Ile Pro Ile Leu Glu Lys Met Asp</td>
</tr>
<tr>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>365</td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>ggg acc gag gaa ctg ccc aag aag gag gag gac ctg cgg Gly Thr Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg</td>
</tr>
<tr>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>380</td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>aag cag cgg acc ttc gcc aac gcc aac cct gcc aac ccc ctc gag悦 ctc gtc aac cc aac ctc ctc gcc Lys Glu Arg Thr Phe Aen Asn Gly Ser Ile Pro His Glu Ile His Leu</td>
</tr>
<tr>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>395</td>
<td>400</td>
</tr>
<tr>
<td>400</td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>gga gag ctg cag ccc att ctg cgg cgg cag gaa gat ttc cca ttc Gly Glu Leu His Ala Ile Leu Arg Leu Glu Gln Asp Phe Tyr Pro Phe</td>
</tr>
<tr>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>415</td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>ctt aag gac aag cgg gaa aag atc gag aag atc cgc ttc ggc gtc aac ctg Leu Lys Aep Aen Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile</td>
</tr>
<tr>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>430</td>
<td></td>
</tr>
<tr>
<td>435</td>
<td>ccc tac tac gtc ggc cct ctc ggc aag gga aac agc aag ttc gcc tgg Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp</td>
</tr>
<tr>
<td>440</td>
<td>445</td>
</tr>
<tr>
<td>445</td>
<td></td>
</tr>
<tr>
<td>450</td>
<td></td>
</tr>
<tr>
<td>455</td>
<td></td>
</tr>
</tbody>
</table>
Thr Thr Gln Lys Gly Gln Lys Aen Ser Arg Glu Arg Met Lys Arg Ile 770 775 780

gaa gag ggc atc aag gag ctc ggc gac atc ctc aag gaa cac ccc 2400
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gin Ile Leu Lys Glu His Pro 785 790 795 800

gtg gaa acc acc cag ctc gac aag gac cag tac ctc tac tac ctc 2448
Val Glu Asn Thr Gln Leu Gln Asp Lys Val Lys Tyr Tyr Tyr Leu 805 810 815

cag aat ggg ggc gat att tac gtc gac cag gaa ctg gac atc aac cgg 2496
Gln Asn Gly Arg Asp Met Tyr Val Asp Gin Glu Asp Ile Aen Arg 820 825 830 835 840 845

cag ctc tac gat gtt gac ggc atc gtc cct cag aac ttt ctc aag 2544
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gin Ser Phe Leu Lys 845 850 855 860

gac gac ctc atc aag ggc aag atc gtc acc aga aac aag gg cc gg 2592
Asp Asp Ser Ile Gln Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg 860 865 870 875 880

ggc aag aqc gac aac gtt ccc tca gaa gag gtc gtc aag aag atg aag 2640
Gly Lys Ser Asp Aen Val Pro Ser Glu Val Val Lys Pro Met Lys 880 885 890 895 900 905

aac tac tgt cgg cag ctc ctc aac ggc aag cag att acc cag aga aag 2688
Aen Tyr Trp Gln Ala His Val Leu Aen Ala Ile Thr Gln Arg Lys 905 910 915 920 925

ttc gac gat ctc acc aag ggc gag aag ggc ggc ctt aag gaa ctc gat 2736
Phe Asp Asp Thr Thr Ala Glu Lys Glu Gly Leu Ser Gin Leu Asp 925 930 935 940

aag gcc ggc ttc atc aag aga cag ctc gta gaa acc cgg cag atc aca 2784
Lys Ala Gly Phe Ile Lys Arg Glu Leu Val Glu Thr Arg Gin Ile Thr 940 945 950 955 960

aag cac gtc gca cag atc ctc gac ccc cgg atg aac aat aag tac gac 2832
Lys His Val Ala Glu Ile Leu Asp Ser Arg Met Aen Thr Lys Tyr Asp 960 965 970 975 980 985 990

gag aat gac aag ctc ggg gaa gct aag atg aat acc acc ctc aag tcc 2880
Glu Asp Asp Arg Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser 990 995 1000 1005 1010

aag ctt gtt ccc gac gaa gat cag ttt cag aag gaa gat ttc cgg gaa 2928
Lys Leu Val Ser Asp Phe Arg Lys Phe Gin Phe Tyr Lys Arg Val 1010 1015 1020 1025 1030 1035 1040 1045 1050

gac gtc acc tac ctc ccc gcc ctc gcc ctc ctc ctc gac gtc gaa gat ttc 2976
Glu Ile Aen Asn Tyr His Ala Ala Tyr Leu Aen Ala Val 1050 1055 1060 1065 1070 1075 1080

gtt gga acc ggc ctc tct aaaaa aag tac cct aag gaa aag cgg gtc 3024
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe 990 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080
<table>
<thead>
<tr>
<th>cgg aag</th>
<th>gtt ctc</th>
<th>agc atg</th>
<th>ccc</th>
<th>cag gtc</th>
<th>aat ctc</th>
<th>gta aag</th>
<th>acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg Lys</td>
<td>Val Leu</td>
<td>Ser Met</td>
<td>Pro</td>
<td>Gin Val</td>
<td>Arg Ile</td>
<td>Val Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>1085</td>
<td>1090</td>
<td>1095</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gag gtc</td>
<td>cag aca</td>
<td>ggc ggc</td>
<td>ttc</td>
<td>aag aag</td>
<td>gag tct</td>
<td>atc gtc</td>
<td>ccc</td>
</tr>
<tr>
<td>Arg Glu</td>
<td>Val Thr</td>
<td>Gly Gly</td>
<td>Phe</td>
<td>Ser Lys</td>
<td>Glu Ser</td>
<td>Ile Leu</td>
<td>Pro</td>
</tr>
<tr>
<td>1100</td>
<td>1105</td>
<td>1110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aag aac</td>
<td>agc gat</td>
<td>aag ctc</td>
<td>gcc</td>
<td>aca aag</td>
<td>aag gag</td>
<td>aag gac</td>
<td>tgg</td>
</tr>
<tr>
<td>Arg Ser</td>
<td>Ser Asp</td>
<td>Lys Leu</td>
<td>Ile</td>
<td>Ala Arg</td>
<td>Lys Asp</td>
<td>Trp Asp</td>
<td>Pro</td>
</tr>
<tr>
<td>1115</td>
<td>1120</td>
<td>1125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aag cag</td>
<td>tac gcc</td>
<td>ggc ttc</td>
<td>gcc</td>
<td>tat</td>
<td>gce</td>
<td>ccc ttc</td>
<td>gcc</td>
</tr>
<tr>
<td>Lys Tyr</td>
<td>Gly Gly</td>
<td>Phe Ser</td>
<td>Pro Thr Val Ala Tyr Ser Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1130</td>
<td>1135</td>
<td>1140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ctc gtc</td>
<td>gtc gcc</td>
<td>aa aag</td>
<td>gta gaa</td>
<td>aag ggc</td>
<td>aag tct</td>
<td>aag aag</td>
<td>cag</td>
</tr>
<tr>
<td>Leu Val</td>
<td>Val Ala</td>
<td>Lys Val</td>
<td>Glu</td>
<td>Lys Gly</td>
<td>Lys Ser</td>
<td>Lys Leu</td>
<td>Lys</td>
</tr>
<tr>
<td>1145</td>
<td>1150</td>
<td>1155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aag gcc</td>
<td>ctc gcc</td>
<td>ggc ggc</td>
<td>atc</td>
<td>aag</td>
<td>aag aag</td>
<td>aag gac</td>
<td>aag</td>
</tr>
<tr>
<td>Ser Val</td>
<td>Gly Leu</td>
<td>Leu Gly</td>
<td>Ile</td>
<td>Thr Ile</td>
<td>Met Gly</td>
<td>Arg Ser</td>
<td>Ser</td>
</tr>
<tr>
<td>1160</td>
<td>1165</td>
<td>1170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttc ggc</td>
<td>gac aat</td>
<td>ccc atc</td>
<td>gcc</td>
<td>ttc</td>
<td>ctt gca</td>
<td>gca gca</td>
<td>gcc</td>
</tr>
<tr>
<td>Phe Glu</td>
<td>Lys Asp</td>
<td>Val Leu</td>
<td>Gln Ile</td>
<td>Thr Gly</td>
<td>Lys Ser</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>1175</td>
<td>1180</td>
<td>1185</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gac gtc</td>
<td>gac aag</td>
<td>gag ctg</td>
<td>atc</td>
<td>aag</td>
<td>ctc ccc</td>
<td>cag tct</td>
<td>ctg</td>
</tr>
<tr>
<td>Glu Val</td>
<td>Lys Tyr</td>
<td>Leu Ser</td>
<td>Arg Leu</td>
<td>Ile</td>
<td>Gly Leu</td>
<td>Pro Lys</td>
<td>Ser</td>
</tr>
<tr>
<td>1190</td>
<td>1195</td>
<td>1200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttc ggc</td>
<td>ggt gaa</td>
<td>gag gcc</td>
<td>cag</td>
<td>aag</td>
<td>gca tgc</td>
<td>gcc tcc</td>
<td>gcc</td>
</tr>
<tr>
<td>Phe Glu</td>
<td>Leu Glu</td>
<td>Arg Gly</td>
<td>Arg Glu</td>
<td>Met</td>
<td>Leu Ala</td>
<td>Ser Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>1205</td>
<td>1210</td>
<td>1215</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gag gcc</td>
<td>ctg aag</td>
<td>ggg gag</td>
<td>ggc</td>
<td>tct</td>
<td>gcc ggc</td>
<td>gcc tcc</td>
<td>gcc</td>
</tr>
<tr>
<td>Glu Leu</td>
<td>Glu Leu</td>
<td>Glu Asp</td>
<td>Arg</td>
<td>Lys</td>
<td>Lys Val</td>
<td>Lys Ser</td>
<td>Val</td>
</tr>
<tr>
<td>1220</td>
<td>1225</td>
<td>1230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aac gcc</td>
<td>ggc cag</td>
<td>aag tgc</td>
<td>gtc</td>
<td>cac</td>
<td>tct cgc</td>
<td>aag ctc</td>
<td>tgt</td>
</tr>
<tr>
<td>Asn Phe</td>
<td>Tyr Leu</td>
<td>Ala Ser</td>
<td>His</td>
<td>Tyr Glu</td>
<td>Lys Ser</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>1235</td>
<td>1240</td>
<td>1245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ccc ggc</td>
<td>gat aat</td>
<td>gag tag</td>
<td>cag</td>
<td>aag</td>
<td>cag ctc</td>
<td>gta gaa</td>
<td>aag</td>
</tr>
<tr>
<td>Pro Glu</td>
<td>Asp Arg</td>
<td>Glu Gin</td>
<td>Lys</td>
<td>Glu</td>
<td>Phe Val</td>
<td>Phe</td>
<td>Lys</td>
</tr>
<tr>
<td>1250</td>
<td>1255</td>
<td>1260</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cac ggc</td>
<td>gtc gac</td>
<td>gag ctg</td>
<td>atc</td>
<td>aag</td>
<td>gtc aag</td>
<td>gcc gcc</td>
<td>gcc</td>
</tr>
<tr>
<td>His Tyr</td>
<td>Leu Asp</td>
<td>Glu Ile</td>
<td>Ile</td>
<td>Arg</td>
<td>Gin Ile</td>
<td>Ser</td>
<td>Glu</td>
</tr>
<tr>
<td>1265</td>
<td>1270</td>
<td>1275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aag atc</td>
<td>ctc atc</td>
<td>gcc gac</td>
<td>gcc</td>
<td>aat</td>
<td>gtc gcc</td>
<td>aag</td>
<td>tgc</td>
</tr>
<tr>
<td>Arg Val</td>
<td>Ile Leu</td>
<td>Asp Ala</td>
<td>Ala</td>
<td>Tyr</td>
<td>Leu Val</td>
<td>Asp</td>
<td>Ala</td>
</tr>
<tr>
<td>1280</td>
<td>1285</td>
<td>1290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tac aag</td>
<td>cag cag</td>
<td>ggt aag</td>
<td>ccc</td>
<td>atc</td>
<td>aag gag</td>
<td>cag ggc</td>
<td>aat</td>
</tr>
<tr>
<td>Tyr Aen</td>
<td>Lys His</td>
<td>Arg Asp</td>
<td>Lys</td>
<td>Ile</td>
<td>Gly Leu</td>
<td>Arg Glu</td>
<td>Ala</td>
</tr>
<tr>
<td>1295</td>
<td>1300</td>
<td>1305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aag ccg</td>
<td>ctc ttg</td>
<td>ggg gaa</td>
<td>ccc</td>
<td>atc</td>
<td>gtc gga</td>
<td>gcc gcc</td>
<td>gct</td>
</tr>
<tr>
<td>Ile Ile</td>
<td>His Leu</td>
<td>Phe Thr</td>
<td>Thr</td>
<td>Arg</td>
<td>Leu Gly</td>
<td>Ala Glu</td>
<td>Aen</td>
</tr>
<tr>
<td>1310</td>
<td>1315</td>
<td>1320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttc aag</td>
<td>tac ttt</td>
<td>gac ccc</td>
<td>atc</td>
<td>ggc</td>
<td>aag gag</td>
<td>cag cag</td>
<td>aag</td>
</tr>
<tr>
<td>Phe Lys</td>
<td>Tyr Phe</td>
<td>Asp Thr</td>
<td>Thr</td>
<td>Ser</td>
<td>Ser</td>
<td>His Arg</td>
<td>Lys</td>
</tr>
<tr>
<td>1325</td>
<td>1330</td>
<td>1335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aag ctc</td>
<td>ggc ggt</td>
<td>ctg gcc</td>
<td>gcc</td>
<td>aag</td>
<td>cag cag</td>
<td>cgc aag</td>
<td>aag</td>
</tr>
<tr>
<td>Thr Lys</td>
<td>Glu Val</td>
<td>Leu Asp</td>
<td>Ala</td>
<td>Thr</td>
<td>Leu Ile</td>
<td>His Glu</td>
<td>Ser</td>
</tr>
<tr>
<td>1340</td>
<td>1345</td>
<td>1350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ggc gcc</td>
<td>tac gga</td>
<td>aca cgg</td>
<td>aca</td>
<td>gtc</td>
<td>tgt cag</td>
<td>ctc ctg</td>
<td>gga</td>
</tr>
<tr>
<td>Gly Leu</td>
<td>Tyr Glu</td>
<td>Thr Arg</td>
<td>Ile</td>
<td>Asp</td>
<td>Ser Leu</td>
<td>Glu Leu</td>
<td>Gly</td>
</tr>
</tbody>
</table>
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 504

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1  5  10  15
Gly Trp Ala Val Ile Thr Asp Tyr Lys Val Pro Ser Lys Lys Phe
20  25  30
Lys Val Leu Gly Asn Thr Arg Arg His Ser Ile Lys Lys Asn Leu Ile
35  40  45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50  55  60
Lys Arg Thr Ala Arg Arg Tyr Thr Arg Tyr Lys Asn Arg Arg Ile Cys
65  70  75  80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85  90  95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr Thr Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Ala Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gin Leu Val Gin Thr Tyr
180 185 190
Asn Gin Leu Phe Glu Glu Asp Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gin Leu Pro Gly Glu Lys Asn Gin Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Ser Asn Phe
245 250 255
Aasp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Aasp Asp Leu Asp Asn Leu Leu Ala Gin Ile Gly Asp Gin Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Ser Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His Gin Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gin Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Aasp Gin Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405  410  415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420  425  430
Leu Lys Asp Arg Arg Glu Ile Glu Lys Ile Leu Thr Phe Arg Ile
435  440  445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450  455  460
Met Thr Arg Lys Ser Glu Gly Thr Ile Thr Pro Trp Asn Phe Glu Glu
465  470  475  480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485  490  495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500  505  510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515  520  525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530  535  540
Lys Lys Ala Ile Val Asp Leu Phe Lys Thr Asn Arg Lys Val Thr
545  550  555  560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565  570  575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580  585  590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Asp Phe Leu Asp
595  600  605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610  615  620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Tyr Thr Ala
625  630  635  640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645  650  655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660  665  670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675  680  685  690
Ala Asn Arg Asn Phe Met Gin Leu Ile His Asp Ser Leu Thr Phe
695  700
Lys Glu Asp Ile Gin Lys Ala Gin Val Ser Gly Gin Glu Asp Ser Leu
705  710  715  720
His Gin His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725  730  735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740  745  750
Arg His Lys Pro Glu Asn Ile Val Ile Ala Met Ala Arg Glu Asn Gin
755  760  765
Thr Thr Gin Lys Gly Gin Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770  775  780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gin Ile Leu Lys Glu His Pro
785  790  795  800
Val Glu Asn Thr Gin Leu Gin Asn Gin Lys Leu Tyr Leu Tyr Tyr Leu
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gln Asn Gly Arg Asp Met Tyr Val Asp Gin Glu Leu Asp Ile Asn Arg</td>
<td>820</td>
</tr>
<tr>
<td>Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gin Ser Phe Leu Lys</td>
<td>840</td>
</tr>
<tr>
<td>Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg</td>
<td>850</td>
</tr>
<tr>
<td>Gly Lys Ser Asp Asn Val Pro Ser Glu Val Val Lys Lys Met Lye</td>
<td>865</td>
</tr>
<tr>
<td>Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys</td>
<td>885</td>
</tr>
<tr>
<td>Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp</td>
<td>900</td>
</tr>
<tr>
<td>Lys Ala Gly Phe Ile Lys Arg Gin Leu Val Glu Thr Arg Gin Ile Thr</td>
<td>915</td>
</tr>
<tr>
<td>Lys His Val Ala Gin Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp</td>
<td>930</td>
</tr>
<tr>
<td>Glu Asn Asp Lys Leu Ile Arg Gin Val Val Ile Thr Leu Lys Ser</td>
<td>945</td>
</tr>
<tr>
<td>Lys Leu Val Ser Asp Phe Arg Lys Gin Lys Phe Tyr Lys Val Arg</td>
<td>960</td>
</tr>
<tr>
<td>Glu Ile Asn Asn Tyr His His Ala His Ala Tyr Leu Asn Ala Val</td>
<td>975</td>
</tr>
<tr>
<td>Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Phe</td>
<td>995</td>
</tr>
<tr>
<td>Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala</td>
<td>1010</td>
</tr>
<tr>
<td>Lys Ser Glu Gin Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe</td>
<td>1025</td>
</tr>
<tr>
<td>Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala</td>
<td>1040</td>
</tr>
<tr>
<td>Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu</td>
<td>1055</td>
</tr>
<tr>
<td>Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val</td>
<td>1070</td>
</tr>
<tr>
<td>Arg Lys Val Leu Ser Met Pro Gin Val Asn Ile Val Lys Lys Thr</td>
<td>1085</td>
</tr>
<tr>
<td>Glu Val Gin Thr Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys</td>
<td>1100</td>
</tr>
<tr>
<td>Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Thr Asp Pro</td>
<td>1115</td>
</tr>
<tr>
<td>Lys Lys Tyr Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val</td>
<td>1130</td>
</tr>
<tr>
<td>Leu Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys</td>
<td>1145</td>
</tr>
<tr>
<td>Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser</td>
<td>1160</td>
</tr>
<tr>
<td>Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys</td>
<td>1175</td>
</tr>
<tr>
<td>Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu</td>
<td>1190</td>
</tr>
<tr>
<td>Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly</td>
<td>1205</td>
</tr>
<tr>
<td>Glu Leu Gin Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val</td>
<td>1220</td>
</tr>
</tbody>
</table>
<210> SEQ ID NO: 505
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 505

cagagaagaagaggc

<210> SEQ ID NO: 506
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 506

cacaatgggaaacagctgatggtcattcctaatagactaggttgtggggcagc

<210> SEQ ID NO: 507
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 507
cctgagccacattc

<210> SEQ ID NO: 508
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 508

catacgatgtcacctcaaatgtaagcttctagcgttggtgacaaccacaaacct
ccttttacct cccatgcgctct cccgacgcaac gcaactctga

<SEQ ID NO 510
<LENGTH: 4269
<TYPE: DNA
<ORGANISM: Artificial Sequence
<FEATURE: Description of Artificial Sequence: Synthetic polynucleotide

<SEQUENCE: 510
atggactata aagcagcaaca gggagactac aagagatca tgatgatta caaagagcat

gagcataagaa tggccacaa gaagaacgg gaagttcggta tacacgaggt cccagcagcc

gacaagagat aagcagcagt cctggagactg gcaaccaact ctgatgggtgc gaagctgagc

ggtagtggtac aagcagagct cctggagactg gcaagcagcc
tacagctctg cccgacgcaac gcaactctga

407
408
-continued
gtgctctcagagttttctgaaagacgactcctctcgacactacagacgggtgctgacacagacagacagcgc
2700
aagaacgccgggcaagagcagcacaaggtcgtccctcgcgttaaagaaactgaagaacagcacagc
2760
tacgtggtgaaggtgtcgtggcggcgaacactgtgattaccacagagaaagttcgacactctgaccc
2820
aacggcagagggcgtgctgggaacacagctctgtcgtgccggcacggcaggacgtgctgctgaagct
2880
gtggacaacggagcagcaacggagcgacagtgcaaggtgaaacgttactacacatggtcaccagct
2940
atgcctgctgggttgctgagatgtcggctctcatccgaagcgtggtcactacagtgtctgcaaaagtac
3000
cctgtgttctgaatttcggaagattcccagtttacaagtgagccgacagatacaagctacaaagatc
3060
caccacaaccaagaacccgctaacttgccaggtttgaggaacccgcctggtctggacaagcactcc
3120
cctatatggaaagcagagtctgtgtacgagactacaaggttaaggagtggagagagagtgtgtcagaaagcagacggaagcgacagtctcagacgggactcctgctggagaaaaagacagtgttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgagtgtgtgtgagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg
tccttcaggg acaagggcag ctcacaagacc cgccgagcct tgaagttcga ggccgccacc 420
tggtgtaacc gataacgttt gactggcatc gatctccagag acaaacgca gactctggat 480
cacaagttgg agatcaacta caacacggc cacatacgta tacattgcca caagcagaga 540
acaaggcata caaagtagccg cacaacatcg ccgagagggc cagtcgtccag 600
gcggggacct acacagggaa catacccccag cagcagcctgt cagttggtgct geccagacaac 660
cactaccgct cagccacccgt gccccctgac aagaccccaa acgcgaagcgc ccgtccctag 720
gttgctgtag agttctgtag acgcgtgcggg atcaacttcc gcatggaagag gtgtacag 780

<210> SEQ ID NO 512
<211> LENGTH: 597
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 512

cgatattca caaacctgatt acaaaatttt gtaaagatgg actagtgatt tttaacttgtt 60
tgtctttttactgatgct gatagcctgc ttaaattgcct tgtatcatgt cttattgcttt 120
ccgtagatttt ctctctgta aataaacttgg ttgtgggtgct tttatgagaa 180
ogttgctgagc gcttgctgagc aagctggtgtgt ggttgctgac ggttgctgctgct aagctgaccc 240
caactgttgg ggcattggtc gcaactcccgt gacocctttcc gggactttgctg cttctccctgc 300
cacttttgcc cagccggcag ctcaggggagat cgtctggtgc cttgctggga caggggtctg 360
gctggtgggg acctgctgacttc cgtctggtgct gttgcggaaa ctcaggtgct ttcctttggtt 420
gcgtggctgt gttgcaactcccc gatggtctgc cgggagttcc tttgctgact ttcctgctgg 480
cctcactgca ggcagcctc ttcagcgcctgc cttgctggcg cttgctggcg cttgctggcg 540
tttcgcttc ggcctggaag atctctttgg gcccctccccc ccgatcgtg 597

<210> SEQ ID NO 513
<211> LENGTH: 210
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 513
cgacetegag actgcttcct actgctcag ccatactggtg ttcgctcctcc cccgtgctctg 60
ccttgactc gaaaggtggtc accatcctgg ttcctttctta attaatgac gaattgtcct 120
cgatcgtctc gatgtctcccc cccctcctct cgggggttgg ggctggccag cacgcctcag 180
gggagagtg agaaggacact accgagccctg 210

<210> SEQ ID NO 514
<211> LENGTH: 906
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (109), (195)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (139), (135)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 514

ataactttgt ataattgtatg ctatactgaag ttattcagca tgaataaaatg aaagctttga
60
gacgctgcag ctagcggtat atgcagcttc aaggttaatg aactcgacct accacatatt
120
gttaggagttt ttcaggtcctt ctatactttg atagccaagt tgcataaaga
180
atgaaactca ctaggtcact tctttcctag gtctgccagt ttttctcttg
240
aatgcactca ctaagacttca aataaagaga ttttttcttc tgcataaaga ttttctcttg
300
tcacaactca taattacttgc ttaaaggtctg catttacttg ctttttttcag
360
goctacacat attttagatg gttttacttg ctttttttca ccttccccaa cctctcttc
420
aacgccaaa ttaaaatgaa gcaatgtttg tttgtaacct gtttatctgca gtttataatg
480
gttacacaa aagcaataac atccaaactt ctcaaaattt aacactttttt tcaaactttt
540
ttagttgat tttgctccaa cttcctaatg tattttaata tcttgcagct tgcagcttca
600
gcgactaat aatagctgtc tgggtctttg gtcagctca ccttccccaa aatgttctttt tgcagcttca
660
aacccctttc ctaggctca aacactttttt tagagttttta ctgggtttgt aacgcctcttc
720
tgcattcttt gtcttctttc tgcagcgttttt tgggtctttt tgcagcgttttt
780
atatccatat ctcctaaatg tgctcttctt tgcagctttgt tgcagcttttt
840
gattcgcgctt gagtaataag cccgagtaa ttttttcttt tgcagcttttt
900
agtatat
960

<210> SEQ ID NO: 515
<211> LENGTH: 1079
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 515

cacgccctcc gacgctgttg gacgctgtct cacgacacgc cgggtacag tctctgacgt
60
gagggggaac agoctgtgtg gcggcaggaat gcagctcgcc ctgcagccac cggaggggga
120
ggsgaggggac agggcggatt gtcgcaccc gagggcggatt gcagcttcgg cggaggggga
180
gocgagggcag gtcggcagct gcggcagccc gcggcgcccg cgggggaggcc gcggcgcccag
240
tcttggagaga ctcttggagata tcttggagata tcttggagata tcttggagata tcttggagata
300
ggctgtgacc ggtgtactgtg ggtgtactgtg ggtgtactgtg ggtgtactgtg ggtgtactgtg
360
gtgccgagg gtcgacgacc tctgctgagt tctggtgctt tctgctgagt tctggtgctt
420
cccactgag ctgctgagt tctggtgctt tctgctgagt tctggtgctt tctgctgagt
480
tgttggagata tcttggagata tcttggagata tcttggagata tcttggagata tcttggagata
540
tggtcctctgc ctcttggagata tcttggagata tcttggagata tcttggagata tcttggagata
600
gggggaggg ggcggccggc ttcgggccct tctggccgct ccggcgccgc ggcggcgcag cgggtctgcc
660
ggcgcgcgcg ccggcagaggt gcggcgcttg gcggcgcttg gcggcgcttg gcggcgcttg
720
tggcgggtcg tggcgggtcg tggcgggtcg tggcgggtcg tggcgggtcg tggcgggtcg
780
ctgctgaggg gggcagactc gcttcgctgag gggcagactc gcttcgctgag gggcagactc
840
gggggaggg ggcggccggc ttcgggccct tctggccgct ccggcgccgc ggcggcgcag cgggtctgcc
900
gggttggagata tcttggagata tcttggagata tcttggagata tcttggagata tcttggagata
960
<420> SEQ ID NO: 516
<421> NAME: DNA
<422> ORIGIN: Artificial Sequence
<423> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
-continued

gccagagaca ggcaatcttc tgaattttgg accacaactgc tacactaccc aagttcttac
1980
tagatagcgg agaataacac acaacaaacc tgtgaggag aggagacttg agaatccata
2040
aaattataaat taagagacct cctaaagtgc caactatggt gtggctaaat cccgtctctct
2100
ttcctgtggt atttgggtat taccttttc ctgcttcgct tgtgtgggtgt cttttgtacc
2160
tggagtgtct tttaaaagcc tccttcttcgt gatgtgtgtt gtgttgtctg actactctcc
2220
cgcatctcaaa gctggtggcc ttttgccttc ataagggagg ctgcagagat tggacgctg
2280
atttgaggg tgggggtgagg aaattttttaa acaacatttaa ttgtttttta attttttttt
2340
taaaaatggt gttattataa aacaactcata aaaaatgttat gaggagttgag gtggactaat
2400
atataaagag tcctcttctt ataaaaagae tattaagget ttgtctttta tacattacata
2460
tttttaaaaa tgcctttttt tcagaacaaa ggttctttaa gtttttgatg acaagagacag
2520
tgcctgcctt taatacagtt ttcttgggttc aaatacagaa aataaacaatttt cttctagcgc
2580
aagtcaaanta taagaatttt tcggcttgaac ggttcttcttg tttaacattg atatatata
2640
gtagtcttttt ttgagggtcag gttctctacta tgtatctctg ctgtagtctgc aacaaagat
2700
tctgaactctc ggctttctggt cttttgtcttc ctgcttaacag agtataaatct gtgatatac
2760
ttttggaact actgaggtctg atcttttttta aggagacaca cttaagggag tgggtggtata
2820
gttgtgaact gggtttattg gtagaataag actaatgttt tttgtttttt tgtttaatgc
2880
aaaaaagag ccagctgcttt tagtagagatg ttaagggatc octagagact gcgtgcagaca
2940
cagagatgt tacagcttgg gcagtcgcttg aagacatttta cttgaaaaa aacagagagga
3000
ccagaaacct acggttaagg ttgcctttgtct ttaacaccct tgcataacact cgcaggtat
3060
aaaaaattaac aataatatttt cccttacttac acaatctgagga actgtatcaggttcttc
3120
tcgggtgaat cggcagagtt gcagataaaa ccattgctgg tttctttgtt tgtagaatgt
3180
tttctcagtt taagtttaat ccctttgagat acaacaagaca gttgtttttat cagagttcct
3240
tgtgttatac tcgaaggaac ttcttggggt tttctgcctac tgtctctgtatc gttgcctagg
3300
cctacttcat gcgtgtgttt gttctaaatct actatatgttt ctgctcgagtt atttgtatat
3360
gatattcata acaagagcaca aacagtgtt aatagtttaac aatrtatctt taaaggtaaa
3420
gttaggtcttt ttgagttttt aatatttaatt tctaaatcct cgggctaatc agctatcattt
3480
aatggtttca cagttggttg aatgttagaa ttggcttttc gcagccagac tctgcactt
3540
agtaacccta cctataagcg acttctgtgg tgcacagaac gcagagctaa agcattcacc
3600
tctggaggaag gcagttttaat tctggttatac tgcagttttt tgcagagagc atgtttatac
3660
agaagaattt cggcaatatt tttagctgg cttaaggnatt gattctgtatg acgactcattag
3720
gttctgttctt acgtgcctt atgctgaatg aagttctcttac tgtgctatct aataataagt
3780
ggctttttttt aaaccagcct ttcttttttc ctctataact atgtttttcactttttactc
3840
ggaatacttg ttgctagctt ttttcggcaca ctaattgtgt cttcgctgaoc atatttttttt
3900
tctttttttttttttttt ggttcttctg cctcagagtt gctgtgtacc ccctgtgctgt
3960
cctgaacttc acctttggaa cccggtggtgc ctgaacctca gaaatagccgc tgcctcttgc
4020
tctgtactgc ccggatatatt ccgcggacccca ccaacagctg gttactgtttt gatatttca
4080
ataatactt aaactaaaac aacaaactgtct gtctgagtat cttcagcttt gttgagtttcta
4140
agtaggtttt attgctgctct cattaaact aacccctctgt tgggatatt tttttactgct
4200
tctgagactac ttacagcact ttctagcctg atacagcag ttaaggggtc agggattagaga
4260
actgactagag gctttttgtt acaactcata tgtgccttca ggttagaataat
4320
tcaggaactc tagctt

ccccgtgg gcacctgc gagaatc tggacactgc gtttaagcag 60
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 120
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 180
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 240
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 300
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 360
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 420
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 480
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 540
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 600
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 660
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 720
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 780
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 840
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 900
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 960
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1020
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1080
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1140
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1200
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1260
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1320
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1380
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1440
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1500
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1560
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1620
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1680
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1740
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1800
ccccgtgg gcacctgc gagaatc tggacactgc tccacactca 1846
<210> SEQ ID NO: 519
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 519

gagggccat ttcgccatgat tcc 23

<210> SEQ ID NO: 520
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 520

cctgtgaaaa ggaagaaaac cc

<210> SEQ ID NO 521
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified base
<222> LOCATION: (41) .. (23)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 521

aasnnnnnn nnnnnnnnn nnggttgtt cgtcttttcc acaag

<210> SEQ ID NO 522
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 522

cotgagtgtt gaggccocac tgtctgct

<210> SEQ ID NO 523
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 523

acgggcccag agtgcgtgctt gctgtgggcc aggcccc

<210> SEQ ID NO 524
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 524

catcaggcct tcagctccac ctgaggttgtg aggccctgcct gcgcagggcc ctggctgggc

ccaaagctg

<210> SEQ ID NO 525
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 525

aggcgcatt ttcccatgtat tcttca
<210> SEQ ID NO 526
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (84)..(103)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 526

aaaaaagca cggacctgt gcgccttttt caagtgata acggactgc ctatattaa 60
cgtctatatt ctatctetaa aacnnnnnnn nnnnnnnnn nnnggttctt cgctcttttee 120
acaaag 125

<210> SEQ ID NO 527
<211> LENGTH: 111
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9) ..(29)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 527
gaaacacmnn nnnnnnnnn nnnnnnnngt ttagagcta gaatagcga aatggataaa 60
ggtgtagc cgttctactt gaaaaagttt gcggagtcg gtgccttttt t 111

<210> SEQ ID NO 528
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 528

agctgtttta ctgctggct 20

<210> SEQ ID NO 529
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 529

aatggataca cctgggcgaa 20

<210> SEQ ID NO 530
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 530

caatggatac acctggcgc 20
What is claimed is:

1. A method of altering expression of at least one gene product comprising introducing into a eukaryotic cell containing and expressing a DNA molecule having a target sequence and encoding the gene product an engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) system comprising one or more vectors comprising:

   a) a first regulatory element operable in a eukaryotic cell operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with the target sequence, and

   b) a second regulatory element operable in a eukaryotic cell operably linked to a nucleotide sequence encoding a Type-II Cas9 protein, wherein components (a) and (b) are located on same or different vectors of the system, wherein the Cas9 protein comprises one or more mutations in a catalytic domain whereby the Cas9 protein is a nickase that cleaves only one strand of the DNA molecule, whereby the guide RNA targets the target sequence and the Cas9 protein nicks the DNA molecule, whereby expression of the at least one gene product is altered; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.

2. The method of claim 1, wherein the method further comprises the insertion of a recombination template into the nicked DNA molecule.

3. The method of claim 1, wherein the expression of two or more gene products is altered.

4. The method of claim 1, wherein the CRISPR-Cas system comprises a trans-activating cr (tracr) sequence.

5. The method of claim 1, wherein the Cas9 protein comprises a mutation, wherein the mutation comprises D10A, E762A, H840A, N854A, N863A or D986A with reference to the position numbering of a Streptococcus pyogenes Cas9 protein.

6. The method of claim 1, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

7. The method of claim 1, wherein the eukaryotic cell is a mammalian or human cell.

8. The method of claim 1, wherein the expression of one or more gene products is altered by genome editing.

9. The method of claim 1, wherein the expression of one or more gene products is increased.

10. The method of claim 1, wherein the expression of one or more gene products is decreased.

11. The method of claim 1, wherein the one or more vectors are viral vectors.

12. The method of claim 1, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

13. A CRISPR-Cas system-mediated genome editing method comprising introducing into a eukaryotic cell containing and expressing a DNA molecule having a target sequence and encoding at least one gene product an engineered, non-naturally occurring CRISPR-Cas system comprising one or more vectors comprising:

   a) a first regulatory element operable in a eukaryotic cell operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with the target sequence, and

   b) a second regulatory element operable in a eukaryotic cell operably linked to a nucleotide sequence encoding a Type-II Cas9 protein, wherein components (a) and (b) are located on same or different vectors of the system, wherein the Cas9 protein comprises one or more mutations in a catalytic domain whereby the Cas9 protein is a nickase that cleaves only one strand of the DNA molecule, whereby expression of the at least one gene product is altered through the CRISPR-Cas system acting as to the DNA molecule comprising the guide RNA directing sequence-specific binding of the CRISPR-Cas system, whereby there is genome editing; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.

14. The method of claim 13, wherein the method further comprises the insertion of a recombination template into the nicked DNA molecule.

15. The method of claim 13, wherein the expression of two or more gene products is altered.

16. The method of claim 13, wherein the CRISPR-Cas system comprises a tracr sequence.

17. The method of claim 13, wherein the Cas9 protein comprises a mutation, wherein the mutation comprises D10A, E762A, H840A, N854A, N863A or D986A with reference to the position numbering of a Streptococcus pyogenes Cas9 protein.

18. The method of claim 13, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

19. The method of claim 13, wherein the eukaryotic cell is a mammalian or human cell.

20. The method of claim 13, wherein the expression of one or more gene products is increased.

21. The method of claim 13, wherein the expression of one or more gene products is decreased.

22. The method of claim 13, wherein the one or more vectors are viral vectors.

23. The method of claim 13, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

24. An engineered, programmable, non-naturally occurring Type II CRISPR-Cas system comprising a Cas9 protein and at least one guide RNA that targets and hybridizes to a target sequence of a DNA molecule in a eukaryotic cell, wherein the DNA molecule encodes and the eukaryotic cell expresses at least one gene product and wherein the Cas9
protein comprises one or more mutations in a catalytic domain whereby the Cas9 protein is a nickase that cleaves only one strand of the DNA molecule, whereby expression of the at least one gene product is altered; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.  

25. The CRISPR-Cas system of claim 24, wherein the CRISPR-Cas system comprises a tract sequence.  

26. The CRISPR-Cas system of claim 24, wherein the Cas9 protein comprises a mutation, wherein the mutation comprises D10A, E762A, H840A, N854A, N863A or D986A with reference to the position numbering of a Streptococcus pyogenes Cas9 protein.  

27. The CRISPR-Cas system of claim 26, wherein the Cas9 protein comprises the mutation D10A.  

28. The CRISPR-Cas system of claim 26, wherein the Cas9 protein comprises the mutation H840A.  

29. The CRISPR-Cas system of claim 24, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.  

30. The CRISPR-Cas system of claim 24, wherein the eukaryotic cell is a mammalian or human cell.