CRISPR-CAS COMPONENT SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION

Applicants: The Broad Institute, Inc., Cambridge, MA (US); Massachusetts Institute of Technology, Cambridge, MA (US); President and Fellows of Harvard College, Cambridge, MA (US)

Inventors: Le Cong, Cambridge, MA (US); Feng Zhang, Cambridge, MA (US)

Assignees: The Broad Institute Inc., Cambridge, MA (US); Massachusetts Institute of Technology, Cambridge, MA (US); President and Fellows of Harvard College, 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.: 14/259,420
Filed: Apr. 23, 2014

Prior Publication Data

Related U.S. Application Data
Continuation of application No. 14/105,035, filed on Dec. 12, 2013.


Int. Cl.
C12Q 1/68 (2006.01)
C12N 15/85 (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)

U.S. Cl.
CPC .................................................. C12N 15/85 (2013.01)
USPC .............. 435/6.1; 435/6.13; 435/195; 435/199; 435/220; 435/320.1; 424/94.1; 424/94.6; 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

FOREIGN PATENT DOCUMENTS
WO WO/2012/164656 12/2012
WO WO/2013/008244 7/2013
WO WO/2013/176772 11/2013
WO WO/2014/065956 5/2014
WO WO/2014/089290 6/2014

OHER PUBLICATIONS
U.S. Appl. No. 61/734,256, filed Dec. 6, 2013, Fuqiang Chen et al.

Primary Examiner — Anne Gussow
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. 1
FIG. 2B
FIG. 2D
E

Target locus

5' - AGCTGGAGGGAGGAGGGCCCTGAATCCGAGGAAGAAGGAGGCTCCCACT-3'

3' - TCGACCTCCCTCTCTCCGGGGACTAGGCTGCTTCTTCTCTCCGGGCTG-5'

crRNA.

5' - AAGUCCGGAGGAGAACGAAGAGUUGUAGAGC-3'

indel

AGCTGGAGGGAGGAGGGCCCTGAATCCGAGGAAGAAGGAGGCTCCCACT

F

human EMX1 protospacer target (mutation in 5 of 43 sequenced clones = 11.6%)

WT

5' - CTGGAGGGAGGGGCTGAGTCCGAGGAAGAAGAAGGCTCCCATCAG-3'

Δ1

CTGGAGGAGGAAGGGGCTGAGTCCGAGGAAGAAGGCTCCCATCAG

+1

CTGGAGGAGGAAGGGGCTGAGTCCGAGGAAGAAGGCTCCCATCAG

Δ3

CTGGAGGAGGAAGGGGCTGAGTCCGAGGAAGAAGGCTCCCATCAG

m1, Δ6

CTGGAGGAGGAAGGGGCTGAGTCCGAGGAAGAAGGCTCCCATCAG

FIG. 2E-F
FIG. 3
A

human EMX1 locus

wt crRNA

mismatch-containing guide sequences

B

mismatched spacers

694bp

357bp

317bp

indel (%) 5.6 7.5 8.8 9.7

C

left TALEN binding site

right TALEN binding site

FIG. 4A-C
FIG. 4D
C

human EMX1 locus

HR Template

D

hSpCas9 + − −
hSpCas9n − + −
HR template + + +

2281bp

1189bp
1092bp

HR (%) 0.70 0.46

E

AATGACAAGCTTGCAGCGGTGG

HindIII NheI

FIG. 5C-E
FIG. 5G
<table>
<thead>
<tr>
<th>Cas9</th>
<th>target 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 + tracrRNA)</th>
<th>% indel (chimeric RNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Homo sapiens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMX1</td>
<td></td>
<td>1</td>
<td>GAGGAGGCGCTGATCGCGACGAGAAGAA</td>
<td>GGG</td>
<td>+</td>
<td>293FT</td>
<td>20 ± 1.8</td>
<td>6.7 ± 0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>CATGGAGGGCTGATCGCGACGAGAAGAA</td>
<td>TGG</td>
<td>−</td>
<td>293FT</td>
<td>2.1 ± 0.31</td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>GAAATAGGCTGATCGCGACGAGAAGAA</td>
<td>TGG</td>
<td>+</td>
<td>293FT</td>
<td>14 ± 1.1</td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>CAGGAGGCGCTGATCGCGACGAGAAGAA</td>
<td>TGG</td>
<td>−</td>
<td>293FT</td>
<td>11 ± 1.7</td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>CTCTGAGGCTGATCGCGACGAGAAGAA</td>
<td>TGG</td>
<td>−</td>
<td>293FT</td>
<td>4.3 ± 0.48</td>
<td>2.1 ± 0.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>TGGGAGGCGCTGATCGCGACGAGAAGAA</td>
<td>TGG</td>
<td>−</td>
<td>293FT</td>
<td>4.0 ± 0.65</td>
<td>0.41 ± 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>TCAAGCTGTCCTGCTGACGAGAAGAA</td>
<td>CAG</td>
<td>−</td>
<td>293FT</td>
<td>1.5 ± 0.12</td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>GAGGAGGCGCTGATCGCGACGAGAAGAA</td>
<td>AGG</td>
<td>−</td>
<td>293FT</td>
<td>7.8 ± 0.83</td>
<td>2.3 ± 1.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. pyogenes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SF370 type II</td>
<td>PV1B</td>
<td>9</td>
<td>AGGGCCGAGATTGCTCTGACGACGACGAG</td>
<td>AGG</td>
<td>+</td>
<td>293FT</td>
<td>21 ± 2.6</td>
<td>5.3 ± 0.32</td>
</tr>
<tr>
<td></td>
<td>CRISPR</td>
<td>PV1B</td>
<td>10</td>
<td>ATGACGAGAGGCGAGGAGGGCGAGAAGAT</td>
<td>TGG</td>
<td>+</td>
<td>293FT</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PV1B</td>
<td>11</td>
<td>GTCTAGGACGGAGGCGAGGAGGGCGAGAAG</td>
<td>AGG</td>
<td>+</td>
<td>293FT</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>Homo sapiens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Th</td>
<td>12</td>
<td>CGCTGACTGCTGCTGACGAGAAGAAAGAG</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></td>
<td>Th</td>
<td>13</td>
<td>ATGACGAGAGGAGGAGGAGGAGGAGAAGAG</td>
<td>GGG</td>
<td>−</td>
<td>Neuro2A</td>
<td>4.8 ± 1.2</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Th</td>
<td>14</td>
<td>AGCGCTGAGAGGAGGAGGAGGAGGAGAAGAG</td>
<td>AGG</td>
<td>+</td>
<td>Neuro2A</td>
<td>11.3 ± 1.3</td>
<td>N.D.</td>
</tr>
<tr>
<td></td>
<td>Mus musculus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EEX1</td>
<td>15</td>
<td>GAGCGAGGCTGATCGCGACGAGAAGAAAGA</td>
<td>GTAGAAT</td>
<td>−</td>
<td>293FT</td>
<td>14 ± 0.88</td>
<td>N.T.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EEX1</td>
<td>16</td>
<td>AGATGGTAGAGGAGGAGGAGGAGAAGAAAGAAGA</td>
<td>CTGAGA</td>
<td>−</td>
<td>293FT</td>
<td>7.8 ± 0.77</td>
<td>N.T.</td>
</tr>
</tbody>
</table>

FIG. 6
<table>
<thead>
<tr>
<th>SpCas9</th>
<th>long tracrRNA</th>
<th>short tracrRNA</th>
<th>DSN-EMX1(1)-DH</th>
</tr>
</thead>
<tbody>
<tr>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>~</td>
<td>~</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. 12A-C**
FIG. 13A-C
**FIG. 14A-B**

### Table A

<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>

### Figure B

- **U6**
FIG. 15A-B
<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</td>
<td>EMX1</td>
<td>AAAACCACCTCTCTCTCTG6C</td>
</tr>
<tr>
<td>Sp-EMX1-R</td>
<td>SURVEYOR</td>
<td>EMX1</td>
<td>GGAGATGGAGACAGAAGAG</td>
</tr>
<tr>
<td>Sp-PVALB-F</td>
<td>SURVEYOR</td>
<td>PVALB</td>
<td>CTGGAAGGCAATGCCCTGAC</td>
</tr>
<tr>
<td>Sp-PVALB-R</td>
<td>SURVEYOR</td>
<td>PVALB</td>
<td>GGCAAGAAAACCTCCCTTGCTCC</td>
</tr>
<tr>
<td>Sp-Th-F</td>
<td>SURVEYOR</td>
<td>Th</td>
<td>GTGCTTTGAGAGGCTTACC</td>
</tr>
<tr>
<td>Sp-Th-R</td>
<td>SURVEYOR</td>
<td>Th</td>
<td>CCTGGAGGCGATGCGAGA6GT</td>
</tr>
<tr>
<td>St-EMX1-F</td>
<td>SURVEYOR</td>
<td>EMX1</td>
<td>ACCCTCTGTGTTTCCACCATYC</td>
</tr>
<tr>
<td>St-EMX1-R</td>
<td>SURVEYOR</td>
<td>EMX1</td>
<td>TTTGGGA6GTGACAGACTTC</td>
</tr>
<tr>
<td>Sp-EMX1-RFLP-F</td>
<td>RFLP,</td>
<td>EMX1</td>
<td>GGCTCCCTG6GTCAAAAGTA</td>
</tr>
<tr>
<td>Sp-EMX1-RFLP-R</td>
<td>RFLP,</td>
<td>EMX1</td>
<td>AGGAGGGCTG6ATGCTCGTAA</td>
</tr>
<tr>
<td>Pb_EMX1_sp1</td>
<td>Northern</td>
<td>Not applicable</td>
<td>TAGCTCTAAAACCTCTCTCTGCTGGAC</td>
</tr>
<tr>
<td>Pb_matureRNA</td>
<td>Northern</td>
<td>Not applicable</td>
<td>CTA6GCTATTATTTAAGTGTGCTGTGTTT</td>
</tr>
</tbody>
</table>

FIG. 17
a

U6

chimeric RNA

CBh

NLS SpCas9 bGH pA

chimeric RNA architecture

guide sequence (20 bp)

5' - NNNNNNNNNNNNNNNNNNNNGUUUUGAGCUAG

+54

+448

A

A

GUUCAACUAUUUGCCUGAGCGAAUAAAAUUGAAUA

+1

+47

A

A

AAAGUGGCCACCCA

+85

A

A

GUUGGCU

3' - UUGCGUGGC

FIG. 18A
Protoscaler 1 (EMX1)

Protoscaler 2 (EMX1)

Protoscaler 3 (EMX1)

Protoscaler 4 (PVALB)

Protoscaler 5 (PVALB)

FIG. 21
### FIG. 23A-B

**Diagram a:** Targeting construct and editing template, followed by transformation and Kanamycin selection leading to the edited genome.

**Table b:**

<table>
<thead>
<tr>
<th>Editing Construct</th>
<th>cfu/ug of crR6M DNA</th>
<th>Ed. Un.</th>
<th>Kan&lt;sup&gt;+&lt;/sup&gt; transformants</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>2.6 ± 0.9</td>
<td></td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>R6(srtA)</td>
<td>14 ± 1.8</td>
<td></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>R6&lt;sup&gt;370.1&lt;/sup&gt;</td>
<td>25 ± 2.6</td>
<td></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>
FIG. 27A-D
FIG. 29A-F
FIG. 32
FIG. 34A-D
FIG. 35A-C
FIG. 36
FIG. 38A-B
FIG. 41B
FIG. 41C
FIG. 41D
FIG. 41E
FIG. 41L
hSpCas9

GACCTGTGTACGCTGGRAGCGGRC

D L S Q L G G D

1301 1302 1303 1304 1305 1306 1307 1308

FIG. 41M
FIG. 43A
FIG. 43C
FIG. 43D
FIG. 44A
FIG. 45
FIG. 46
GCACCAACGGCAACGCTGCCCTGGCCACCCCTCTGATCAACCCAGAGCACTTGGACGTG
GGCTGGATGCTGGTCAACGGCCCTCCACCCCTGGGAGACGAGGACGAGGACGAGGACGAGG
CAGCTCTCGAAAGAGCATCCTGGTGAGTCGGACCACGGTGGCACTGGGACGAGGACGAGG
AGTCAGAGCTTGGGCTCGGCAAGAAGAAGCTTGGACGTG
CAGCGTCTCTCGGAGTCGGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGG
ACGACACTATGCAGCAGAAGCAGTGCTCACTCAAGGGATCGAGGTGGAGAGCTTGAG

WPRE
Cgtactttcaacctctggattttcaaaaaatgcgttggtgattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
FIG. 50G
FIG. 50H
Validation of Cas9 nuclease activity by Surveyor
<table>
<thead>
<tr>
<th>Construct</th>
<th>Clone 1</th>
<th>Clone 2</th>
<th>Clone 3</th>
<th>Clone 4</th>
<th>Average</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCAG-loxp(pA)loxp-NLS-hSpCas9-NLS-2A-GFP</td>
<td>32.1</td>
<td>27.3</td>
<td>35.9</td>
<td>39.0</td>
<td>34.8</td>
<td>7.1</td>
</tr>
<tr>
<td>pCAG-NLS-hSpCas9-NLS-2A-GFP</td>
<td>26.9</td>
<td>33.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 54**
FIG. 56

AtplF-Rosa26

1 min exposure

10xpl-Atplp-10xpl-CAG-Cas9 targeting ES

AtplF-Rosa26

3 min exposure

CAG-Cas9 targeting ES

1 min exposure
left guide RNA target (L)  spacer (0 - n BP)  right guide RNA target (R)

5' - NNNNGGG . . . CCNNNNN - 3'
3' - NNNNNCC . . . GNNNN - 5'

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

FIG. 58
FIG. 63A-C
FIG. 64
### Table 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>gRNA1</th>
<th>gRNA2</th>
<th>gRNA3</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCN2A</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>ND</td>
</tr>
<tr>
<td>KATNAL2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>4.6</td>
</tr>
<tr>
<td>KCTD13</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>6.0</td>
</tr>
<tr>
<td>CHD8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>6.8</td>
</tr>
<tr>
<td>CACNA1C</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>ND</td>
</tr>
<tr>
<td>CACNB2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>9.4</td>
</tr>
<tr>
<td>GIT1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>6.1</td>
</tr>
<tr>
<td>MECP2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>2.7</td>
</tr>
<tr>
<td>SHANK3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>20.2</td>
</tr>
</tbody>
</table>

**Note:**
- 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 - agctgttttactggtcgct
Chd8.2 - aatggatacaccctgtcga
Chd8.3 - caatggatacaccctgtcga

FIG. 66
Cas9

Cas9 with N-term NLS

Cas9 with C-term NLS

Cas9 with Double NLS

FIG. NLS1

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

Percentage of Endogenous Genome Cleavage

Construct NLS positions

FIG. 68
b

FIG. 69A-B
NLS architecture optimization for SpCas9

FIG. 70
FIG. 71

FIG. 72
CRISPR-CAS COMPONENT SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION

RELATED APPLICATIONS AND INCORPORATION 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.


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.

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 Feb. 14, 2014, is named 44790.07.2003_SL.txt and is 305,328 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 eukaryote 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 a 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 tracer mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracer 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 tracer mate sequence that is hybridized to the tracer 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 tracer sequence downstream of the tracer 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 sequence 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 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 publicly and commercially available alignment algorithms and programs such as, but not limited to, ClustalW, Smith-Waterman in matlub, 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. Without wishing to be bound by theory, it is believed 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 nuclear 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 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 varieties 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 sequences 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 ENZYMOLOGY 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 III 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-1 (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, the 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 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 sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr 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 comprised with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr 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 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 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%, or 95%, or 99% of sequence complementarity along the length of the tracr 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.
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 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 mutations 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 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. 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 with an increase in the risk of having or developing a disease. In some embodiments, the method comprises introducing one or more vectors into an eukaryotic cell, 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 mutations result 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 with an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) contacting a test compound with a model cell of any one of the described embodiments; and (b) 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 a 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: a 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 counter-selection 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 making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. §112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve the right and hereby disclose a disclaimer of any previously described product, process 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 protospecky nucleotide motif (PAM; magenta), and Cas9 mediates a double-stranded break (DSB) ~3 bp upstream of the PAM (red triangle).

**FIGS. 2A-F** show 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 NO. 279-280, respectively, in order of appearance. FIG. 2D discloses SEQ ID NO. 281-283, respectively, in order of appearance. FIG. 2E discloses SEQ ID NO. 284-288, 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 NO. 289-298, respectively, in order of appearance).

**FIGS. 4A-D** show results of an evaluation of SpCas9 specificity for an example target. FIG. 4A discloses SEQ ID NO. 299, 302 and 300-310, respectively, in order of appearance. FIG. 4C discloses SEQ ID NO. 299.

**FIGS. 5A-G** show an exemplary vector system and results for its use in directing homologous recombination in eukaryotic cells. FIG. 5E discloses SEQ ID NO. 311. FIG. 5F discloses SEQ ID NO. 312-313, respectively, in order of appearance. FIG. 5G discloses SEQ ID NO. 314-318, respectively, in order of appearance.

**FIG. 6** provides a table of protospacer sequences (SEQ ID NO. 33, 32, 31, 322-327, 35, 34 and 330-334, 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).

**FIGS. 7A-C** show a comparison of different tracrRNA transcripts for Cas9-mediated gene targeting. FIG. 7A discloses SEQ ID NO. 335-336, 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.

**FIGS. 9A-B** show exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells. FIG. 9A discloses SEQ ID NO. 337-339, respectively, in order of appearance. FIG. 9B discloses SEQ ID NO. 340-342, respectively, in order of appearance.

**FIG. 10A-D** shows a bacterial plasmid transformation interference assay, expression cassettes and plasmids used therein, and transformation efficiencies of cells used therein. FIG. 10A discloses SEQ ID NO. 343-345, respectively, in order of appearance. FIG. 10B discloses SEQ ID NO. 346-347, respectively, in order of appearance. FIG. 10C discloses SEQ ID NO: 348.

**FIGS. 11A-C** show histograms of distances between adjacent S. pyogenes SF370 locus 1 PAM (NGG) (FIG. 10A) and S. thermophilus LMD9 locus 2 PAM (NNAGAAW) (FIG. 10B) in the human genome; and distances for each PAM by chromosome (Chr) (FIG. 10C).

**FIGS. 12A-C** show an exemplary CRISPR system, an example adaptation for expression in eukaryotic cells, and results of tests assessing CRISPR activity. FIG. 12B discloses SEQ ID NO. 346-347, respectively, in order of appearance. FIG. 12C discloses SEQ ID NO: 348.

**FIGS. 13A-C** show exemplary manipulations of a CRISPR system for targeting of genomic loci in mammalian cells. FIG. 13A discloses SEQ ID NO: 349. FIG. 13B discloses SEQ ID NO. 350-352, respectively, in order of appearance. FIGS. 14A-B show the results of a Northern blot analysis of crRNA processing in mammalian cells. FIG. 14A discloses SEQ ID NO: 353.

**FIG. 15A-B** shows a selection of protospacers in the human PVALB and mouse Th loci. FIG. 15A discloses SEQ ID NO: 354. FIG. 15B discloses SEQ ID NO: 355.

**FIG. 16** shows example protospacer and corresponding PAM sequence targets of the S. thermophilus CRISPR system in the human EMX1 locus (SEQ ID NO: 348).

**FIG. 17** provides a table of sequences for primers and probes (SEQ ID NO. 36-39 and 356-363, respectively, in order of appearance) used for Surveyor, RFLP, genomic sequencing, and Northern blot assays.

**FIGS. 18A-C** show 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: 364, respectively, in order of appearance.
FIGS. 19A-B show a graphical representation of the results of SURVEYOR assays for CRISPR system activity in eukaryotic cells (SEQ ID NO 365-443, respectively, in order of appearance). FIG. 20 shows an exemplary visualization of some S. pyogenes Cas9 target sites in the human genome using the UCSC genome browser. FIG. 2I shows predicted secondary structures for exemplary chimeric RNAs comprising a guide sequence, tracr mate sequence, and tracr sequence (SEQ ID NO 444-463, respectively, in order of appearance). FIG. 22 shows exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells (SEQ ID NO 464 and 341-342, respectively, in order of appearance). FIG. 23A-B 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. tracr, trans-activating CRISPR RNA; aphA-3, kanamycin resistance gene. (b) Transformation of crR6M DNA in R68232.5 cells with no editing template, the R6 wild-type srTA or the R6370.1 editing templates. Recombination of either R6 srTA 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 srTA locus of strain R68232.5; "Ed." shows the editing template. R68232.5 and R6370.1 targets are distinguished by restriction with EcoI.

FIG. 24A-C shows analysis of PAM and seed sequences that eliminate Cas9 cleavage. (a) PCR products with randomized PAM sequences or randomized seed sequences were transformed in crR6 cells (SEQ ID NO 465-469, respectively, in order of appearance). These cells expressed Cas9 loaded with a crRNA that targeted a chromosomal region of R68232.5 (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: 470). 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. 25A-F 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 471-475 and 472, 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 BigI (R-→A) and Tsel (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 srTA and bgaA) and was co-transformed with two different editing templates (f) PCR analysis for 8 transformants to detect deletions in srTA and bgaA loci. 6/8 transformants contained deletions of both genes.

FIG. 26A-D 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 476-479, respectively, in order of appearance). (c) Fraction of erythromycin-resistant (erm8) cfu calculated from total or kanamycin-resistant (kan8) 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±s.d. for three independent experiments.

FIG. 27A-D 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 hME63 recombining strain containing a chloramphenicol-resistant plasmid harboring cas9 and tracr (pCas9), together with an oligonucleotide specifying the mutation. (b) A K427 mutation conferring streptomycin resistance was introduced in the rpsL gene (SEQ ID NO 480-483, respectively, in order of appearance) (c) Fraction of streptomycin-resistant (strpm) cfu calculated from total or kanamycin-resistant (kan8) 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±s.d. for three independent experiments.

FIG. 28A-B 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 CRISPRR01 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 apha-A gene encodes kanamycin resistance. Protospeakers from the streptococcal bacteriophages φ8232.5 and φ730.1 were fused to a chloramphenicol resistance gene (cat) and integrated in the srTA gene of strain R6 to generate strains R68232.5 and R6870.1. (b) Left panel: Transformation of crR6 and crR6M genomic DNA in R68232.5 and R6870.1. As a control of cell competence a streptomycin resistant gene was also transformed. Right panel: PCR analysis of 8 R68232.5 transformants with crR6 genomic DNA. Primers that amplify the srTA locus were used for PCR. 7/8 genotyped colonies replaced the R68232.5 srTA locus by the WT locus from the crr6 genomic DNA.

FIG. 29A-F provides chromatograms of DNA sequences of edited cells obtained in this study. In all cases the wild-type and mutant protospeakers 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) Chromatogram for the introduction of a PAM mutation into the R632 target (FIG. 23D) (SEQ ID NOS 484-485, respectively, in order of appearance). (b) Chromatograms for the introduction of the R->A and NE->AA mutations into β-galactosidase (bgaA) (FIG. 25c) (SEQ ID NOS 471-475 and 472, respectively, in order of appearance). (c) Chromatogram for the introduction of a 6664 bp deletion within bgaA ORF (FIGS. 25c and 25f). The dotted line indicates the limits of the deletion (SEQ ID NOS 486-488, respectively, in order of appearance). (d) Chromatogram for the introduction of a 729 bp deletion within srtA ORF (FIG. 25f). The dotted line indicates the limits of the deletion (SEQ ID NOS 489-491, respectively, in order of appearance). (e) Chromatograms for the generation of a premature stop codon within ermAM (FIG. 33) (SEQ ID NOS 492-495, respectively, in order of appearance). (f) rpsL editing in E. coli (FIG. 27) (SEQ ID NOS 480-483, respectively, in order of appearance).

FIG. 30A-C 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 pD399-108 (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. NGGs 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. 33A-D illustrates CRISPR-mediated editing of the ermAM locus using genomic DNA as targeting construct. To use genomic DNA as targeting construction 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 492-495, 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 srtA 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 or control templates. In the presence of the control template 5.4x10^6 cfu/ml were obtained, and 4.3x10^7 cfu/ml when the editing template was used. This difference indicates an editing efficiency of about 99% ([(4.3x10^5-5.4x10^5)/4.3x10^7]). (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. 34A-D 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 srtA-targeting construct fused to cat for chloramphenicol selection of edited cells, along with an editing construct for a ΔsrtA in-frame deletion. Strain crR6 ΔsrtA is generated by selection on chloramphenicol. Subsequently, the ΔsrtA strain is co-transformed with a bgaA-targeting construct fused to aphA-3 for kanamycin selection of edited cells, and an editing construct containing a ΔbgaA 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 bgaA protospacer (pD897), and selection on spectinomycin. (b) PCR analysis for 8 chloramphenicol (Cam)-resistant transformants to detect the deletion in the srtA locus. (c) β-galactosidase activity as measured by Miller assay. In S. pneumoniae, this enzyme is anchored to the cell wall by sortase A. Detection of the srtA gene results in the release of β-galactosidase into the supernatant. ΔbgaA mutants show no activity. (d) PCR analysis for 8 spectinomycin (Spec)-resistant transformants to detect the replacement of the CRISPR locus with wild-type IS1167.

FIG. 35A-C 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^R 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^-3. (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 496-499, 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 343, 500 and 127, respectively, in order of appearance). Oligonucleotide design is shown at bottom, pCas9 carried chloramphenicol resistance (CamR) and is based on the low-copy pACYC184 plasmid backbone. pCRISPR is based on the high-copy number pZET21 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 (WS524) carrying a point mutation that both confers streptomycin resistance and abolishes CRISPR immunity, together with a plasmid targeting ppsL (pCRISPR:ppsl) 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. 38A-B illustrates the background mutation frequency of CRISPR in E. coli HME63. (a) Transformation of the pCRISPR::O or pCRISPR::ppsl 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 escapers showed that 8/8 have deleted the spacer.

FIGS. 39A-D show 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).

FIGS. 40A-F show 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 501-502, 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. 43A-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 indicates sgRNA cleavage site; PCR primers for genotyping (Tables J 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 503-505, 503-506 and 505 respectively, in order of appearance). Arrows indicate positions of expected fragment sizes.

FIGS. 44A-B shows single vector designs for SpCas9. FIG. 44A discloses SEQ ID Nos 320-321 and 328, respectively, in order of appearance. FIG. 44B discloses SEQ ID NO: 329.

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

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

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

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

FIG. 49A shows the Conditional Cas9, Rosa26 targeting vector map.

FIG. 49B shows the Constitutive Cas9, Rosa26 targeting vector map.

FIG. 50A-H show the sequences of each element present in the vector maps of FIGS. 49A-B (SEQ ID Nos 507-516, respectively, in order of appearance).

FIG. 51 shows a schematic of the important elements in the Constitutive and Conditional Cas9 constructs.

FIG. 52 shows the functional validation of the expression of Constitutive and Conditional Cas9 constructs.

FIG. 53 shows the validation of Cas9 nuclelease activity by Surveysor.

FIG. 54 shows the quantification of Cas9 nuclelease activity.

FIG. 55 shows construct design and homologous recombination (HR) strategy.

FIG. 56 shows the genomic PCR genotyping results for the constitutive (Right) and conditional (Left) constructs at two different gel exposure times (top row for 3 min and bottom row for 1 min).

FIG. 57 shows Cas9 activation in mESCs.

FIG. 58 shows a schematic of the strategy used to mediate gene knockout via NHEJ using a nickase version of Cas9 along with two guide RNAs.

FIG. 59 shows how DNA double-strand break (DSB) repair promotes gene editing. In the error-prone non-homologous end joining (NHEJ) pathway, the ends of a DSB are processed by endonuclease repair machineries and rejoined together, which can result in random insertion/deletion (indel) mutations at the site of junction. Indel mutations occurring within the coding region of a gene can result in frame-shift and a premature stop codon, leading to gene knockout. Alternatively, a repair template in the form of a plasmid or single-stranded oligodeoxynucleotides (ssODN) can be supplied to leverage the homology-directed repair (HDR) pathway, which allows high fidelity and precise editing.

FIG. 60 shows the timeline and overview of experiments. Steps for reagent design, construction, validation, and cell line expansion. Custom sgRNAs (light blue bars) for each target, as well as genotyping primers, are designed in silico via our online design tool (available at the website genome-engineering.org/tools). sgRNA expression vectors are then cloned into a plasmid containing Cas9 (PX330) and verified via DNA sequencing. Completed plasmids (pCRISPRs), and optional repair templates for facilitating homology directed repair, are then transfected into cells and assayed for ability to mediate targeted cleavage. Finally, transfected cells can be clonally expanded to derive isogenic cell lines with defined mutations.

FIG. 61A-C shows Target selection and reagent preparation. (a) For S. pyogenes Cas9, 20-bp targets (highlighted in blue) must be followed by 5'-NGG, which can occur in either strand on genomic DNA. We recommend using the online tool described in this protocol in aiding target selection (www-genome-engineering.org/tools). (b) Schematic for co-transfection of Cas9 expression plasmid (PX165) and PCR-amplified U6-driven sgRNA expression cassette. Using a U6 promoter-containing PCR template and a fixed forward primer (U6 Fwd), sgRNA-encoding DNA can be appended onto the U6 reverse primer (U6 Rev) and synthesized as an extended DNA oligo (Ultramer oligos from IDT). Note the guide sequence (blue N's) in U6 Rev is the reverse complement of the 5'-NGG flanking target sequence (SEQ ID Nos 517 and 517-519, respectively, in order of appearance). (c) Schematic for scarless cloning of the guide sequence oligos into a plasmid containing Cas9 and sgRNA scaffold (PX330). The guide oligos (blue N's) contain overhangs for ligation into the pair of BbsI sites on PS330, with the top and bottom strand orientations matching those of the genomic target (i.e. top oligo is the 20-bp sequence preceding 5'-NGG in genomic DNA). Digestion of PX330 with BbsI allows the replacement of the Type IIs 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 520-521 and 328, 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 reannealed slowly to generate heteroduplexes. The reannealed 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 DYRK1A 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 Fwd, 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 520-522, 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 SEQS ID NOS 505-505, 503, 506 and 505, 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 carrying the reverse complement of part of the U6 promoter, the sgRNA(+)5 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 517, 523, 518 and 524-525, 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 strand break at the site of a mutation. Each gRNA was validated in the mouse cell line, Neuro-2a, by liposomal transient co-transfection with hSpCas9. 72 hours post-transfection genomic DNA was purified using QuickExtract DNA 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 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 (SEQ ID NOS 526-528, respectively, in order of appearance).

FIG. 67 shows 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 scar 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 5 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), openly linked to triple NLS and a VP64 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-VP64 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-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.

FIG. 70 depicts NLS architecture optimization for SpCas9.

FIG. 71 shows a QQ plot for NGGNN sequences.

FIG. 72 depicts 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 bgaA expression by dgRNA:cas9g. 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 bgaA promoter and represses transcription. b. The targets used to direct Cas9g to the bgaA promoter are represented (SEQ ID NO: 529). Putative –35, –10 as well as the bgaA start codon are in bold. c. Betagalactosidase activity as measured by Miller assay in the absence of targeting and for the four different targets.
FIG. 74A-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 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 of 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 polynucleotides, 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 nucleic acid structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide 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 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-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 Tijssen (1993). Laboratory Techniques in Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter “Overview of principles of hybridization and the strategy of nucleic acid probe assay”. Elsevier, N.Y.

“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, lipiddation, 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 a reduction in an effect of dysfunction; 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 prophylactic benefit, it 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 achieve 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 may 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 ENZYMOLOGY 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 pRT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.


In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast Saccharomyces cerevisiae include pYepSec1 (Baldari et al., 1987, EMBO J. 6: 229-234), pMP9a (Kuijers and Heskenswitz, 1982, Cell 30: 933-943), pJRY88 (Schultz et al., 1987, Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and pizZ (InVitrogen Corp, San Diego, Calif.).


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 pcDNA8 (Seed, 1987. Nature 329: 840) and pmT2PC (Kaufman et al., 1987. EMBO J 6: 187-195). When used in mammalian cells, the 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 known in the 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 ed., 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

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 SPIDRs (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 *Halofexa mediterranei*, *Streptococcus pyogenes*, *Anaeroba*, and *Mycobacterium tuberculosis* (See, Groenen et al., Mol. Microbiol., 10:1057-1065 [1993]; Hoe et al., Emerg. Infect. Dis., 5:254-263 [1999]; Masepol et al., Biochim. 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 (SSRRs) (Janssen et al., OMICS J. Integr. Biol., 6:23-33 [2002]; and Mojica et al., Mol. Microbiol., 36:244-246 [2000]). In general, the repeat sequences are short and differ 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., 182:2393-2401 [2000]). CRISPR loci have been identified in more than 40 prokaryotes (See, e.g., Janssen et al., Mol. Microbiol., 43:1565-1575 [2002]; and Mojica et al., [2005]) including, but not limited to *Aeropyrum*, *Pyrobaculum*, *Sulfobolus*, *Archeoglobus*, *Haloarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azuribacterium*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Mycococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pastenrella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Trepomonas*, 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 tracer (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing 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 tracr 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 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 a 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 sites being located between two tracr mate sequences so as to allow insertion of a guide sequence at each site. In such an arrangement, 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 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 Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csa2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr5, Cmr6, Cas1b, Cas2b, Csb3, Csx17, Csx14, Csx10, Csx16, Cas9, Csx3, Csx1, Csx15, Casf1, Casf2, Casf3, 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 Q99Z2W. 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 0.01%, 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, 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 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 Gene Forge (Aptagen; Jacobs, Pa.), are also available. In some embodiments, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, 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, 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 or near the amino-terminus and one or more NLSs at or near 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 sequences derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV (SEQ ID NO: 1); the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPAATKKAQQKAK (SEQ ID NO: 2)); the c-mye NLS having the amino acid sequence PAAKVRKVL (SEQ ID NO: 3) or RQRNRNELKRSP (SEQ ID NO: 4); the hRNP A1 M9 NLS having the sequence NQSRSFGPMKGGGNGFRRSSGPFYQQYAKPRNGGYY (SEQ ID NO: 5); the sequence RMIZFTK KNGGKDVTALRRRVEVSLVEKLKKDQEIQKLKRNV (SEQ ID NO: 6) of the IBB domain from importin-alpha; the sequences VSRKPRPR (SEQ ID NO: 7) and PPKKARED (SEQ ID NO: 8) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO: 9) of human p53; the sequence SAILKKKMKMAP (SEQ ID NO: 10) of mouse c-abl 1V; the sequences DRLKR (SEQ ID NO: 11) and PKQKKRK (SEQ ID NO: 12) of the influenza virus NS1; the sequence RKJLKJJKK (SEQ ID NO: 13) of the Hepatitis virus delta antigen; the sequence REFKKKLKRR (SEQ ID NO: 14) of the mouse Mx1 protein; the sequence KRRGDEVDGDEVAVKSSKK (SEQ ID NO: 15) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTK (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 (HA 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 not 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 complementarity to a target polynucleotide sequence to hybridize 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, Nvovalign (Novocraft Technologies, ELAND ( Illumina, San Diego, Calif.), SOAP (available at soop.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 MMNNNNMNNNNNNNNNNNXXGG (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.
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 the 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 transcript termination:

In general, a tracr mate 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 mate sequences in a cell containing the corres-ponding tracr sequence; and (2) formation of a CRISPR complex at a target sequence, wherein the CRISPR complex comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate 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, 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. Examples 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
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 sequence (e.g., about or more than 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: methylase activity, demethylase activity, transcription 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-glucuronidase, luciferase, green fluorescent protein (GFP), HisRed, 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) VP16 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 or 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 cells 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 can 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); Nobel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillen, 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 & Perriacaudet, British Medical Bulletin 51(1):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, polycation or lipid-nucleic acid conjugates, naked DNA, artificial virions, 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., Transfection™ and Lipofectin®). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, 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); Blase et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Cao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 5,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

The use of RNA or DNA 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 vector 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 transfer 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-acting LTRs 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 Immuno deficiency virus (SW), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Duschker et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol.
A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-7R, CHO-K1, CHO-K2, CHO-T, CHO-DaDa1, COR-L-23, COR-L23/CPR, COR-L-23/S010, COR-L23/R3, COS-7, CON-343, CML T1, CMT, CT26, D17, DH82, DU145, DU01A, EL4, EM2, EM3, EMT6/AR1, EMT6/AR1.0, FM3, H1299, H169, HB54, HB55, ICA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Kn 92, KCL22, KGI, KO1, LNCap, 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, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-1A/ PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, SF-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties 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 or plant is algae. 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 for methods 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 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 that said 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 sequence which in turn hybridizes to a tracr sequence.
tions 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. 7,868,149—Plant Genome Sequences and Uses Thereof and US 2009/0100536—Transgenic Plants with Enhanced Agromatic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Merrill et al. “Crop genomics advances and applications” Nat Rev Genet. 2011 Dec 29; 13(2):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, mutatis mutandis, to plant cells unless otherwise apparent.

In one aspect, the invention provides methods for modifying a target nucleotide 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 microalgae).

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. dianthii Puccinia graminis f. sp. triticci attacks only wheat. 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 one balance changes in 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 tracr mate sequence and one or more insertion sites for inserting a guide sequence 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. Elements may be provided 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 operably link the guide sequence and a regulatory element. In some embodiments, the kit comprises a homologous recombination template nucleotide.

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 nucleotide. The CRISPR complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target nucleotide 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 nucleotide. The guide sequence is linked to a tracr mate sequence, which in turn hybridizes to a tracr sequence.

The target nucleotide of a CRISPR complex can be any nucleotide endogenous or exogenous to the eukaryotic cell. For example, the target nucleotide can be a polynucleotide residue 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 enzyme.
The target polynucleotide of a CRISPR complex may include a number of disease-associated genes and polynucleotides as well as signaling biochemistry pathway-associated genes and polynucleotides as listed in U.S. provisional patent applications 61/736,527 and 61/748,427, both entitled SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION filed on Dec. 12, 2012 and Jan. 2, 2013, respectively, the contents 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 biochemistry 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 transcribed 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 biochemistry pathway-associated genes and polynucleotides are listed in Table C.

Mutations in these genes and pathways may 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 application 61/736,527 filed on Dec. 12, 2012 and 61/748,427 filed on Feb. 2, 2013. Such genes, proteins and pathways may be the target polynucleotide of a CRISPR complex.

<table>
<thead>
<tr>
<th>TABLE A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DISEASE/DISORDERS</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Neoplasia</td>
</tr>
<tr>
<td>Age-related Macular Degeneration</td>
</tr>
<tr>
<td>Schizophrenia</td>
</tr>
<tr>
<td>Disorders</td>
</tr>
<tr>
<td>Trinucleotide Repeat Disorders</td>
</tr>
<tr>
<td>Fragile X Syndrome</td>
</tr>
<tr>
<td>Other Related Disorders</td>
</tr>
<tr>
<td>Prion-related disorders</td>
</tr>
<tr>
<td>ALS</td>
</tr>
<tr>
<td>Drug addiction</td>
</tr>
<tr>
<td>Autism</td>
</tr>
<tr>
<td>Alzheimer’s Disease</td>
</tr>
<tr>
<td>Inflammation</td>
</tr>
<tr>
<td>Parkinson’s Disease</td>
</tr>
<tr>
<td>Blood and coagulation diseases and disorders</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Cell dysregulation and oncology diseases and disorders</td>
</tr>
<tr>
<td>Inflammation and immune related diseases and disorders</td>
</tr>
<tr>
<td>Metabolic, liver, kidney and protein diseases and disorders</td>
</tr>
<tr>
<td>Muscular/Skeletal diseases and disorders</td>
</tr>
<tr>
<td>Neurological and neuronal diseases and disorders</td>
</tr>
</tbody>
</table>
### Table B-continued

1. GLO1, MEC2P, RIT, PPMX, MRX16, MRX79, NCLG3, NCLG4, KIAA1200, AUTS2X2; Fragile X Syndrome (FM2R, FXR1, FXR2, MGLUR5); Huntington’s disease and disease like disorders (HD, IT15, PRKN, PTP, PI3, JP3, HLD2, TBP, SCA17); Parkinson disease (NRA42, NURR1, NCOA2, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, D11, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK3, SNCA, NACP, PARK1, PARK4, PARK4, PRKN, PARK2, PD1, DBH, NDUFA2); Rett syndrome (MECP2, RIT, PPMX, MEX16, MRX79, CDKL5, STK9, MECP2, RIT, PPMX, MEX16, MRX79, x-Synuclein, Dj-1); Schizophrenia (Neuregulin 1 (Nrg1), ErbB (receptor for Neuregulin), Complexin1 (Cpx1), Tph1 Tyrptophan hydroxylase, Tph2, Tyrptophan hydroxylase 2, Neuroxin 1, GSK3, GSK3a, GSK3b, S-HT (Slc6a4), COMT; DRD5 (Drd5); DLG1, DLG3, DLP1, DTPB1, Duo (Duo1)); Secretase Related Disorders (APH1-1 (alpha and beta), Presenilin (Psen1), nicaestrin, (Nest), PEN2, Nasa1, Papi1, Nasa1, Nasa2); Trinucleotide Repeat Disorders (Htt (Huntington’s Dx), SBMA/SMX1/ARK (Kennedy’s Dx), FXN/Ne (Friedreich’s Ataxia), ATX3 (Machado- Joseph’s Dx), ATXN1 and ATXN2 (spinocerebellar ataxia), DMPK (myotonic dystrophy), Atrophin 1 and Ain1 (DRPLA Dx), CBP (Creb-BP - global instability), VLDLR (Alzheimer’s), Ato7, Atun10).

### Ocular diseases and disorders

Age-related macular degeneration (Aber, Cei2, Ce2, cp (centioporin), Tin6p, cathexin10, Vdfr, Ce2); Cataract (CryAA, CryA1, CryB2, CryB2, PitaX3, Bsp2, Sph1, CP49, CP74, CryA, CryAA, CryA1, PAX6, AN2, MchA, CryA1, CryB1, CryOC, CryG3, Cll, LIM2, Mki9, CryG3, CryG4, Bsp2, CP49, CP47, HSf4, CMT, HSf4, CMT, MIP, AQP9, CryAB, CryA2, CTP2, CryB2, CryG3, CryG4, CryG5, CryG6, Ccp3, Cae3, Cm31, Cm, Krtrial1); Corneal clouding and dystrophy (ApoA1, TGFB2, CD2, CDG61, CSD, BIGH3, CD2, TACE, TRO2, M1S1, VSN1, RINX, PPC2, PPD, KTCN, COL6A2, FCD, PPC2, PLP1, CDF); Cornea plana congenital (KERA, CAN2); Glaucoma (MYOC, TGF, GLC1A, JOAG, GPOA, OPTN, GLC2E, FIP2, Itip1, NR2F, CryB1, GLC3A, OPA1, DTG, NPG, CryB1, GLC3A); Leber congenital amaurosis (CRB1, Rho1, CROX, CROD2, CRD, PPOPI, LCA6, CORD9, RPE65, RP20, APL1, LCA4, GUCY2D, CUC2D, LCA1, CORD6, RDI12, LCA3); Macular dystrophy (ELOVL4, ADMD, STGD2, STGD3, RD8, RP7, PRP12, PRP11, AOMD, AOMD, VMD2).

### Table C

<table>
<thead>
<tr>
<th>Cellular Function</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI3K/AKT Signaling</td>
<td>PRKCE, ITGAM, ITG5, IRAK1, PRKAA2, EIF2AK2, PTK2, EIF4E, PTK2, GRK6, MAPK1, PLC1, AKT2, IKKb, PIK3CA, CDK8, CDK7, NFkB2, BCL2, PIK3CB, PIP2; MAPK1, BCL2L1, MAPK3, TSC2, ITGAT, KRAS, EIF4EBP1, RELA, PRKCD, NOX3; PKA2, MAP8, CD2, PPP2CA, PIM1, ITGB7, YWHAE, ILK, TP53, RAF1, IKKB, RELB, Dyrk1a, CDK5/1A, ITGB1, MAP2K2, JAK1, AKT1, IAK2, PIK3R1, CTJUK, PKP1, PPP2R5C, CTVN81, MAPK1, NFKB1, PKA3, ITGB3, CCND1, GSK3A, FRA1, SFN, ITGA2, TTK, CSNK1A1, BRAF, GSK3B, AKT3, FOXO1, SGK, IKB2; RPS5, KIDH</td>
</tr>
<tr>
<td>ERK/MAPK Signaling</td>
<td>PRKCE, ITGAM, ITG5, HSBB1, IRAK1, PRKAA2, EIF2AK2, RAC1, RAP1A, TNL1, EIF4E, ELK1, GRK6, MAPK1, RAC1, PLC1, AKT2, PIK3CA, CDK8, CREB1, PIK3C, PIP2; MAPK3, MAP3K, ITGA1, TSC2, KRAS, MYCN, EIF4EBP1, PPA1, PRKCD, PKA2, MAPK9, SRC, CDK2, PPP2CA, PIM1, PIK3CA, ITGB7, YWHAE, PPP1CC, KSR, FKN, RAF1, FYN, Dyrk1a, ITGB1, MAPK2, P4K, P4K1, STAT3, PPP2R5C, MAPK2, PKA3, ITGB3, ESR1, ITGA2, MYC, TTK, CSNK1A1, CRKL, BRAF, ATF4, PIK3CA, SRC1, STAT1, EGK</td>
</tr>
<tr>
<td>Glucocorticoid Receptor Signaling</td>
<td>RAC1, TAF4, EP300, SMAD2, TRAF6, PCAF, ELK1, MAPK1, SMAD3, AKT2, IKKB, NCO2, RUB1, PIK3CA, CREB1, FOS, HSBB5, NFkB2, BCL2, MAPK14, STAT3B, PIK3C, PIK3C, MAPK8, BCL2L1, MAPK3, TSC2, MAP1K, NR1I1, KRAS, MAPK13, RELA, STAT5A, MAPK9, NOS2A, PKX1, NR3C1, PIK3CA, CDK11, TRAF2, SERPINE1, NCOA3, MAPK4, TNF, RAF1, IKKB, MAPK7, CREBBP, CDK11A, MAP2K2, JAK1, IL8, NCOA2, AKT1, JAK2</td>
</tr>
<tr>
<td>CELLULAR FUNCTION</td>
<td>GENES</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Axonal Guidance Signaling</td>
<td>PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ERK1; SMAD4; CEBPβ; JUN; AR; AKT3; CCL2; MMP1; STAT1; IL6; EISP20AA1; PRKCE; ITGA8; ROCK1; ITGAX; CXCX4; ADAM1; IGF1; RAC1; RAF1; AIF4; PKRC2; NR1P1; NTRK2; ARHGEF7; SOS1; ROCK2; MAPK1; PGE; RAC2; PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCE; PTK2; CFL1; GNAQ; PIK3CB; CXCL2; PIK3C3; WNT11; PRKD1; GNB2L1; ABL1; MAPK3; ITGAL; KRAS; RHOA; PRKCD; PIK3C2A; ITGAF; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1; GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA</td>
</tr>
<tr>
<td>Ephrin Receptor Signaling</td>
<td>PRKCE; ITGA2; ROCK1; ITGA5; CXC4; IRAK1; PRKAA2; EIF2AK2; RAC1; RAF1; AIF4; ERK6; GRK6; ROCK2; MAPK1; RAC2; PIK3C2; CDK8; PT2; CFL1; PIK3CB; MYH9; DIA1F1; PIK3C3; MAPK8; F2K; MAP3K3; SLCA1; ITGAL; KRAS; RHOA; PRKCD; PRKAA1; MAPK8; CDK2; PIM1; ITGAB; PXN; RAF1; FYN; DRYK1A; ITGB1; MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; TTK; CSNK1A1; CRKL; BRF1; PTPN13; AT14; AKT1; SGK</td>
</tr>
<tr>
<td>Actin Cytoskeleton Signaling</td>
<td>ACTN4; PRKCE; ITGAM; ROCK1; ITGAX; IRAK1; PRKAA2; EIF2AK2; RAC1; IRS1; ARHGEF7; ERK6; ROCK2; MAPK1; RAC2; PIK3C2; CDK8; PT2; CFL1; PIK3CB; MYH9; DIA1F1; PIK3C3; MAPK8; F2K; MAP3K3; SLCA1; ITGAL; KRAS; RHOA; PRKCD; PRKAA1; MAPK8; CDK2; PIM1; PIK3CA; ITGB1; PPP1CC; PXN; VHL; RAF1; GSN; DRYK1A; ITGB1; MAP2K2; PAK4; ITK5XL1; PIK3R1; MAP2K1; PAK3; ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRF1; VAV3; SGK</td>
</tr>
<tr>
<td>Huntington’s Disease Signaling</td>
<td>PRKCE; IGF1; EP300; ROC1; PRKCE; HDAC4; TMC2; MAPK1; CAPNS1; AKT2; DORF; NCOA2; SP1; CAPN2; PIK3CA; HDAC5; CREB1; PIK3C1; EISP20A5; REST; GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3; CAPS8; HDAC2; HDAC7A; PIK3CB; HDAC1; MAP5; HDAC9; PIK3CA; HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1; PDKP1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX; AT14; AKT3; PRKCA; C10G10G1; SGK; HDAC6; CAPS3</td>
</tr>
<tr>
<td>Apoptosis Signaling</td>
<td>PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1; BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NEFB2; BCL2; MAP3K4; MAPK8; BCL2L1; CAPN1; MAPK3; CAPS8; KRAS; RELA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNN; RAF1; IKBKG; RELB; CASP9; DRYK1A; MAP2K2; CHUK; APAF1; MAP2K1; NEK1B; PAK3; LMAA; CAPS2; BIRC3; TTK; CSNK1A1; BRF1; BAX; PRKCA; SGK; CASP3; BIRC3; PAPAR</td>
</tr>
<tr>
<td>B Cell Receptor Signaling</td>
<td>RAC1; PTEN; LYN; ELK1; MAPK1; ROCK2; PTPN11; AKT2; IKBKB; PIK3CA; CREB1; SYK; NEFB2; CAMK2A; MAPK14; RIK3B; PIK3C; MAPK8; BCL2L1; ABL1; MAPK3; ELK1; MAPK13; RELA; PTEN; PIK3C2; EGFR1; PIK3CA; BTK; MAPK14; RAF1; IKBKG; RELB; MAPK7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NEK8B1; CDC42; GSK3A; FRAF; BCL6; BCL10; JUN; GSK3B; AT14; AKT3; VANT3; RPS6KB1</td>
</tr>
<tr>
<td>Leukocyte Extravasation Signaling</td>
<td>ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXC4; CYBA; RAC1; RAP1A; PRKCE; ROCK2; RAC2; PTPN11; MMP14; PIK3CA; PRKX1; PT2; PIK3CB; CXCL12; PIK3C3; MAPK8; PRKX1; ABL1; MAPK12; PIK3R3; HDAC3; RHOA; PIK3D; PIK3K; SRC; PIK3CA; BTK; MAPK14; NOX1; PXN; VHL; VASP; ITGAB1; MAP2K2; CTNNB1; PIK3R1; CTNNB1; C1DN1; CDC42; FBL1; ITK; CRKL; VAV3; CTNN; PRKCA; MMP1; MMP9</td>
</tr>
<tr>
<td>Integrin Signaling</td>
<td>ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A; TLN1; ARIHGEF7; MAPK1; ROCK2; CAPN1; AKT2; PIK3C3; PT2; PIK3CB; PIK3C3; MAPK8; CASV1; CAPN1; ABL1; MAPK3; ITGAX; KRAS; RHOA; SRC; PIK3CB1; ITGAF; PPI1CC; ILK; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1; TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL; BRF1; GSK3B; AKT3</td>
</tr>
<tr>
<td>CELLULAR FUNCTION</td>
<td>GENES</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Acute Phase Response Signaling</td>
<td>IRAK1, SOD2, MYD88, TRAF6, ELK1, MAPK1, PTPN11, AKT2, JNK, PI3KCA, FOS, NFkB1, MAPK14, PI3KC3B, MAPK8, RIIK1, MAPK3, IL6ST, KRAS, MAPK13, IL6R, RELA, SOCS1, MAPK5, FTI, NRRC1, TRAF2, SERPINE1, MAPK14, TNF, RAF1, PDK1, IKBKG, RELB, MAPK37, MAPK22, AKT1, JAK2, PIK3R1, CHUK, STAT3, MAP2K1, NFKB1, FRAP1, CEBPB, JUN, AKT3, ILIR1, IL6</td>
</tr>
<tr>
<td>PTEN Signaling</td>
<td>ITGAM, ITGAV, RAC1, PTEN, PIKKZ, BCL2L11, MAPK3, RAC2, AKT2, EFR, IKBKB, CBL, PI3KCA, CDKN1B, PTK2, NFkB2, BCL2, PI3KC3B, BCL2L11, MAPK3, ITGAL, KRAS, ITGB7, ILK, PDGFRB, INS, RAF1, IKBKG, CASP9, CDKN1A, ITGB1, MAP2K2, AKT1, PIK3R1, CHUK, PDGFRB, PDK1, MAPK13, NFKB1, ITGB3, CDC42, CCND1, OSK3A, ITGA2, GSK3B, AKT3, FOXO1, CASP3, RPS6KB1</td>
</tr>
<tr>
<td>p53 Signaling</td>
<td>PTEN, EP300, BRC3, P53AIP1, P53BP1, BRC4, AKT2, PI3KCA, CHEK1, TP53INPI, BCL2, PI3KC3B, PI3KC3A, MAPK8, THBS1, ATR, BCL2L1, E2F1, PMAPK11, CHEK2, TNFRSF10B, TP53, RB1, HDAO9, CD2K, PI3KC2A, MAPK14, TSP53, LRRD3, CDKN1A, IKBK2, AKT1, PIK3R1, RRM2B, APAF1, CTNNB1, SIRT1, CCND1, PRKDC, ATM, SFN, CDKN2A, JUN, SNA2I, GSK3B, BAX, AKT3</td>
</tr>
<tr>
<td>Acyl Hydrocarbon Receptor Signaling</td>
<td>HSPP1, EP300, FASN, TGM2, RXRA, MAPK1, NQO1, NCOA2, SPI1, ARNT, CDKN1B, FOS, CHEK1, SMARC4A, NFKB2, ALDH1A1, ATR, E2F1, MAPK3, NRP1, CHEK2, RELA, TP53, GSTP1, RB1, SRC, CD2K, AKR, NF2L2, NCOA3, TSP53, TNF, CDKN1A, NCOA2, APAF1, NFKB1, CCND1, ATM, ESR1, CDKN2A, MYC, JUN, ESR2, BAX, IL6, CYP1B1, HSPP1</td>
</tr>
<tr>
<td>Xenobiotic Metabolism Signaling</td>
<td>PRKCE, EP300, PRKCE, RXRA, MAPK1, NQO1, NCOA2, PI3KC3A, ARNT, PRKCE, NFkB2, CAMK2A, PI3KC3B, PPP2R1A, PI3KC3A, MAPK6, PI3KC3A, ALDH1A1, MAPK3, NRP1, KRAS, MAPK13, PRKCD, GSTP1, MAPK9, NOS2A, AIBC1, AKR, PP2CA, FTI, NFE2L2, PI3KC2A, PPP1GCA1A, MAPK14, RAF1, CREBBP, MAP2K2, PIK3R1, PPP2RC5, MAPK21, NFKB1, KEAP1, PRKCA, E2F2, A3, IL6, CYP1B1, HSPP1</td>
</tr>
<tr>
<td>SAPK/IKK Signaling</td>
<td>PRKCE, IRAK1, PRKAA2, E2F2, AK2, RAC1, ELK1, GRK6, MAPK1, GADD45A, RAC2, PLK1, AKT2, PI3KCA, FADD, CDK8, PI3KC3B, PI3KC3A, MAPK8, RIIK1, GNB2L1, IRS1, MAPK3, MAPK10, JAXX, KRAS, PRKCD, PRKAA1, MAPK9, CDK2, PIM1, PI3KC3A, TRAf2, TSP53, LCK, MAPK7, Dyrk1a, MAP2K2, PIK3R1, MAP2K1, PAK3, CDC42, JUN, TTK, CSNK1A1, CRL1, B Raf, SOK</td>
</tr>
<tr>
<td>PPAR/RXR Signaling</td>
<td>PRKAA2, EP300, INS, SMAD2, TRAF6, PPARa, FASN, RXRA, MAPK1, SMAD3, GNAS, IKBKB, NCO2, AHCAP1, GNAQ, NFKB2, MAPK14, STAT3B, MAPK8, IRS1, MAPK3, KRAS, RELA, PIK3R1, PPARGCA1A, NCOA3, MAPK14, INS, RAF1, IKBKG, RELB, MAPK7, CREBBP, MAP2K2, JAK2, CHUK, MAP2K1, NFKB1, TGFBR1, SMAD4, JUN, ILIR1, PRKCA, IL6, HSPP1AA1, A3DPOQ</td>
</tr>
<tr>
<td>NF-κB Signaling</td>
<td>IRAK1, E2F2, AK2, EP300, INS, MYD88, PRKCE, TRAF6, TBK1, AKT2, EGRF, IKBKB, PI3KCA, BTRC, NFkB2, MAPK14, PI3KC3B, PI3KC3A, MAPK8, RIIK1, HADAC2, KRAS, RELA, PI3KC3A, TRAF2, TLR4, PDGFRB, TNF, INS, LCK, IKBKG, RELB, MAPK7, CREBBP, AKT1, PIK3R1, CHUK, PDGFRB, NFKB1, TL2, BCL6, OSK3B, AKT3, TNFAIP3, ILIR1</td>
</tr>
<tr>
<td>Neuregulin Signaling</td>
<td>ERBB4, PI3KC3A, ITGAM, ITGAV, PTEN, PRKCE, ELK1, MAPK1, PTPN11, AKT2, EGRF, IKBKB, PI3KCA, BTRC, NFkB2, MAPK14, PI3KC3B, PI3KC3A, MAPK8, RIIK1, HADAC2, KRAS, RELA, PI3KC3A, TRAF2, TLR4, PDGFRB, TNF, INS, LCK, IKBKG, RELB, MAPK7, CREBBP, AKT1, PIK3R1, PDK1, MAP2K1, TGFBR3, EREG, FRA1, B3EN1, ITGAV, MYC, NRG1, CRKL, AKT3, PRKCA, HSPP1AA1, PPS6KB1</td>
</tr>
<tr>
<td>Wnt &amp; Beta catenin Signaling</td>
<td>C4D4, EP300, LRIP, DVL3, C3NSK1E2, GJA1, SMO, AKT2, P21, CDH1, BTRC, GNAQ, MARK2, PI3KC3A, WNT11, SRC, DKK1, P2P2CA, SOX6, SFRP2, ILK, LEF1, SOX9, TSP5, MAPK7, CREBBP, TCF7L2, AKT1, PI3KC3A, WNT5A, LRPS, CTNNB1, TGFBR1, CCND1</td>
</tr>
<tr>
<td>GENES</td>
<td>CELLULAR FUNCTION</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
</tr>
<tr>
<td>GSX3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOD2</td>
<td>Insulin Receptor Signaling</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>HSP90A1</td>
<td>IL-6 Signaling</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>CELLULAR FUNCTION</td>
<td>GENES</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Cell Cycle: G1/S</td>
<td>HDAC4, SMAD3, SUV39H1, HDAC5, CKDN1B, BTRC, ATR, ABU1, EZF1, HDAC2, HDAC7A, RB1, HDAC11, HDAC9, CDK2, EZF1, HDAC3, TP53, CDKN1A, CCND1, EZF4, ATM, RBL2, SMAD4, CDKN2A, MYC, NRG1, GSK3B, RBL1, HDAC6</td>
</tr>
<tr>
<td>Checkpoint Regulation</td>
<td>RAC1, ELK1, MAPK1, JNK1, BCL6, PI3KCA, FOXO, NFkB2, PI3KCB, PI3KC3, MAPK8, MAPK1, KRAS, RELA, PI3KC3A, BTK, JNK1, RAF1, IKK, RELB, FYN, MAP2K2, PI3K3, CHUK, MAP2K1, NFkB1, IKK, BCL10, JUN, JUNB, JUN</td>
</tr>
<tr>
<td>T cell Receptor Signaling</td>
<td>CRADD, HSPSB1, BID, BIRC4, TBK1, IKK, KAP1, FADD, FAS, NFkB2, BCL2, MAPK1, MAPK8, RIPK1, CASP9, DAXX, TNFRSF10B, RELA, TRAF2, TNF, IKK, RELB, CASP9, CHUK, APAF1, NFkB1, CASP8, BIRC2, CASP3, BIRC3</td>
</tr>
<tr>
<td>FGF Signaling</td>
<td>RAC1, FGFR1, MET, MAPKAPK2, MAPK1, PTEN, AKT2, PI3KCA, CREB1, PI3KC3, MAPK8, MAP3K1, PTEN, PI3KC3A, MAPK14, RAF1, AKT1, PI3K3, STAT3, MAPK1, FGFR4, CRKL, ATM, AKT3, PRKCA, HGF</td>
</tr>
<tr>
<td>GM-CSF Signaling</td>
<td>LYN, ELK1, MAPK1, PTEN, AKT2, PI3KCA, CAMK2A, STAT3, PI3KCB, PI3KC3, GNB2L1, BCL2L1, MAPK3, ETS1, KRAS, RUNX1, PIM1, PI3KCA, RAF1, MAP2K2, AKT1, JAK2, PI3K1, STAT3, MAPK2K1, CCND1, AKT3, STAT5</td>
</tr>
<tr>
<td>Amyotrophic Lateral Sclerosis Signaling</td>
<td>BID, JGF1, RAC1, BIRC4, PGF, CAPN1, CAPN2, PI3KCA, BCL2, PI3KC3, PI3KC3A, TP53, ASPP, PI3K1, RB1, CASP9, APAF1, VEGFA, BIRC2, BAX, AKT3, CASP9, BIRC3</td>
</tr>
<tr>
<td>JAK-Stat Signaling</td>
<td>PTEN, MAPK1, PTEN, AKT2, PI3KCA, STAT3, PI3K, PI3KC3, PI3KC3A, GNB2L1, BCL2L1, MAPK3, ETS1, KRAS, RUNX1, PIM1, PI3KCA, RAF1, MAP2K2, AKT1, JAK2, PI3K1, STAT3, MAPK2K1, FRA1, AKT3, STAT5</td>
</tr>
<tr>
<td>Nicotinate and Nicotinamide Metabolism</td>
<td>PRKCE, IRAK1, PRKAA2, EIF2AK2, GRK6, MAPK1, PLK1, AKT2, CDK8, MAPK8, MAPK3, PRKAA1, P38, MAPK9, CDK2, PI3K, DYRK1A, MAP2K2, MAPK1, PI3K, NT5E, TTK, CSNK1A1, B Raf, SGK</td>
</tr>
<tr>
<td>Chemokine Signaling</td>
<td>CXCR4, ROCK2, MAPK1, PTK2, FOS, CFL1, GRNAQ, CAMK2A, CXCL12, MAPK9, MAPK3, KRAS, MAPK3, RHOA, CCR3, SRC, PP1, MAPK14, NOX1, RAF1, MAP2K2, MAPK1, JUN, CCL2, PRKCA</td>
</tr>
<tr>
<td>IL-2 Signaling</td>
<td>ELK1, MAPK1, PTEN, AKT2, PI3KCA, SYK, FOS, STAT5, PI3KCB, PI3K3, MAPK8, MAPK3, KRAS, SOCS1, STAT5A, PI3KCA, LCK, RAF1, MAP2K2, JAK1, AKT1, PI3K1, MAP2K1, JUN, AKT3</td>
</tr>
<tr>
<td>Synaptic Long Term Depression</td>
<td>PRKCE, IGF1, PRKCZ, PRDX6, LYN, MAPK1, GRNAQ, PRKCE, GRNAQ, PPP2R5C, MAPK1, PRKCA</td>
</tr>
<tr>
<td>Estrogen Receptor Signaling</td>
<td>TAF4B, EP300, CARM1, PCAF, MAPK1, NCOA2, SMARCA6, MAPK3, NRP1, KRAS, SRC, NCOA1, HDAC3, PPARG, RBM9, NCOA3, RAF1, CREBBP, MAP2K2, NCOA2, MAPK1, PRKDC, ESR1, ESR2</td>
</tr>
<tr>
<td>Protein Ubiquitination Pathway</td>
<td>TRAF6, SMURF1, BIRC4, BIRC4, UCHL1, NEDD4, CBL, UBE2J1, BTRC, HSPA5, USP7, USP10, FBXW7, USP9X, STUB1, UBE2, IKBKCD, BIRC2, PAR2, USP8, USP1, HSP90AA1, BIRC3</td>
</tr>
<tr>
<td>IL-10 Signaling</td>
<td>TRAF6, CCR1, ELK1, IKK, SP1, FOS, NFkB2, MAPK14, MAPK8, MAPK13, RELA, MAPK14, TNF, IKK, RELB, MAPK7, JAK1, CHUK, STAT3, NFKB1, JUN, JUNB, JUN</td>
</tr>
<tr>
<td>VDR/RR Activation</td>
<td>PRKCE, EP300, PRKCB, RXRA, GADD45A, IKB1, NCOA2, SPI1, PI3K, CDKN1B, PRKD1, PRKCD, RUNX2, KL4, YY1, NCOA3, CDKN1A, NCOA2, SPP1, LRPS, CREBBP, FOXO1, PRKCA</td>
</tr>
<tr>
<td>TGF-beta Signaling</td>
<td>EP300, SMAD2, SMURF1, MAPK1, SMAD5, SMAD1, FO, MAPK8, MAPK3, KRAS, MAPK9, RUNX2, SERINE, RAF1, MAPK7, CREBBP, MAP2K2, MAP2K1, TGFBR1, NCOA4, JUN, NCOA5</td>
</tr>
<tr>
<td>Toll-like Receptor Signaling</td>
<td>IRAK1, EIF2AK2, MYD88, TRAF6, PPARA, ELK1, IKK, FOS, NFkB2, MAPK14, MAPK8, MAPK13, RELA, TLR4, MAPK1, IKK, RELB, MAPK7, CHUK, NFKB1, TLR2, JUN</td>
</tr>
<tr>
<td>p38 MAPK Signaling</td>
<td>HS3BP1, IRAK1, TRAF6, MAPKAPK2, ELK1, FADD, FAS, CREB1, DODT3, RPS6KA4, DAXX, MAPK13, TRAF2</td>
</tr>
<tr>
<td>Cellular Function</td>
<td>Genes</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Neurotrophin/Trk Signaling</td>
<td>MAPK14, TNF, TFAMK17, TGFBR1, MYC, AIF4, IL1R1, SRF, STAT1</td>
</tr>
<tr>
<td>Fxr/Rxr Activation</td>
<td>TRK2, MAP1, PTPN11, PIK3CA, CREB1, FOS, PIK3CB, PIK3C3, MAPK8, MAPK3, KRAS, PIK3CA2, RAF1, MAP2K2, AKT1, PIK3R1, PDK1, MAP2K1</td>
</tr>
<tr>
<td>Synaptic Long Term Potentiation</td>
<td>CDC42, JUN, ATF4</td>
</tr>
<tr>
<td>Calcium Signaling</td>
<td>INS, PPARA, FASN, RXRA, AKT2, SDC1, MAPK8, APOB, MAPK10, PPARG, MTTP, MAPK9, PPARG2C1A, TNE, CREBBP, AKT1, SREBF1, FOXM1, AKT3, FOXO1</td>
</tr>
<tr>
<td>EGFR Signaling</td>
<td>RA2A, EP300, HDAC4, MAPK1, HDAC5, CREB1, CAMK2A, MYH9, MAPK3, HDAC2, HDAC7A, HDAC11, HDAC9, HDAC3, CREBBP, CALR, CAMKK2, ATF4, HDAC6</td>
</tr>
<tr>
<td>Hypoxia Signaling in the Cardiovascular System</td>
<td>ELK1, MAPK1, EGFR, PIK3CA, FOS, PIK3CB, PIK3C3, MAPK8, MAPK3, PIK3CA2, RAF1, JAK1, PIK3R1, STAT3, MAPK1, JUN, PRL1, CREB1, SRF, STAT1</td>
</tr>
<tr>
<td>IL-6/11/L-1 Mediated Inhibition of Xcr1 Signaling</td>
<td>IRAK1, MYD88, TRAF6, PPARA, RXRA, ABCA1</td>
</tr>
<tr>
<td>LXR/Rxr Activation</td>
<td>MAPK6, ALDH1A1, GSK3B, MAPK9, ABC1, TRAF2, TLR4, TNF, MAPK3, NR1H2, SREBF1, JUN, IL1K</td>
</tr>
<tr>
<td>Ankylosing Processing</td>
<td>FASN, RXRA, NOCS1, ABCA1, NFIB, IRE1, RELA, NOS2A, TLR4, TNF, RELB, LDR, NR1H2, NFKB1, SREBF1, IL1R1, CCL2, IL6, MMP9</td>
</tr>
<tr>
<td>IL-4 Signaling</td>
<td>PIK3C1, CSNK1E, MAPK1, CAPN811, AKT2, CAPN2, CAPN1, MAPK3, MAPK13, MAPT, MAPK14, AKT1, PSEN1, CSN1K1, GSK3B, AKT3, APP</td>
</tr>
<tr>
<td>Cell Cycle, G2/M DNA Damage Checkpoint</td>
<td>PI3K, PAF, BRCA1, GADD45A, PLK1, BTRC</td>
</tr>
<tr>
<td>Nitric Oxide Signaling in the Cardiovascular System</td>
<td>VEGFA, ATM, SPT fantastic, Cdk2</td>
</tr>
<tr>
<td>Purine Metabolism</td>
<td>NME1, SMAC4A, MYD9, RRM2, ADAR, EIF2AK4, P53M2, ENTPD1, RAD51, RRM2B, TF2, RAD51C, NT5E, POLD1, NME1</td>
</tr>
<tr>
<td>cAMP-Mediated Signaling</td>
<td>RAP1A, MAP1, GNAS, CREB1, CAMK2A, MAPK3, SRC, RAF1, MAP2K2, STAT3, MAP2K1, BRAF, ATF4</td>
</tr>
<tr>
<td>Mitochondrial Dysfunction</td>
<td>SOX2, MAPK8, CASP8, MAPK10, MAPK9, CASP9</td>
</tr>
<tr>
<td>Notch Signaling</td>
<td>HES1, JAG1, NUMB, NOTCH1, ADAM17, NOTCH2, PSEN1, NOTCH3, NOTCH1, Dll4</td>
</tr>
<tr>
<td>Endoplasmic Reticulum Stress Pathway</td>
<td>ITP2A, MAPK8, XBP1, TRAF2, ATF6, CASP9, AT4F, EIF2AK3, CASP3</td>
</tr>
<tr>
<td>Parkinson's Signaling</td>
<td>UC111, MAPK8, MAPK13, MAPK14, CASP9, PARK7, PARK2, CASP3</td>
</tr>
<tr>
<td>Adrenergic Signaling</td>
<td>GNAS, GNAQ, PPP2R1A, GNB2L1, PPP2CA, PPP1CC</td>
</tr>
<tr>
<td>Glycogenolysis</td>
<td>HK2, GC, GF, ALDH1A1, PKM2, LDLA, HIK</td>
</tr>
<tr>
<td>Interferon Signaling</td>
<td>IRF1, SOCS1, JAK1, JAK2, IFITM1, STAT1, IFIT3</td>
</tr>
<tr>
<td>Sonic Hedgehog Signaling</td>
<td>ARRB2, SMO, GLI2, Dyrk1A, GLI1, GSK3B, Dyrk1B</td>
</tr>
<tr>
<td>Glycophosphatidyl Inositol Metabolism</td>
<td>PLD1, GRN, GPA4, YWHAZ, SPHK1, SPHK2</td>
</tr>
<tr>
<td>Phospholipid Degradation</td>
<td>PRDX6, PLD1, GRN, YWHAZ, SPHK1, SPHK2</td>
</tr>
<tr>
<td>Tryptophan Metabolism</td>
<td>SIAH2, PRMT5, NEDD4, ALDH1A1, CYP1B1, SIAH1</td>
</tr>
<tr>
<td>Lysine Degradation</td>
<td>SVE3H1, EHM2T, N2D1, SETD7, PPP2R3C</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>UC111, HK2, GCK, GP1, HIK</td>
</tr>
<tr>
<td>Aminoglycoside Metabolism</td>
<td>NQO1, HK2, GCK, HIK</td>
</tr>
<tr>
<td>CELLULAR FUNCTION</td>
<td>GENES</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Arachidonic Acid Metabolism</td>
<td>PRDX6; GRN; YWHAZ; CYP1B1</td>
</tr>
<tr>
<td>Circadian Rhythm Signaling</td>
<td>CSNK1E; CREB1; ATF4; NR1D1</td>
</tr>
<tr>
<td>Coagulation System</td>
<td>BDKRB1; F2R; SERPINE1; 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; SPHK1; SPHK2</td>
</tr>
<tr>
<td>Linoleic Acid Metabolism</td>
<td>PRDX6; GRN; YWHAZ; CYP1B1</td>
</tr>
<tr>
<td>Methionine Metabolism</td>
<td>DNM1T1; DNM3T3b; AHCY; DNM3T3a</td>
</tr>
<tr>
<td>Pyruvate Metabolism</td>
<td>GLO1; ALDH1A1; PRK2; LDHA</td>
</tr>
<tr>
<td>Arginine and Proline Metabolism</td>
<td>ALDH1A1; NOS3; NOS2A</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>Stilbene, Coumarine 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>NQO1; DHCR7</td>
</tr>
<tr>
<td>Butanoate Metabolism</td>
<td>ALDH1A1; NLGN1</td>
</tr>
<tr>
<td>Citrate Cycle</td>
<td>IDH2; IDH1</td>
</tr>
<tr>
<td>Fatty Acid Metabolism</td>
<td>ALDH1A1; CYP1B1</td>
</tr>
<tr>
<td>Glycosphingolipid Metabolism</td>
<td>PRDX6; CHKA</td>
</tr>
<tr>
<td>Histidine Metabolism</td>
<td>PRMT5; ALDH1A1</td>
</tr>
<tr>
<td>Inositol Metabolism</td>
<td>ERO1L; 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>PRMT5; AHCA</td>
</tr>
<tr>
<td>Sphingolipid Metabolism</td>
<td>SPHK1; SPHK2</td>
</tr>
<tr>
<td>Aminophosphonate Metabolism</td>
<td>PRMT5</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>FASN</td>
</tr>
<tr>
<td>Glutamate Receptor Signaling</td>
<td>GNB2L1</td>
</tr>
<tr>
<td>NRF2-mediated Oxidative Stress Response Pathway</td>
<td>PRDX1</td>
</tr>
<tr>
<td>Pentose Phosphate Pathway</td>
<td>GPI</td>
</tr>
<tr>
<td>Pentose and Glucuronate 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>Ubiquitine 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>
<tr>
<td>Pain Tolerance</td>
<td>TRPM5; TRPA1</td>
</tr>
<tr>
<td>Pain</td>
<td>TRPM7; TRPC5; TRPC6; TRPC1; Car1; car2; Grk2; Trpa1; Pnoc; Cgrp; Crf; Pnla; Ntr2b; TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a</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. McIvor E L, Polak U, Napierala 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 EMP2A and EMP2B genes that have been identified to be associated with Lafora disease. Lafora disease is an autosomal recessive condition which is characterized by progressive myoclonic 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 Paeudiatic Neurology; 20 (2009).

In yet another aspect of the invention, 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 Trivalent Repeat Disorders, Leigh’s Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Collpcephaly. 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 Truncinotid 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, PINK1, 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 Igf receptor IIB (FCGR2B, also termed CD32) encoded by the FCGR2B gene, or the Fc epsilon R1g (FCER1G) protein encoded by the FcER1G gene, for example.

Examples of cardiovascular diseases associated proteins may include IL1B (interleukin 1 beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin I2 (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 (UBEL1C) encoded by the UBE1C 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 protein (AFF2) 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 (ABC1) 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) Ligand 2 protein (CCL2) encoded by the CCL2 gene, for example.

Examples of proteins associated Schizophrenia may include NRG1, ErbB4, CRLF1, TP1, TP2, NRXN1, GSK3α, BDNF, DISC1, GSK3β, 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 (v-erb-b2 erythroblast leukemia viral oncogene homolog 2), ERBB3 (v-erb-b2 erythroblast leukemia viral oncogene homolog 3), ERBB4 (v-erb-b2 erythroblast leukemia viral oncogene homolog 4), Notch 1, Notch2, Notch 3, or Notch 4, for example.

Examples of proteins associated with a secretory disorder may include PSENEN (presenilin enhancer 2 homolog C (C. elegans)), CTSC (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective 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), VAFGA (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 associated with prion diseases may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAFGA (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 (Alphas-2-Macroglobulin), AAFT (Apopotosis antagonizing transcription factor), ACPP (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 adrenoreceptor), for example.

Examples of proteins associated with Immunodeficiency may include A2M (Alphas-2-Macroglobulin); AANAT [arylalkylamine N-acetyltransferase]; ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1]; ABCA2 [ATP-binding cassette, sub-family A (ABC1), member 2]; or ABCA3 [ATP-binding cassette, sub-family A (ABC1), 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 (atxin 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 HTRC (5-hydroxytryptamine (serotonin) receptor 2C), for example.

Examples of neurodevelopmental-associated sequences include A2BP1 [Ataxin-2 binding protein 1]; AADAT [aminooxidase aminotransferase], AANAT [arylalkylamine N-acetyltransferase], ABAT [4-aminobutyrate aminotransferase], ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1], or ABCA13 [ATP-binding cassette, sub-family A (ABC1), member 13], for example.

Further examples of preferred conditions treatable with the present system include may be selected from: Aicardi-Goutières Syndrome; Alexander Disease; Allan-Herndon-Dudley Syndrome; PolQ-Related Disorders; Alpha-Mannosidosis (Type II and III); Alström Syndrome; Angelman Syndrome; Ataxia-Telangiectasia; Neuronal Ceroid-Lipofuscinoses; Beta-Thalassemia; Bilateral Optic Atrophy and (Infantile) Optic Atrophy Type 1; Retinoblastoma (bilateral); Canavan Disease; Cerebrooculofacioskeletal Syndrome 1 (COSF1); Cerebrotendinous Xanthomatosis; Cornelia de Lange Syndrome; MAPT-Related Disorders; Genetic Prion Diseases; Dravet Syndrome; Early-Onset Familial Alzheimer Disease; Friedreich Ataxia [FRDA]; Fröhlich Syndrome; Puecosidosis; Fukuyama Congenital Muscular Dystrophy; Galactosialidosis; Gaucher Disease; Organic Acidemias; Hemophagocytic Lymphohistiocytosis; Hutchinson-Gilford Progeria Syndrome; Mucoepidermoid II; Infantile Free Sialic Acid Storage Disease; PLA2G6-Associated Neurodegeneration; Jervell and Lange-Nielsen Syndrome; Junctional Epidermolysis Bullosa; Huntington Disease; Knobbe Disease (Infantile); Mitochondrial DNA-Associated Leigh Syndrome and NARP; Lesch-Nyhan Syndrome; LIS1-Associated Lisenscephaly; 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; PLP1-Related Disorders; Perry Syndrome; Phelan-McDermid Syndrome; Glycogen Storage Disease Type II (Pompe Disease) (Infantile); MAPT-Related Disorders; MECP2-Related Disorders; Rhizomelic Chondrodysplasia Punctata Type 1; Roberts Syndrome; Sandhoff Disease; Schindler Disease—Type 1; Adenosine Deaminase Deficiency; Smith-Lemi-Opitz Syndrome; Spinal Muscular Atrophy; Infantile-Onset Spinoocerebellar 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 polynucleotide 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 therein 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 1

CRISPR Complex Activity in the Nucleus of a Eukaryotic Cell

An example type II CRISPR system is the type II CRISPR loci from Streptococcus pyogenes SF370, which contains a
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 server 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 136(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 CRISP1 locus sufficient for CRISP activity were reconstituted in E. coli using pCrispr plasmid (schematically illustrated in Fig. 10A). pCrispr contained ttrRNA, SpCas9, and a leader sequence driving the crRNA array. Spacers (also referred to as “guide sequences”) were inserted into the crRNA array between BsaI sites using annealed oligonucleotides, as illustrated. Challenges 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 pCrispr 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 trypsinization 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 Naondrop (Thermo Scientific) and normalized to same concentration.

Northern Blot Analysis of crRNA and ttrRNA Expression in Mammalian Cells

RNAs were mixed with equal volumes of 2x loading buffer (Ambion), heated to 95°C for 5 min, chilled on ice for 1 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 limit. 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 Stratalink (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 [gamma-³²P]ATP (Perkin Elmer) with T4 poly-
ucleotide kinase (New England Biolabs). The membrane
was washed once with pre-warmed (42° C.) 2×SSC, 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 phorimager (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 Gib-
son Assembly. Two **B**sal type IIS sites were introduced in
between two direct repeats to facilitate easy insertion of spec-
cers (FIG. 9). PCR products were cloned into EcoRV-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 Csn2. Oligos (Integrated DNA Technology) encoding
spacers with complementary overhangs were cloned into the
**B**sal-digested vector pDC3000 (NEB) and then ligated with
17 ligase (Enzymatic) to generate pCRISPR plasmids.
Challenger 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-dig-
gested 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 manufacturer’s
instructions. In the transformation assay, 50 ul aliquots of
competent cells carrying pCRISPR were thawed on ice and
transformed with 1 ng of spacer plasmid or pUC19 on ice for
30 minutes, followed by 45 second heat shock at 42° C. and 2
minutes on ice. Subsequently, 250 ul SOC (Invitrogen) was
added followed by shaking incubation at 37° C. for 1 hr, and
100 ul of the post-SOC outgrowth was plated onto double
selection plates (12.5 mg/ml chloramphenicol, 100 mg/ml
ampicillin). To obtain cfi/ng of DNA, total colony numbers
were multiplied by 3.

To improve expression of CRISPR components in mamma-
lian cells, two genes from the SF370 locus I of *Strepto-
coccus pyogenes* (*S. pyogenes*) were codon-optimized, Cas9
(SpCas9) and RNase III (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
case, the C-terminal NLS was that of nucleoplasmin
(KRPAATKKGQAKKKK (SEQ ID NO: 2)), and the C-ter-
minal NLS was that of the SV40 large-T antigen (PKKKRRKV
(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 subse-
quently processed into identical 75 nt mature tracrRNAs. The
shorter 89 nt tracrRNA was selected for expression in mamma-
lian 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 tracr-
RNA are detected on the Northern blot.

To promote precise transcriptional 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
(39mers, 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 mamma-
lian chromosones, HEK 293FT cells were transfected with
combinations of CRISPR components. Since DSBs in mam-
aalian 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 cleav-
age activity at the target EMX1 locus (FIG. 8) (see e.g.
Guschin et al., 2010, Methods Mol Biol 649: 247). Co-transf-
fection of all four CRISPR components was able to induce up
to 5.0% cleavage in the protospacer (see FIG. 2D). Co-transf-
fection 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
able to degrade pre-crRNA and 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. 4-7, 12, and 13). These results define a
three-component system for efficient CRISPR-mediated
genome modification in mammalian cells. To optimize the
cleavage efficiency, Applicants also tested whether different
isofoms 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 II 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 III (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 Pol3 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 nuclease array for SpCas9-mediated minor insertions and deletions. SpRNase III was expressed with and without SpRNase III, tracrRNA, and a pre-crRNA array carrying the EMX1-targeted spacer. FIG. 2E illustrates a schematic representation of base pairing between target locus and EMX1-targeting crRNA, 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 adopted, 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 transfected cells (FIGS. 3A and 8). In parallel, the bicistronic vectors were used to express a pre-crRNA (DR-guide sequence-DR) with SpCas9, to induce processing into crRNA with a separately expressed tracrRNA (compare FIG. 13B top and bottom). FIG. 9 provides schematic illustrations of bicistronic expression vectors for pre-crRNA 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 (GTTTTAGCCTTA (SEQ ID NO: 27)) and a partial tracrRNA sequence (TAGCAAGTTAAATAGGCTAGTCCTGGTTTTT (SEQ ID NO: 28)). Guide sequences can be inserted between BbsI sites using annealed oligonucleotides. Sequence design for the oligonucleotides are shown below the schematic illustrations in FIG. 9, with appropriate ligation adapters indicated. WPPE 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 protospacers in human PVALB (FIG. 15A) and mouse Th (FIG. 15B) loci. Schematics of the gene loci and the location of three protospacers within the last exon of each are provided. The underlined sequences include 30 bp of protospacer sequence and 3 bp at the 3' end corresponding to the PAM sequences. Protospacers 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-crRNA 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 protospacers, indicated by the underlined sequences. FIG. 13B provides a schematic of the pre-crRNA/tracrRNA complex showing hybridization between the direct repeat region of the pre-crRNA and tracrRNA (top), and a schematic of a chimeric RNA design comprising a 20 bp guide sequence, and tracrRNA sequences containing partial direct repeats and tracrRNA sequences hybridized in a hairpin structure (bottom). Results of a Surveyor assay comparing the efficacy of Cas9-mediated cleavage at five protospacers in the human EMX1 locus is illustrated in FIG. 13C. Each protospacer is targeted using either processed pre-crRNA/tracrRNA complex (crRNA) or chimeric RNA (chIRNA).

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 Research, 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 protospacer 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 chimeric crRNA (chIRNA) as well as SpCas9. The chimeric guide RNA
contains a 20-bp guide sequence corresponding to the protospacer 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 amplicon 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 grayscale. 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 CbB 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 efficiency of protospacer-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 transduced with challenge plasmids containing the corresponding protospacer 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 further 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 TALEN (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 NEHEJ 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 NEHEJ 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, hSpCas9 or hSpCas9n, as well as a HR template to introduce a pair of restriction sites (HindIII and NheI) 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 following 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 NEHEJ 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 gel showing efficient 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. Jinck 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 CRISPR1 of Streptococcus thermophilus LMD-9 for heterologous expression in mammalian cells to achieve CRISPR-mediated genome editing. FIG. 12A provides a Schematic illustration of CRISPR 1 from S. thermo-
philus 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 EF1\alpha promoter. Mature versions of tracrRNA and crRNA are expressed using the U6 promoter to promote transcription initiation. Sequences from the mature crRNA and tracrRNA are illustrated. A single base indicated by the lower case “u” 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 PAM sequence targets of the S. thermophilus CRISPR system in the human EMX1 locus. Two protospacer sequences are highlighted and their corresponding PAM 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 the 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’-N,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 NNAGAAW, may be identified by searching for 5’-N,NNAGAAW-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 NGGNG, may be identified by searching for 5’-N,NGGNG-3’ both on the input sequence and on the reverse-complement of the input. The value “x” in N, 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 CBlα 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 GUUUUGAGCUA (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). Protospacer IDs and their corresponding genomic target, protospacer sequence, PAM sequence, and strand location are provided in Table D. Guide sequences were designed to be complementary to the entire protospacer sequence in the case of separate transcripts in the hybrid system, or only to the underlined portion in the case of chimeric RNAs.

<table>
<thead>
<tr>
<th>protospacer ID</th>
<th>genomic target sequence (5’ to 3’)</th>
<th>PAM strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EMX1 GGAATTCGATGCTACCTGGAATG</td>
<td>TGG +</td>
</tr>
<tr>
<td>2</td>
<td>EMX1 CATTGGACGCTACCTGGAATG</td>
<td>TGG -</td>
</tr>
<tr>
<td>3</td>
<td>EMX1 GGAAGGGCCGCTACCTGGAATG</td>
<td>GGG +</td>
</tr>
<tr>
<td>4</td>
<td>PVALB GCTGGACGCTACCTGGAATG</td>
<td>AGG +</td>
</tr>
<tr>
<td>5</td>
<td>PVALB ATGCAAGGGCCGCTACCTGGAATG</td>
<td>TGG +</td>
</tr>
</tbody>
</table>

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 Technolo-
gies), 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 E), 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 minutes, 95°C to 85°C ramping at -2°C/minute, 85°C to 25°C at -0.25°C/minute, 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-rad). Quantification was based on relative band intensities.

<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-P</td>
<td>EMX</td>
<td>AAAAAACCTTCTCTTCGTGCG</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 36)</td>
<td></td>
</tr>
<tr>
<td>Sp-EMX1-R</td>
<td>EMX</td>
<td>GGGAGAGTGGAGCAACGGAGAGG</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 37)</td>
<td></td>
</tr>
<tr>
<td>Sp-PVALB-P</td>
<td>PVALB</td>
<td>CTGAAAAGGCCATGCTGAC</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 38)</td>
<td></td>
</tr>
<tr>
<td>Sp-PVALB-R</td>
<td>PVALB</td>
<td>GCCAGCAACACTTCCTGTGCTT</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 this example, each SpCas9 target site was operationally defined as a 20 bp sequence followed by an NGG protospacer adjacent motif (PAM) sequence, and we identified all sequences satisfying this 5’-N₂₀-NGG-3’ 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’-MNKNÖNNÖNNÖNG-3’ 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 m5. Zebrafish genome danRer7. D. melanogaster genome dm6 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 chriRNA 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 chriRNA(+) indicate that up to the +n 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 (chriRNA(+67) and chriRNA(+85)) mediated DNA cleavage at all three EMX1 target sites, with chriRNA(+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. 18a 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 chriRNAs. chriRNA(+67) and chriRNA(+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.ribi.univie.ac.at/cgi-bin/RNAfold.cgi) using minimum free energy and partition function algorithm. Pseudoknots for each base (reproduced in grayscale) indicates the probability of pairing. Because chriRNAs 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 site-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 bicistronic 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 CBh promoter and terminated with the bGH polyA signal (bGH PA). The expanded sequence illustrated immediately below the schematic corresponds to the region surrounding the guide sequence insertion site, and includes, from 5’ to 3’ of the U6 promoter (first shaded region), BbsI cleavage sites (arrows), partial direct repeat (tracr mate sequence GTTITAGACTA (SEQ ID NO: 27), underlined), loop sequence GAAA, and +85 tracr sequence (underlined sequence following loop sequence). An exem-
NLS-SpCas9n-NLS (the D10A nickase mutation is lowercase):

SEQ ID NO: 50

MDYKMDGDYKEDHIDYDDKKMAPPPKKEKKNGHVGVAPADKYSTISGAI

GTNSVGKAVTDYKYPSSFKVPLQNTDRHSI1NKELIGALLFDGQETARA

TRLRATTAYKRTERKHDKYCLQF1QFSQKSMKADVSFSPHLELSFVEEDK

KHERHPQCNVQDEVAYNEKTTYHFLKREVLDSVVTDLRSALYILALAHM

1IFKHPFLKEDGHPNDSDMVDKLYPLQVTQVNLQFENHIPASGDVDAIKAI

LAS1LGSKSRRLNLIQA1QLPGKEKGLPQNLISLQGLTIPENFDLADEB

AQKLQECSDTVDDLNLNLQIGDQALPLAANSLNLDDLNSLDLUNTNE

ITEAALPSASHKMDYDHQDQTTLKLALVRQFLEKEIFDPQDKSYYGAY

YIDQAGASQGSF2Q1F1PFK1F1FQDEELQXNLKSRDLELRKQFDTQDSIP

NLS-SmCherry-SpRNase3:

SEQ ID NO: 48

MFLFLSTSLSSLSSSRLLVSKGGEEDMA1IKEPFRKPKHVEVGSVHGEFEL

KEEKKKEPQFQTKLKVQKGPIPFLDNLSSCPQMGYKAYKHPAD1

PDYKLPSFEGFQWEKRYKHFEDGQVTQVDSQLQDGEF1KYCVLRGTFNP

PSGQVQMGKTKMKEASMERSMPEGDAGKELQKRPQVQAHYDWKCYAKVT

TYKAKVKQPGAYNWK1IKLMTSHMEDITVYEQYERABBROSSTGQDSEL

YKGSQKELLESLTSTDINQNDLTLLETAPHTSYNHKR1LMWNSHERE

FLQDAVLQILLISSEYLFKPEKTBDMSKSLIVPRERBLSAGSRSCFSDP

AYIKOGSEQQGKQNDTILDGFEPARGALLDLGDAVVLKQVMI

PVQKSNHFRVYKTECLQFEPLQTKGDVADIYIQVISKHGPAHAEFVSI

VNVGAILSELQEGKSRQAPDQGAMNLASELY

SpRNase3-mCherry-NLS:

SEQ ID NO: 49

MKQOLELLETSFDIPQNDTLDLLETAPHTSYNHKR1LMWNSHERENFLG

DAVLQILISSEYLFKPEKTBDMSKSLIVPRERBLSAGSRSCFSDP

AYIKOGSEQQGKQNDTILDGFEPARGALLDLGDAVVLKQVMI

PVQKSNHFRVYKTECLQFEPLQTKGDVADIYIQVISKHGPAHAEFVSI

VNVGAILSELQEGKSRQAPDQGAMNLASELY

NLS-135

SEQ ID NO: 51

hEZX1-1R-Template-HindII-Nhel:

GAATCGCTCCCTACGACCCCTCCTCTCCCTCCCTCCTGCTGCTAAAGAG

AATGAGCAGCTACTGTGATTTGCAGACTCTTGAGGAGTGACCTGACCTG
Chimeric RNA Containing 48 Tracr RNA (S. Pyogenes SF370)

(SEQ ID NO: 55)

aggccgtatattgcatcggcataagaagaccagacgactgtgc

tgtagattaattttaattatcaacacatattataatcatgttccgcttcat

tccccctgcgcgttttctttgtttttttttttttttt

(first underline = guide sequence; second underline = tracr sequence; bold = terminator)

Chimeric RNA Containing 54 Tracr RNA (S. Pyogenes SF370)

(SEQ ID NO: 56)

aggccgtatattgcatcggcataagaagaccagacgactgtgc

tgtagattaattttaattatcaacacatattataatcatgttccgcttcat

tccccctgcgcgttttctttgtttttttttttttttt

(first underline = guide sequence; second underline = tracr sequence; bold = terminator)

Chimeric RNA Containing 67 Tracr RNA (S. Pyogenes SF370)

(SEQ ID NO: 57)

aggccgtatattgcatcggcataagaagaccagacgactgtgc

tgtagattaattttaattatcaacacatattataatcatgttccgcttcat

tccccctgcgcgttttctttgtttttttttttttttt

(first underline = guide sequence; second underline = tracr sequence; bold = terminator)
Example Chimeric RNA for *S. Thermophilus* LMD-9 CRISPR Cas9 (with PAM of NNAGAAW)
Example Chimeric RNA for *S. Thermophilus* LMD-9 CRISPR1Cas9 (with PAM of NNAGAAW)

**catttatgacagggttttcgattatattaaaa**
(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)

Example Chimeric RNA for *S. Thermophilus* LMD-9 CRISPR1Cas9 (with PAM of NNAGAAW)

**cattatatgacagggttttcgattatattaaaa**
(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)

Example Chimeric RNA for *S. Thermophilus* LMD-9 CRISPR1Cas9 (with PAM of NGGNG)

**cattatatgacagggttttcgattatattaaaa**
(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)

Codon-Optimized Version of Cas9 from *S. Thermophilus* LMD-9 CRISPR3 Locus (with an NLS at Both 5’ and 3’ Ends)

**cattatatgacagggttttcgattatattaaaa**
(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)
Example 5

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 65% when used in combination with recombining in E. coli. Applicants exhaustively analyzed Cas9 target requirements to define the range of tar-
getable 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 muclense 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 recombination, a technique that promotes homologous recombination of linear DNA or oligonucleotides. However, because there is no selection of mutations, recombination 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 dsDNA 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 recombination 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 φ232.5. This target was integrated into the srlA chromosomal locus of a second strain R6. An altered target sequence containing a mutation in the PAM region was integrated into the srlA locus of a third strain R6, rendering this strain ‘immune’ to CRISPR cleavage (FIG. 28a). Applicants transformed R6 and R6 cells with genomic DNA from crR6 cells, expecting that successful transformation of R6 cells should lead to cleavage of the target locus and cell death. Contrary to this expectation, Applicants isolated R6 transformants, albeit with approximately 10-fold less efficiency than R6 transformants (FIG. 28b). Genetic analysis of eight R6 transformants (FIG. 28b) revealed that the majority are the product of a double recombination event that eliminates the toxicity of Cas9 targeting by replacing the φ232.5 target with the crR6 genome’s wild-type srlA 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. 23).

To create a simplified system for genome editing, Applicants modified the CRISPR locus in strain crR6 by deleting cas1, cas2 and csn2, genes which have been shown to be dispensable for CRISPR targeting, yielding strain crR6M (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 cells with PCR products of the wild-type srlA gene or the mutant R6 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 crR6DNA alone (FIG. 23a). The efficiency of editing was also substantially increased, with 8/8 transformants tested containing a wild-type srlA copy and 7/8 containing the PAM mutation present in the R6 target (FIG. 23b). 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 abolition 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 heterogeneous 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 2×10^4 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 NNGGN 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 NAG to NAG or NNNGN 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 spe1 target present in R6^232.25 cells (FIG. 23a) was randomized and transformed into crR6 and R6 cells (FIG. 24c). 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 contrast, mutations to any nucleotide in the proximal positions (6 to 1, except 3) eliminated Cas9 activity, although at different levels for each particular substitution. At position 3, only two substitutions affected CRISPR activity and with different strength. Applicants concluded that, although seed sequence mutations can prevent CRISPR targeting, there are restrictions regarding the nucleotide changes that can be made in each position of the seed. Moreover, these restrictions can most likely vary for different spacer sequences. Therefore Applicants believe that mutations in the PAM sequence, if possible, should be the preferred editing strategy. Alternatively, multiple mutations in the seed sequence may be introduced to prevent Cas9 nuclease activity.

Cas9-Mediated Genome Editing in S. Pneumoniae:
To develop a rapid and efficient method for targeted genome editing, Applicants engineered strain crR6Rk, a strain in which spacers can be easily introduced by PCR (FIG. 33). 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 (CGT 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 are 218 nt downstream of the protospacer region (AAA to GCA, generating a Tscl 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. 25). Applicants genotyped 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. 29). 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. 25e). 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 rRNA 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::O) (FIG. 26a,b). In the absence of kanamycin selection, the fraction of edited colonies was on the order of \(10^{-6}\) (erythromycin-resistant cftu/total cftu) (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 \(10^{-3}\) (kanamycin- and erythromycin-resistant cftu/kanamycin-resistant cftu) (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 (cftu) obtained after co-transformation of JEN53 cells with the CRISPR::ermAM(stop) or CRISPR::O constructs.

Applicants counted 5.3 times less kanamycin-resistant colonies after transformation of the ermAM(stop) construct \((2.5 \times 10^4/4.7 \times 10^3, \text{FIG. 35a})\), 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 introduce 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::O construct (FIG. 26b), indicating that there was a modest induction of recombination. Taken together, these results showed that co-selection of transformable 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::O cftu, \(2.6 \times 10^{-8}/7.1 \times 10^{-7}\) 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 was 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 recombining, using Cas9-induced cell death to select for the desired mutations. The pCas9 plasmid was introduced into the recombining strain HME63 (31), which contains the Gam, Exo and Beta functions of the ϕred phage. The resulting strain was co-transformed with the pCRISPR::rpsl plasmid (or a pCRISPR::O control) and the W542 oligonucleotide (FIG. 27a). The recombining efficiency was \(5.3 \times 10^{-5}\), 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. 27a and 29). Applicants observed that the number of cftu was reduced by about three orders of magnitude after transformation of the pCRISPR::rpsl plasmid than the control plasmid \((4.8 \times 10^7/5.3 \times 10^5, \text{FIG. 38a})\), 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::O cftu, was \(2.5 \times 10^{-8}/1.2 \times 10^{-7}\). 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 recombining efficiency of the rpsl mutation, \(5.3 \times 10^{-5}\), 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:mpy or pCRISPR:o1 (Fig. 27D). As in the case for S. pneumoniae, Applicants observed a modest induction of recombination, about 6.7 fold (2.0×10⁻⁶/3.0×10⁻⁶). Taken together, these results indicated that the CRISPR system provided a method for selecting mutations introduced by recombination.

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 recombiningogenic 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 recombiningogenic of mutagenic oligonucleotides. To use this methodology in microbes where recombiningogenic is not a 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’ arise 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 flanking 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 specific 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 recombiningogenic by selecting a subpopula-
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 ul of overnight *S. pneumoniae* cultures were pelleted, resuspended in 60 ul of lysozyme solution (2 mg/ml) and incubated 30 min at 37°C. The genomic DNA was extracted using QIAprep Spin Miniprep Kit (Qiagen).

Strain Construction.

All primers used in this study are provided in Table G. To generate *S. pneumoniae* crf6M, an intermediate strain, LAM226, was made. In this strain the aphA-3 gene (providing kanamycin resistance) adjacent to the CRISPR array of *S. pneumoniae* crf6 strain was replaced by a cat gene (providing chloramphenicol resistance). Briefly, crf6 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* crf6 cells and chloramphenicol-resistant transformants were selected. To generate *S. pneumoniae* crf6M, *S. pneumoniae* crf6Rc 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* crfR6Rc, *S. pneumoniae* crf6M 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* crf6M cells and chloramphenicol-resistant transformants were selected.

To generate *S. pneumoniae* crfR6Rk, *S. pneumoniae* crf6Rk 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* crfR6Rc cells and kanamycin-resistant transformants were selected.

To generate JEN37, *S. pneumoniae* crfR6Rk genomic DNA was amplified by PCR using primers L430/W536 and W537/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 pFW15 using primers L457/L458. Each PCR product was gel-purified and all three were fused by SOEing PCR with primers L422/L426. The resulting PCR product was transformed into competent *S. pneumoniae* crfR6Rc 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* crfR6Rk 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* crfR6Rc cells and kanamycin-resistant transformants were selected.

Plasmid Construction.

pD97 was constructed through phosphorylation and annealing of oligonucleotides B296/B297, followed by ligation in pLZ2spec digested by EcoRI/BamHI. Applicants fully sequenced pLZ2spec and deposited its sequence in genebank (accession: KC112384).

pD98 was obtained after cloning the CRISPR leader sequence was cloned together with a repeat-spacer-repeat unit into plLZ2spec. This was achieved through amplification of crfR6Rc DNA with primers B298/B320 and B299/B321, followed by SOEing PCR of both products and cloning in plLZ2spec with restriction sites BamHI/EcoRI. In this way the spacer sequence in pD98 was engineered to contain two Bsal restriction sites in opposite directions that allow for the scar-less cloning of new spacers.

pD99 to pDB108 were constructed by annealing of oligonucleotides B300/B301 (pD99), B302/B303 (pD100), B304/B305 (pD101), B306/B307 (pD102), B308/B309 (pD103), B310/B311 (pD104), B312/B313 (pD105), B314/B315 (pD106), B315/B317 (pD107), B318/B319 (pD108), followed by ligation in pD98 cut by Bsal.

The pCas9 plasmid was constructed as follow. Essential CRISPR elements were amplified from *Streptococcus pyogenes* 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 Bsal type IIIS sites were introduced in 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 B352+B353 and cloning in the Bsal 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 G). Editing constructs were made by SOEing PCR fusing PCR products A (PCR A), PCR products B (PCR B) and PCR products C (PCR C) when applicable (Table G). The CRISPR:G and CRISPR::erm AM(stop) targeting constructs were generated by PCR amplification of JEN62 and crfR6 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 R6232.5 genomic DNA with primers W377/L426. This PCR product was then assembled with the cat gene and the srtA 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 crfR6. 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 cut gene and srtA 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 crfR6. After transformation,
cells were plated on chloramphenicol selection. For each sample more than $2 \times 10^7$ cells were pooled together in 1 ml of THY/1 genomic DNA was extracted with the Promega Wizard kit. Primers B251/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 crR6 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 crR6 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 crR6 sample over the R6 sample was computed and is denoted $r_{pAM}$ where $l_j,k,l$ in are one of the 4 possible bases. The following statistical model was constructed:

$$
\log(r_{pAM}) = \beta_0 + \beta_1 \cdot b_2 + \beta_2 \cdot b_3 + \beta_3 \cdot b_4 + \beta_4 \cdot b_2 \cdot b_3 + \beta_5 \cdot b_2 \cdot b_4 + \beta_6 \cdot b_3 \cdot b_4 + \epsilon
$$

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, $b_2b_3$ is the interaction between the second and third bases, $b_3b_4$ between the third and fourth bases. An analysis of variance was performed:

**ANOVA table**

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>b3</td>
<td>3</td>
<td>151,693</td>
<td>50,564</td>
<td>601,8450</td>
<td>&lt;2.2e-16 ***</td>
</tr>
<tr>
<td>b2</td>
<td>3</td>
<td>90,251</td>
<td>30,174</td>
<td>359,1454</td>
<td>&lt;2.2e-16 ***</td>
</tr>
<tr>
<td>b4</td>
<td>3</td>
<td>1,881</td>
<td>0.627</td>
<td>7,4023</td>
<td>0.070e-05***</td>
</tr>
<tr>
<td>b2b3</td>
<td>9</td>
<td>228,940</td>
<td>25,438</td>
<td>302,7738</td>
<td>&lt;2.2e-16 ***</td>
</tr>
<tr>
<td>b3b4</td>
<td>9</td>
<td>3,010</td>
<td>0.334</td>
<td>3,9800</td>
<td>5,227e-05***</td>
</tr>
<tr>
<td>Residuals</td>
<td>996</td>
<td>83,680</td>
<td>0.084</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When added to this model, b4 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.

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

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G-G-C</td>
<td>-2.79911</td>
<td>-2.97511</td>
<td>-2.62311</td>
</tr>
<tr>
<td>G-G-T</td>
<td>-2.7809</td>
<td>-2.9595</td>
<td>-2.6049</td>
</tr>
<tr>
<td>G-A-C</td>
<td>-2.81643</td>
<td>-2.99244</td>
<td>-2.64043</td>
</tr>
<tr>
<td>G-C-C</td>
<td>-2.77903</td>
<td>-2.95504</td>
<td>-2.60303</td>
</tr>
<tr>
<td>G-C-G</td>
<td>-2.68467</td>
<td>-2.82468</td>
<td>-2.47267</td>
</tr>
<tr>
<td>G-C-T</td>
<td>-2.79718</td>
<td>-2.97319</td>
<td>-2.62118</td>
</tr>
<tr>
<td>G-A-C</td>
<td>-2.67068</td>
<td>-2.84668</td>
<td>-2.49468</td>
</tr>
<tr>
<td>G-C-G</td>
<td>-2.73525</td>
<td>-2.91125</td>
<td>-2.55925</td>
</tr>
<tr>
<td>G-C-T</td>
<td>-2.7976</td>
<td>-2.92159</td>
<td>-2.9736</td>
</tr>
<tr>
<td>G-C-T</td>
<td>-2.86414</td>
<td>-2.66513</td>
<td>-3.01714</td>
</tr>
<tr>
<td>G-G-T</td>
<td>-2.76409</td>
<td>-2.58809</td>
<td>-2.94009</td>
</tr>
<tr>
<td>G-G-T</td>
<td>-2.76781</td>
<td>-2.59181</td>
<td>-2.94381</td>
</tr>
<tr>
<td>G-G-A</td>
<td>-2.13964</td>
<td>-2.31565</td>
<td>-1.96364</td>
</tr>
<tr>
<td>G-A-A</td>
<td>-0.62511</td>
<td>-0.80111</td>
<td>-0.4491</td>
</tr>
<tr>
<td>G-A-C</td>
<td>-0.65947</td>
<td>-0.85347</td>
<td>-0.48346</td>
</tr>
<tr>
<td>G-A-T</td>
<td>-0.40126</td>
<td>-0.46525</td>
<td>-0.81726</td>
</tr>
<tr>
<td>G-C-A</td>
<td>-0.67679</td>
<td>-0.80078</td>
<td>-0.85319</td>
</tr>
<tr>
<td>G-C-C</td>
<td>-0.63939</td>
<td>-0.46339</td>
<td>-0.81539</td>
</tr>
<tr>
<td>G-C-G</td>
<td>-0.50903</td>
<td>-0.33303</td>
<td>-0.68303</td>
</tr>
<tr>
<td>G-A-C</td>
<td>-0.65734</td>
<td>-0.48154</td>
<td>-0.83354</td>
</tr>
<tr>
<td>G-C-A</td>
<td>-0.53104</td>
<td>-0.35593</td>
<td>-0.70704</td>
</tr>
<tr>
<td>G-C-C</td>
<td>-0.59561</td>
<td>-0.4196</td>
<td>-0.77161</td>
</tr>
<tr>
<td>G-C-G</td>
<td>-0.67959</td>
<td>-0.48195</td>
<td>-0.83396</td>
</tr>
<tr>
<td>G-A-T</td>
<td>-0.62763</td>
<td>-0.45163</td>
<td>-0.80363</td>
</tr>
<tr>
<td>G-C-C</td>
<td>-0.70149</td>
<td>-0.52549</td>
<td>-0.8775</td>
</tr>
<tr>
<td>G-C-G</td>
<td>-0.64245</td>
<td>-0.44844</td>
<td>-0.86045</td>
</tr>
<tr>
<td>G-A-T</td>
<td>-0.62817</td>
<td>-0.45216</td>
<td>-0.80417</td>
</tr>
</tbody>
</table>

**Pairwise comparisons of the effect of b4 on NYGN sequences using t-tests with pooled SD Data + b4**

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>9.2e-05</td>
<td>2.4e-06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0.31</td>
<td>1.00</td>
<td>1.2e-08</td>
<td></td>
</tr>
</tbody>
</table>

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

**Tukey multiple comparisons of means:**

<table>
<thead>
<tr>
<th></th>
<th>diff</th>
<th>hwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>b3b4</td>
<td></td>
<td></td>
<td></td>
<td></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 (FIG. 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 said to significantly disrupt interference. A normal distribution was fitted to these sequences using the fitdistr function of the MASS R package. The 0.99 quantile of the fitted distribution is shown as a dotted line in FIG. 24c. FIG. 72 shows a histogram of the data density with fitted normal distribution (black line) and 0.99 quantile (dotted line).

### TABLE F

<table>
<thead>
<tr>
<th>2nd position</th>
<th>3rd position</th>
<th>4th position</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>AAA 1.04</th>
<th>ACA 1.12</th>
<th>AGA 0.73</th>
<th>ATG 1.10</th>
<th>ATA 1.10</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAC 1.07</td>
<td>ACC 1.04</td>
<td>AGC 0.64</td>
<td>ATC 0.97</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGC 1.00</td>
<td>AGG 1.09</td>
<td>AGT 0.61</td>
<td>ATG 1.07</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAT 0.98</td>
<td>ACT 1.02</td>
<td>ATT 0.65</td>
<td>ATT 1.01</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>CCA 1.05</td>
<td>CGA 0.99</td>
<td>CTA 1.07</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAC 1.04</td>
<td>CCC 1.02</td>
<td>CGC 0.98</td>
<td>CTC 1.04</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGA 1.08</td>
<td>CGC 1.08</td>
<td>CGG 0.81</td>
<td>CGT 1.05</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT 1.13</td>
<td>CCT 1.05</td>
<td>CGT 0.70</td>
<td>CTT 1.08</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>GCA 0.97</td>
<td>GCC 0.95</td>
<td>GGA 0.08</td>
<td>GTA 0.99</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>GAC 0.92</td>
<td>GGC 1.00</td>
<td>GGC 0.05</td>
<td>GTG 1.15</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAG 0.96</td>
<td>GGG 0.98</td>
<td>GGG 0.07</td>
<td>GTG 0.98</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAT 0.98</td>
<td>GGT 0.59</td>
<td>GGT 0.06</td>
<td>GTC 1.05</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>TAA 1.08</td>
<td>TCA 1.16</td>
<td>TCA 1.05</td>
<td>TTA 1.24</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>TAG 1.02</td>
<td>TCG 1.16</td>
<td>TGT 0.77</td>
<td>TGT 1.02</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAT 1.01</td>
<td>TCT 1.12</td>
<td>TTA 1.21</td>
<td>TTT 1.02</td>
<td>T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE G

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>B217</td>
<td>TCCTAGACAGAGTTTCTGATATTACGCTCACTTATTTTAAAGCCTATGCTGTTTTTG</td>
</tr>
<tr>
<td>B218</td>
<td>GTGACAGATATCGAATCCGATGATGTGTTGTTGACCATTAAACAAGC</td>
</tr>
<tr>
<td>B229</td>
<td>GCGTTCTACAGTTTTCAGTAGAGG</td>
</tr>
<tr>
<td>B230</td>
<td>GCCAAATGCAAGAGAACACCATCAGAC</td>
</tr>
<tr>
<td>B250</td>
<td>NNNNGCGGACCCATGCTCTTAAC</td>
</tr>
<tr>
<td>B251</td>
<td>NNNNTATTCTGCTGAATTTTTGCT</td>
</tr>
<tr>
<td>B255</td>
<td>CTTTCAGCTACAGTAGCAGCAGA</td>
</tr>
<tr>
<td>B256</td>
<td>CAAAAAATTCTAGTCTTTTCCCCCAATTAACCACTCACTTAA</td>
</tr>
<tr>
<td>B257</td>
<td>AGGGTTTTTATTGGCCGGGGCCCAAGAAGCTAAAGCTTATTTTGCCATTCC</td>
</tr>
<tr>
<td>B258</td>
<td>CTTCATGCATACAAATCTCAATTCC</td>
</tr>
<tr>
<td>B269</td>
<td>TGCGTCGATTTTTCAGGATTG</td>
</tr>
<tr>
<td>B270</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
<tr>
<td>B271</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
<tr>
<td>B272</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
<tr>
<td>B273</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
<tr>
<td>B274</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
<tr>
<td>B275</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
<tr>
<td>B276</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
<tr>
<td>B277</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
<tr>
<td>B278</td>
<td>CTTGACGAGAAGGACATGGAATGGAGACACCAAAAGAGGGCAAG</td>
</tr>
</tbody>
</table>

Primers used in this study (SEQ ID NOS 68-183, respectively, in order of appearance).
<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>B279</td>
<td>GCCCTTTTTGCTGATTTTCG</td>
</tr>
<tr>
<td>B280</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B281</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B282</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B283</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B284</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B285</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B286</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B287</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B288</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B289</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B290</td>
<td>CAATGGCTGAAATCGAGGCAAAAAAGCCGAGAAGAAATC</td>
</tr>
<tr>
<td>B291</td>
<td>gatccCTCATCGCTGACAAACCAACACCTTG</td>
</tr>
<tr>
<td>B292</td>
<td>aattcCAGGGTGACCTGACG</td>
</tr>
<tr>
<td>B293</td>
<td>CATGGATCTGACAGGACAGAGAAGAAGAAATC</td>
</tr>
<tr>
<td>B294</td>
<td>CATGGATCTGACAGGACAGAGAAGAAGAAATC</td>
</tr>
<tr>
<td>B295</td>
<td>CATGGATCTGACAGGACAGAGAAGAAGAAATC</td>
</tr>
<tr>
<td>B296</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B297</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B298</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B299</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B300</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B301</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B302</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B303</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B304</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B305</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B306</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B307</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B308</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B309</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B310</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B311</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B312</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B313</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B314</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B315</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B316</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B317</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B318</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B319</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>B320</td>
<td>AAAACATTTTTTCTCCTTCTTCTTCTATTACAAATATA</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence 5'-3'</td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
</tr>
<tr>
<td>B321</td>
<td>TGAGCAGCTCTCGGAAGCTCAAGGCTGTTCGTTTTAAGCTGCTAGCTGTTTTTG</td>
</tr>
<tr>
<td>B352</td>
<td>aacAGTCTTTACCCAGCCGGAGTCTCGGTTTTT</td>
</tr>
<tr>
<td>B353</td>
<td>aaACTCCCCCCGGCTCGGAGTAAGGATTGGCTGAAAGG</td>
</tr>
<tr>
<td>H2006_SP</td>
<td>AAGCCGTTACAGCCGGGCTCTTCGCGGATTACAGAAATCATCCTG</td>
</tr>
<tr>
<td>H2009_SP</td>
<td>GTGACTGCGGATGCTGCGGAAAGAGGATGACACTACTCTCCTT</td>
</tr>
<tr>
<td>H2016_SP</td>
<td>TTAGAAGAAATTCATCTCAAAAATATACCTACTACCTCCTAGTGAC</td>
</tr>
<tr>
<td>H2011_SP</td>
<td>TTTGATTGGATGCGCTACAGGATGGCTGACAGTATATATTTGAGATGAG</td>
</tr>
<tr>
<td>H2014_SP</td>
<td>GAAGACCTTTGAGCTCTCGCGAGACTGCTCTACGTTTTGGAGGAGCTCCATTCAACACGAGCAGATG</td>
</tr>
<tr>
<td>H2015_SP</td>
<td>CAACTTTGCGACACACAGCTGAGTTT</td>
</tr>
<tr>
<td>L403</td>
<td>AGTCACTCCAGGACACAATTAGG</td>
</tr>
<tr>
<td>L409</td>
<td>CGTGTAAGGCTGACAGTGCGGTTG</td>
</tr>
<tr>
<td>L422</td>
<td>Tgcttccttcacaaaacaggg</td>
</tr>
<tr>
<td>L426</td>
<td>AAGCGAAGGGTGTGGCGCACCA</td>
</tr>
<tr>
<td>L430</td>
<td>GTAGCTTATTCAGTCTCTAGG</td>
</tr>
<tr>
<td>L444</td>
<td>CGTTGGTGAGCTATTGGGCTGAAATTACGAACTTCTCTCTGACG</td>
</tr>
<tr>
<td>L445</td>
<td>CGTCAGGCGAGATTCGTAATTTGGACCCCCATTAGTTTACAAACAGG</td>
</tr>
<tr>
<td>L446</td>
<td>QAATTAGGAGCTCTATTGGTGCTGTTTCAATCAAAATTAGG</td>
</tr>
<tr>
<td>L447</td>
<td>CATATAAGTTATTACCTAATAACCCACCAATAGGCTCCATTAAACTC</td>
</tr>
<tr>
<td>L448</td>
<td>ATATATCTTAATACAAATAATAAGG</td>
</tr>
<tr>
<td>L457</td>
<td>CTTGACGTTAGCTAGCTACG</td>
</tr>
<tr>
<td>L458</td>
<td>GCCACCGTGATCATCACGGTCTAGG</td>
</tr>
<tr>
<td>L459</td>
<td>CCTTAGAACGTAGATACGGGTGCAAAATATGAGGCCAACATAAAATATAT</td>
</tr>
<tr>
<td>L461</td>
<td>GCCGTTCCCCAGTGCTATCTGGTTGACAGAATAGGGTGCC</td>
</tr>
<tr>
<td>L461</td>
<td>TTCACACATTTCCCCATTTGCTTC</td>
</tr>
<tr>
<td>L488</td>
<td>CCATAATTCTTAGTTAAAAATGCAAGCTCCAGTCAGCTACCTCC</td>
</tr>
<tr>
<td>W256</td>
<td>AGACGATTCAATAGCACAATAAGG</td>
</tr>
<tr>
<td>W285</td>
<td>GTTGGGACACATCAACACCAGCTAGCTCTTAACACTCTGTGAGAC</td>
</tr>
<tr>
<td>W287</td>
<td>GCTAGCTGTTTTGGATGCTGCCAAAACcatttattttaacacagaggt</td>
</tr>
<tr>
<td>W288</td>
<td>GCTAGCTGTTTTGGATGCTGCCAAAACGCAACATTTAGTCAACAAAGG</td>
</tr>
<tr>
<td>W326</td>
<td>AATTCTTTCTTCATCATCGGTC</td>
</tr>
<tr>
<td>W327</td>
<td>AAAGAAAGATGAAAGATTGTGCTAG</td>
</tr>
<tr>
<td>W341</td>
<td>GCTACTAATCAAAATAGTGAGGAGG</td>
</tr>
<tr>
<td>W354</td>
<td>GTTTTTCAAAACTCAGGGTTG</td>
</tr>
<tr>
<td>W355</td>
<td>AAAAATGAAAAAATGCTAAAGACAC</td>
</tr>
<tr>
<td>W356</td>
<td>ATTCGCTGACTAGCTCCATCTTTTTAAAGGGTTTGCGACCATTTCAACAGG</td>
</tr>
<tr>
<td>W357</td>
<td>TTTAAAGAAACGCAATCGGTTTCAGAARATTGTTTGGAGCTATGCTGTTTTTG</td>
</tr>
</tbody>
</table>

*Note: The table continues with additional primers.*
### TABLE G-continued

Primer used in this study (SEQ ID NOS 68-183, respectively, in order of appearance).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5' - 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>W365</td>
<td>AAAACGTATACGTTTCTTCTTAATTTAATATTGTGGCACCATTCTAAAAAGACG</td>
</tr>
<tr>
<td>W366</td>
<td>AATGGAATTTTTAAAGAAAACGGATAAGCTTGTTTTTTAGAGCTAAGCTGTTTTG</td>
</tr>
<tr>
<td>W370</td>
<td>GTCCTGAAACACGAAAGGATATCAGTTATTTTTTTAACTTAATGAC</td>
</tr>
<tr>
<td>W371</td>
<td>GAACCCGATACCGTTTGGTTTAAGGACAGTAAAGGCACATACAT</td>
</tr>
<tr>
<td>W376</td>
<td>CGATTTCACGGACATTGCTGTC</td>
</tr>
<tr>
<td>W377</td>
<td>GCTTTTGCAGGGAATGCTGAAATCGSNBNNAAAAGCGGACAGGGAATCAAC</td>
</tr>
<tr>
<td>W391</td>
<td>TCGTGCTACACCCCAACACCTCGTCTGTTTGGACACATTCTAAAAAGACG</td>
</tr>
<tr>
<td>W392</td>
<td>GCTCACTAGCGGAAATGGATGATGTTAAGGACACTATGTTGGTATG</td>
</tr>
<tr>
<td>W393</td>
<td>TTGTTCTACTTTCTTTTC</td>
</tr>
<tr>
<td>W397</td>
<td>CAGGTGGTTGCTTCTGATGGTATATGAAAACCATCACCTT</td>
</tr>
<tr>
<td>W398</td>
<td>GGAGTTTACCTCAATGCCAACACCAACCCACAAACCTCAGT</td>
</tr>
<tr>
<td>W403</td>
<td>GCTGATATCTGATGATGAC</td>
</tr>
<tr>
<td>W404</td>
<td>TTACGGAAACGGAAATTATCTGAC</td>
</tr>
<tr>
<td>W405</td>
<td>AAAGCTAGAGTTGCCAACATGG</td>
</tr>
<tr>
<td>W431</td>
<td>GTGGTTTGATGATGTTAATCTCCCACACTCCCTTC</td>
</tr>
<tr>
<td>W432</td>
<td>GATGGGAGTTTACAACACCTCAGTTACACCCACACACCTG</td>
</tr>
<tr>
<td>W433</td>
<td>GTTCCCACTATGGACACCAATGACACATGAGAATAG</td>
</tr>
<tr>
<td>W434</td>
<td>GTGGATCTATGGTGCTGAATAAGTGAGCTATTGGG</td>
</tr>
<tr>
<td>W463</td>
<td>CTGATTTGTATATAATTTCCAGGACTATTTACCTCACCCTTC</td>
</tr>
<tr>
<td>W464</td>
<td>GCAATAGTGCTCAAAATTTAAATAAATACGATTAAAATCAG</td>
</tr>
<tr>
<td>W465</td>
<td>GGTTTGGACACCAATCAAAAACGACTATGCTCAAACGCAAGTGATATATAG</td>
</tr>
<tr>
<td>W466</td>
<td>GTGTTGGAGACGTATGCTTGTTGAAAGTGGTCTCCCAACGCTACTACCGGATG</td>
</tr>
<tr>
<td>W542</td>
<td>ATACTTTACGCAGCGGAGGTTCCGTTTGYAGGGTAGGATGATGATATACAGAGATGTCAT</td>
</tr>
</tbody>
</table>

### TABLE H

Design of targeting and editing constructs used in this study (SEQ ID NOS 184, 184, 184, 185, and 186, respectively, in order of appearance).

<table>
<thead>
<tr>
<th>Targeting Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Edition</strong></td>
</tr>
<tr>
<td>bgaA R &gt; A</td>
</tr>
<tr>
<td>bgaA NR &gt; AA</td>
</tr>
<tr>
<td>AbgaA</td>
</tr>
<tr>
<td>AactA</td>
</tr>
<tr>
<td>emB Stop</td>
</tr>
</tbody>
</table>
TABLE II-continued

Design of targeting and editing constructs used in this study (SEQ ID NO:
184, 184, 184, 185, and 186, respectively, in order of appearance).

| Δacta AbgaA | JEM51 (for K256/W469 W466/W403 same as the ones used for Δacta and AbgaA) | TCG |
| Left PCR | and JEM62 (for right PCR) |

| Editing Constructs |

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Edition</td>
<td>Template DNA</td>
<td>PCR A</td>
<td>PCR B</td>
<td>PCR C</td>
<td>SOEing PCR</td>
<td>Name of resulting strains</td>
</tr>
<tr>
<td>bgaA R &gt; A</td>
<td>R6</td>
<td>W403/W397 W398/W404 N/A</td>
<td>N/A</td>
<td>W403/W404</td>
<td>JEM56</td>
<td>W403/W404</td>
</tr>
<tr>
<td>bgaA NE &gt; AA</td>
<td>R6</td>
<td>W403/W431 W432/W433 W434/W404</td>
<td>W403/W404</td>
<td>JEM60</td>
<td>W403/W404</td>
<td></td>
</tr>
<tr>
<td>AbgaA</td>
<td>R6</td>
<td>B255/B256 B257/B258 N/A</td>
<td>N/A</td>
<td>B255/B258</td>
<td>JEM62</td>
<td>W393/W405</td>
</tr>
<tr>
<td>Δacta</td>
<td>R6</td>
<td>B230/W463 W464/W229 N/A</td>
<td>N/A</td>
<td>B230/B229</td>
<td>JEM51</td>
<td>W422/W426</td>
</tr>
<tr>
<td>Δacta AbgaA same as the ones used for Δacta and AbgaA</td>
<td>L422/W370 W371/L426 N/A</td>
<td>L422/L426</td>
<td>JEM64</td>
<td>same as the ones used for Δacta and AbgaA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example 6

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):

```
GGATCCATTCAAGACGATAGCAAGTTAAATAGTCCTGCTATTC
```

Optimized Direct Repeat 1 (Mutation Underlined):

```
GTGGTACAGCTATATGTGGATCCGACAAACACCC
```

Optimized tracrRNA 2 (Mutation Underlined):

```
GGATCCATTCAAGACGATAGCAAGTTAAATAGTCCTGCTATTC
```

Optimized Direct Repeat 2 (Mutation Underlined):

```
GTGGTACAGCTATATGTGGATCCGACAAACACCC
```

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

Original Guide RNA:

```
(AEQVCTGTCGTTTATACCACTTTAAGAAAAAGTGCACCCAGCTCGGGCTTTT)
```

Optimized Chimeric Guide RNA Sequence 1:

```
(AEQVCTGTCGTTTATACCACTTTAAGAAAAAGTGCACCCAGCTCGGGCTTTT)
```

Optimized Chimeric Guide RNA Sequence 2:

```
(AEQVCTGTCGTTTATACCACTTTAAGAAAAAGTGCACCCAGCTCGGGCTTTT)
```

Optimized Chimeric Guide RNA Sequence 3:

```
(AEQVCTGTCGTTTATACCACTTTAAGAAAAAGTGCACCCAGCTCGGGCTTTT)
```

Applicants showed that optimized chimeric guide RNA works better as indicated in FIG. 3. The experiment was conducted by co-transfecting 293FT cells with Cas9 and a U6-guide RNA DNA cassette to express one of the four RNA forms shown above. The target of the guide RNA is the same target site in the human FmX1 locus: “GTCACCTCCAATGACTAGG (SEQ ID NO: 195)”
Example 7

Optimization of *Streptococcus Thermophilus* LMD-9 CRISPR-Cas9 (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-Cas9 systems are found throughout the bacterial kingdom and highly diverse 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 Kirk 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).

Example 9

**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 Technol-..)
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 ul cleavage reaction, 10 ul of cell lysate with incubated with 2 ul of cleavage buffer (100 mM HEPEs, 50 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 (schematic described in FIG. 73). PCR products were purified using EconoSpin 96-well Filter Plates (Epoch Life Sciences) following the manufacturer’s recommended protocol.

Barcoded and purified DNA samples were quantitated 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 score (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 alignments 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 μ-σ, where μ was the mean quality score of the negative control corresponding to that sample and σ was the standard deviation of same. This yielded whole target-region indel rates for each negative control 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)

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

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=E+Rp, 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

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

Taking all values of the frequency of target-regions with true-indels P to be equally probable a priori, \text{Prob}(n|p)\geq\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, Rp, and the number of reads R. Explicitly, the lower bound I and upper bound U were calculated as

\[
I = \frac{R \cdot p + z^2}{2} - z\sqrt{R \cdot p(1-p) + z^2/4}/(R+z^2)
\]

\[
U = \frac{R \cdot (1-p) + z^2}{2} + z\sqrt{R \cdot (1-p) + z^2/4}/(R+z^2)
\]

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 mirNEasy 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 Muster 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>
<thead>
<tr>
<th>Target site sequences</th>
<th>Target site sequence (5 to 3' .)</th>
<th>PAM strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EMX1</td>
<td>GTCACCTCTGAAGATGACGAGG (SEQ ID NO: 319)</td>
</tr>
<tr>
<td>2</td>
<td>EMX1</td>
<td>GACATCGACGTCCCTCCCAG (SEQ ID NO: 396)</td>
</tr>
</tbody>
</table>
### Table J-continued

<table>
<thead>
<tr>
<th>Primer name</th>
<th>genomic target sequence (5' to 3')</th>
<th>primer sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp-BMX1-R1</td>
<td>AAACACCCCTTTCTCTGGCC</td>
<td>(SEQ ID NO: 230)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCACTTTCTTCTGGCC</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 235)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCACTTTCTTCTGGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- qRT-PCR for Cas9 and sgRNA expression.
- Bisulfite PCR and sequencing.

### Table K

<table>
<thead>
<tr>
<th>primer name</th>
<th>primer sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(SEQ ID NO: 229)</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 230)</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 235)</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 236)</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO: 237)</td>
</tr>
</tbody>
</table>

- primers for testing sgRNA architecture.
### TABLE L

**Target sites with alternate PAMs for testing PAM specificity of Cas9.**

<table>
<thead>
<tr>
<th>Target site sequence (5' to 3')</th>
<th>PAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGGCCCCGATGGTCTGCTT (SEQ ID NO: 234)</td>
<td>NAA</td>
</tr>
<tr>
<td>AACATCAACCGTGACTGACAT (SEQ ID NO: 235)</td>
<td>NAT</td>
</tr>
<tr>
<td>AAGGCGTGCTGGCCAAGCC (SEQ ID NO: 236)</td>
<td>NAC</td>
</tr>
<tr>
<td>GGGAAATGCGAGGAGAAG (SEQ ID NO: 238)</td>
<td>NTA</td>
</tr>
<tr>
<td>CAGCTGTCGAGAAGAAG (SEQ ID NO: 239)</td>
<td>TNT</td>
</tr>
<tr>
<td>GGTGTTGCTGACAACGC (SEQ ID NO: 240)</td>
<td>NTC</td>
</tr>
<tr>
<td>ACCGGAGGAAAGGACCA (SEQ ID NO: 241)</td>
<td>NTC</td>
</tr>
<tr>
<td>TGGTTGCTGAGAAGAAG (SEQ ID NO: 242)</td>
<td>NTC</td>
</tr>
<tr>
<td>TCCGTCGAGAAGAAG (SEQ ID NO: 243)</td>
<td>NTA</td>
</tr>
<tr>
<td>TCAGTCGAGAAGAAG (SEQ ID NO: 244)</td>
<td>NTA</td>
</tr>
<tr>
<td>CAGAAGCTGAGAAGAAG (SEQ ID NO: 245)</td>
<td>NCA</td>
</tr>
<tr>
<td>CATCAGCGGAGAGGACCT (SEQ ID NO: 246)</td>
<td>NTA</td>
</tr>
<tr>
<td>GCGAGACGCTGAGAAG (SEQ ID NO: 247)</td>
<td>NTA</td>
</tr>
<tr>
<td>CACCCGCTGAGAAGAAG (SEQ ID NO: 248)</td>
<td>NTA</td>
</tr>
<tr>
<td>TCTCCTGCTGCTCCTCCT (SEQ ID NO: 249)</td>
<td>NAA</td>
</tr>
<tr>
<td>GGGAAGCAGTGGTGGAC (SEQ ID NO: 250)</td>
<td>NAT</td>
</tr>
<tr>
<td>CAGCGCGAAGCTGAGAAG (SEQ ID NO: 251)</td>
<td>NAC</td>
</tr>
<tr>
<td>GTGGGAGCAGCAGGAC (SEQ ID NO: 252)</td>
<td>NAG</td>
</tr>
<tr>
<td>GGGCGGGCGAAGCAGGAC (SEQ ID NO: 253)</td>
<td>NTA</td>
</tr>
<tr>
<td>GCCTCCAGACGAGAAG (SEQ ID NO: 254)</td>
<td>NTT</td>
</tr>
<tr>
<td>GAGGGGCTGAGTGGAG (SEQ ID NO: 255)</td>
<td>NTC</td>
</tr>
</tbody>
</table>

### Example 10

**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 SF370)*
  
  >U6-Short tracrRNA (Streptococcus Pyogenes SF370) *(seq ID NO: 40)*
  
  gaggctccattttccccatgccacttctatttgcatacagatacaacggtttgtagatgatatgtaattggtaaatag tacacaaatctgagcagtagaagtaatattttatggatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttgatgttg... (seq ID NO: 54)
20220919 started

>Sequencing Amplicon for EMX1 Guides 1.1, 1.14, 1.17

(SEQ ID NO: 264)

>Sequencing Amplicon for EMX1 Guides 1.2, 1.16

(SEQ ID NO: 265)

>Sequencing Amplicon for EMX1 Guides 1.3, 1.13, 1.15

(SEQ ID NO: 266)
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).

293FT cells were used. SpCas9—Wildtype Cas9 and SpCas9—nickase Cas9 (D10A). The chimeric RNA target is the same EMX1 Protoscaler Target 1 as in Examples 5, 9 and 10 and oligos synthesized by IDT using PAGE purification.

FIG. 44 depicts a design of the oligo DNA used as Homologous Recombination (HR) template in this experiment. Long oligos contain 100 bp homology to the EMX1 locus and a HindIII restriction site. 293FT cells were co-transfected with: first, a plasmid containing a chimeric RNA targeting human EMX1 locus and wild-type cas9 protein, and second, the oligo DNA as HR template. Samples are from 293FT cells collected 96 hours post transfection with Lipofectamine 2000. All products were amplified with an EMX1 HR Primer, gel purified, followed by digestion with HindIII to detect the efficiency of integration of HR template into the human genome.

FIGS. 45 and 46 depict a comparison of HR efficiency induced by different combination of Cas9 protein and HR template. The Cas9 construct used were either wild-type Cas9 or the nickase version of Cas9 (Cas9n). The HR template used were: antisense oligo DNA (Antisense-Oligo in above figure), or sense oligo DNA (Sense-Oligo in above figure), or plasmid HR template (HR template in above figure). The sense/antisense definition is that the actively-transcribed strand with sequence corresponding to the transcribed mRNA is defined as the sense strand of genome. HR Efficiency is shown as percentage of HindIII digestion band as against all genomic PCR amplified product (bottom numbers).

Example 12

Autistic Mouse

Recent large-scale sequencing initiatives have produced a large number of genes associated with disease. Discovering the genes is only the beginning in understanding what the gene does and how it leads to a diseased phenotype. Current technologies and approaches to study candidate genes are slow and laborious. The gold standards, gene targeting and genetic knockouts, require a significant investment in time and resources, both monetary and in terms of research personnel. Applicants set out to utilize the hSpCas9 nuclease to target many genes and do so with higher efficiency and lower turnaround compared to any other technology. Because of the high efficiency of hSpCas9 Applicants can do RNA injection into mouse zygotes and immediately get genome-modified animals without the need to do any preliminary gene targeting in mESCs.

Chromodomain helicase DNA binding protein 8 (CHD8) is a pivotal gene in involved in early vertebrate development and morphogenesis. Mice lacking CHD8 die during embryonic development. Mutations in the CHD8 gene have been associated with autism spectrum disorder in humans. This association was made in three different papers published simultaneously in Nature. The same three studies identified a plethora of genes associated with autism spectrum disorder. Applicants’ aim was to create knockout mice for the four genes that were found in all papers, Chd8, Katnal2, Kctd13, and Sce2a. In addition, Applicants chose two other genes
associated with autism spectrum disorder, schizophrenia, and ADHD, GTF1, CACNA1C, and CACNB2. And finally, as a positive control Applicants decide to target MeCP2.

For each gene Applicants designed three gRNAs that would likely knock out the gene. A knockout would occur after the hSpCas9 nuclease makes a double strand break and the error prone DNA repair pathway, non-homologous end joining, corrects the break, creating a mutation. The most likely result is a frameshift mutation that would knock out the gene. The targeting strategy involved finding proto-spacers in the exons of the gene that had a PAM sequence, NNG, 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, NeuroN2a, 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 controls 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 mutation. 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 the most unique within the genome.

RNA (hSpCas9/gRNA RNA) was injected into the pronucleus 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 SURVEYOR. Additionally, PCR products were sent for sequencing. 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 Ch8 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 (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), operably linked to triple-NLS and a VP64 functional domain is expressed by an 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 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-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) EF1a 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 72 h 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-amplified and subjected to the Surveyor assay following manufacturer’s protocol. The genomic cleavage efficiency of different constructs were measured using SBS-PAGE on a 4-12% TBE-PAGE gel (Life Technologies), analyzed and quantified with ImageLab (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 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.

<table>
<thead>
<tr>
<th>Percentage Genome Cleavage as measured by Surveyor assay</th>
<th>Biological Replicate 1 (%)</th>
<th>Biological Replicate 2 (%)</th>
<th>Biological Replicate 3 (%)</th>
<th>Average (%)</th>
<th>Error (S.E.M., standard error of the mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas9 (No NLS)</td>
<td>2.50</td>
<td>3.30</td>
<td>2.73</td>
<td>2.84</td>
<td>0.24</td>
</tr>
</tbody>
</table>
FIG. 70 depicts 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.

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 polymerase using vectors that expresses Cas9 and T7 polymerase 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 promoter driving the guide RNA.

Method 3: Applicants deliver Cas9 mRNA and in vitro transcribed guide RNA to algal 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 Cop 1.

(SEQ ID NO: 275)

TCTTCTTCTCGACATGACACTCCAGCAAAAGGTAGGCGGCGCTCCGTAAGA
CGGCCTCGGGGCCTCCATGCACACGCGATGATCTTCCAGGCCACCCGGAA
CTCCTCTCGGGGCCTCCATGCGGCATGATTGCCTGGCCGTCCACCCGGGAGTGGT
CTAAATAGGCACGGCCCTGAATCCAATGATAGCTACCTCAACAGGCACCTCGAGG
CTTGAGATGGCATCAGTCTGCAAGGGCGCTCTTCCCTTCCTTTTACGCFC
ACACCGCGAAACATGCTACATCGACTGATCTGCAAGGATCGCGCTCACG
AGGAAACGCGGATGCGCACTGCTGCAACTGCAGTACAGTCGCGCTG
GACTCGCCGACCAACTCCTGCGGCGCTGTCGATACGGGCAAGATGCAA
GGTTCGACAGGAAATTGAAAGGCTGTCGACAACAGCGGCGCGAGCA
TCGAGAAAGACACTCGGACAGGCACTCCGTGTCGACAGCGGCGGAAAGCGC
Sequence of guide RNA driven by the T7 promoter (T7 promoter, Ns represent targeting sequence):

SEQ ID NO: 277

Gene Delivery:
Chlamydomonas reinhardtii strain CC-124 and CC-125 from the Chlamydomonas Resource Center will be used for electroporation. Electroporation protocol 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 pChlamy (linearized using PvuI) and selecting for hygromycin resistant colonies. Sequence for pChlamy 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.

pChlamy-Cas9:

SEQ ID NO: 278
For all modified Clamabdomonas rehaindii cells, Applicants used PCR, SURVEYOR nuclease assay, and DNA sequencing to verify successful modification.

**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 *Streptococcus pyogenes* SF370 can be used to direct genome editing in *Streptococcus pneumoniae*. In this study Applicants engineered the crR8K strain containing a minimal CRISPR system, consisting of cas9, the tracrRNA
and a repeat. The D10A-I1840 mutations were introduced into Cas9 in this strain, giving strain cr68r**. Four spacers targeting different positions of the bgA β-galactosidase gene promoter were cloned in the CRISPR array carried by the previously described pDDB9 plasmid. Applicants observed a 10x to 20x reduction in β-galactosidase activity depending on the targeted position, demonstrating the potential of Cas9 as a programmable repressor (FIG. 73).

To achieve Cas9** repression in 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-I1840 mutations into pCas9, a plasmid described 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 RNAAP 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 were targeted, suggesting that Cas9 binding is actually strong enough to represent an obstacle to the running RNAAP. 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. 2B; 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 promoter region (T5 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 dgrNA::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 increased 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 (NOG) 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 mureg gene. Such off-targeting could easily be avoided by a systematic blast of the designed spacers.

Aspects of the invention are further described in the following numbered paragraphs:

1. A vector system comprising one or more vectors, wherein the system comprises:
   a. A first regulatory element operatively linked to a target sequence and one or more insertion sites for inserting a guide sequence upstream of the target 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 target sequence that is hybridized to the target sequence; and
   b. A second regulatory element operatively linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nucleic localization sequence;
   wherein components (a) and (b) are located on the same or different vectors of the system.

2. The vector system of paragraph 1, wherein component (a) further comprises the target sequence downstream of the target sequence under the control of the first regulatory element.

3. The vector system of paragraph 1, wherein 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 sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell.

4. The vector system of paragraph 1, wherein the system comprises the target sequence under the control of a third regulatory element.

5. The vector system of paragraph 1, wherein the target sequence exhibits at least 50% sequence complementarity along the length of the target sequence when optimally aligned.

6. The vector system of paragraph 1, wherein 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.

7. The vector system of paragraph 1, wherein the CRISPR enzyme is a type I/II CRISPR system enzyme.

8. The vector system of paragraph 1, wherein the CRISPR enzyme is a Cas9 enzyme.

9. The vector system of paragraph 1, wherein the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell.
10. The vector system of paragraph 1, wherein the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence.

11. The vector system of paragraph 1, wherein the CRISPR enzyme lacks DNA strand cleavage activity.

12. The vector system of paragraph 1, wherein the first regulatory element is a polymerase II promoter.

13. The vector system of paragraph 1, wherein the second regulatory element is a polymerase II promoter.

14. The vector system of paragraph 4, wherein the third regulatory element is a polymerase III promoter.

15. The vector system of paragraph 1, wherein the guide sequence is at least 15 nucleotides in length.

16. The vector system of paragraph 1, wherein fewer than 50% of the nucleotides of the guide sequence participate in self-complementary base-pairing when optimally folded.

17. A vector comprising a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising one or more nuclear localization sequences, wherein said regulatory element drives transcription of the CRISPR enzyme in an eukaryotic cell such that said CRISPR enzyme accumulates in a detectable amount in the nucleus of the eukaryotic cell.

18. The vector of paragraph 17, wherein said regulatory element is a polymerase II promoter.

19. The vector of paragraph 17, wherein said CRISPR enzyme is a type ICRISPR system enzyme.

20. The vector of paragraph 17, wherein said CRISPR enzyme is a Cas9 enzyme.

21. The vector of paragraph 17, wherein said CRISPR enzyme lacks the ability to cleave one or more strands of a target sequence to which it binds.

22. 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.

23. The CRISPR enzyme of paragraph 22, wherein said CRISPR enzyme is a type ICRISPR system enzyme.

24. The CRISPR enzyme of paragraph 22, wherein said CRISPR enzyme is a Cas9 enzyme.

25. The CRISPR enzyme of paragraph 22, wherein said CRISPR enzyme lacks the ability to cleave one or more strands of a target sequence to which it binds.

26. A eukaryotic host cell comprising:

a. a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting a guide sequence 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.

27. The eukaryotic host cell of paragraph 26, wherein said host cell comprises components (a) and (b).

28. The eukaryotic host cell of paragraph 26, wherein component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell.

29. The eukaryotic host cell of paragraph 26, wherein component (a) further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element.

30. The eukaryotic host cell of paragraph 26, wherein 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 directs sequence-specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell.

31. The eukaryotic host cell of paragraph 26, further comprising a third regulatory element operably linked to said tracr sequence.

32. The eukaryotic host cell of paragraph 26, wherein the tracr sequence exhibits at least 50% of sequence complementarity along the length of the tracr mate sequence when optimally aligned.

33. The eukaryotic host cell of paragraph 26, wherein 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.

34. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme is a type II CRISPR system enzyme.

35. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme is a Cas9 enzyme.

36. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell.

37. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence.

38. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme lacks DNA strand cleavage activity.

39. The eukaryotic host cell of paragraph 26, wherein the first regulatory element is a polymerase III promoter.

40. The eukaryotic host cell of paragraph 26, wherein the second regulatory element is a polymerase II promoter.

41. The eukaryotic host cell of paragraph 31, wherein the third regulatory element is a polymerase III promoter.

42. The eukaryotic host cell of paragraph 26, wherein the guide sequence is at least 15 nucleotides in length.

43. The eukaryotic host cell of paragraph 26, wherein fewer than 50% of the nucleotides of the guide sequence participate in self-complementary base-pairing when optimally folded.

44. A non-human animal comprising a eukaryotic host cell of any one of paragraphs 26-43.

45. A kit comprising a vector system and instructions for using said kit, the vector system comprising:

a. a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting a guide sequence 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.

46. The kit of paragraph 45, wherein said kit comprises components (a) and (b) located on the same or different vectors of the system.

47. The kit of paragraph 45, wherein component (a) further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element.

48. The kit of paragraph 45, wherein 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 sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell.

49. The kit of paragraph 45, wherein the system comprises the truer sequence under the control of a third regulatory element.

50. The kit of paragraph 45, wherein the truer sequence exhibits at least 50% of sequence complementarity along the length of the truer mate sequence when optimally aligned.

51. The kit of paragraph 45, wherein the CRISPR enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable mount in the nucleus of a eukaryotic cell.

52. The kit of paragraph 45, wherein the CRISPR enzyme is a type II CRISPR system enzyme.

53. The kit of paragraph 45, wherein the CRISPR enzyme is a Cas9 enzyme.

54. The kit of paragraph 45, wherein the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell.

55. The kit of paragraph 45, wherein the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence.

56. The kit of paragraph 45, wherein the CRISPR enzyme lacks DNA strand cleavage activity.

57. The kit of paragraph 45, wherein the first regulatory element is a polymerase III promoter.

58. The kit of paragraph 45, wherein the second regulatory element is a polymerase II promoter.

59. The kit of paragraph 49, wherein the third regulatory element is a polymerase III promoter.

60. The kit of paragraph 45, wherein the guide sequence is at least 15 nucleotides in length.

61. The kit of paragraph 45, wherein fewer than 50% of the nucleotides of the guide sequence participate in self-complementary base-pairing when optimally folded.

62. A computer system for selecting a candidate target sequence within a nucleic acid sequence in a eukaryotic cell for targeting by a CRISPR complex, the system comprising:
   a. a memory unit configured to receive and/or store said nucleic acid sequence; and
   b. one or more processors alone or in combination programmed to (i) locate a CRISPR motif sequence within said nucleic acid sequence, and (ii) select a sequence adjacent to said located CRISPR motif sequence as the candidate target sequence to which the CRISPR complex binds.

63. The computer system of paragraph 62, wherein said locating step comprises identifying a CRISPR motif sequence located less than about 500 nucleotides away from said target sequence.

64. The computer system of paragraph 62, wherein said candidate target sequence is at least 10 nucleotides in length.

65. The computer system of paragraph 62, wherein the nucleotide at the 3' end of the candidate target sequence is located no more than about 10 nucleotides upstream of the CRISPR motif sequence.

66. The computer system of paragraph 62, wherein the nucleic acid sequence in the eukaryotic cell is endogenous to the eukaryotic genome.

67. The computer system of claim 62, wherein the nucleic acid sequence in the eukaryotic cell is exogenous to the eukaryotic genome.

68. A computer-readable medium comprising codes that, upon execution by one or more processors, implements a method of selecting a candidate target sequence within a nucleic acid sequence in a eukaryotic cell for targeting by a CRISPR complex, said method comprising: (a) locating a CRISPR motif sequence within said nucleic acid sequence, and (b) selecting a sequence adjacent to said located CRISPR motif sequence as the candidate target sequence to which the CRISPR complex binds.

69. The computer-readable medium of paragraph 68, wherein said locating comprises locating a CRISPR motif sequence that is less than about 500 nucleotides away from said target sequence.

70. The computer-readable medium of paragraph 68, wherein said candidate target sequence is at least 10 nucleotides in length.

71. The computer-readable medium of paragraph 68, wherein said nucleotide at the 3' end of the candidate target sequence is located no more than about 10 nucleotides upstream of the CRISPR motif sequence.

72. The computer-readable medium of paragraph 68, wherein the nucleic acid sequence in the eukaryotic cell is endogenous to the eukaryotic genome.

73. The computer-readable medium of paragraph 68, wherein the nucleic acid sequence in the eukaryotic cell is exogenous to the eukaryotic genome.

74. A method of modifying a target polynucleotide in a eukaryotic cell, the method comprising 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 truer mate sequence which in turn hybridizes to a truer sequence.

75. The method of paragraph 74, wherein said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme.

76. The method of paragraph 74, wherein said cleavage results in decreased transcription of a target gene.

77. The method of paragraph 74, further comprising 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.

78. The method of paragraph 77, wherein said mutation results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence.

79. The method of paragraph 74, further comprising 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 truer mate sequence, and the truer sequence.

80. The method of paragraph 79, wherein said vectors are delivered to the eukaryotic cell in a subject.

81. The method of paragraph 74, wherein said modifying takes place in said eukaryotic cell in a cell culture.

82. The method of paragraph 74, further comprising isolating said eukaryotic cell from a subject prior to said modifying.

83. The method of paragraph 82, further comprising returning said eukaryotic cell and/or cells derived therefrom to said subject.

84. A method of modifying expression of a polynucleotide in a eukaryotic cell, the method comprising: allowing a CRISPR complex to bind to the polynucleotide such that said 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 truer mate sequence which in turn hybridizes to a truer sequence.
85. The method of paragraph 74, further comprising 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.

86. A method of generating a model eukaryotic cell comprising a mutated disease gene, the method comprising:
   a. introducting one or more vectors into a eukaryotic cell, 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.

87. The method of paragraph 86, wherein said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme.

88. The method of paragraph 86, wherein said cleavage results in decreased transcription of a target gene.

89. The method of paragraph 86, further comprising 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.

90. The method of paragraph 89, wherein said mutation results in or more amino acid changes in a protein expressed from a gene comprising the target sequence.

91. A method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene, comprising:
   a. contacting a test compound with a model cell of any one of paragraphs 86-90; and
   b. 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.

92. 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.

93. The recombinant polynucleotide of paragraph 89, wherein the target sequence is a viral sequence present in a eukaryotic cell.

94. The recombinant polynucleotide of paragraph 89, wherein the target sequence is a proto-oncogene or an oncogene.

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 alternatives 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


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

<210> SEQ ID NO 3
<211> LENGTH: 9
<212> TYPE: PRT
<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 Arg 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
Arg Gln Ser Ser Arg 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: ISE 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 Lys Ala 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
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Description of Unknown:
Myoma T protein sequence
<400> SEQUENCE: 8
Pro Pro Lys Lys Ala Arg Glu Asp
1  5

<210> SEQ ID NO 9
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 9
Pro Gln Pro Lys Lys Lys Pro Leu
1  5

<210> SEQ ID NO 10
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<400> SEQUENCE: 10
Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1  5  10

<210> SEQ ID NO 11
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Influenza virus
<400> SEQUENCE: 11
Asp Arg Leu Arg Arg
1  5

<210> SEQ ID NO 12
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Influenza virus
<400> SEQUENCE: 12
Pro Lys Gln Lys Lys Lys Arg Lys
1  5

<210> SEQ ID NO 13
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Hepatitis delta virus
<400> SEQUENCE: 13
Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1  5  10

<210> SEQ ID NO 14
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<400> SEQUENCE: 14
Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1  5  10

<210> SEQ ID NO 15
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 15
Lys Arg Lys G2y Acp Glu Val Acp Gly Val Acp Glu Val Ala Lys Lys
1  5  10  15
Lys Ser Lys Lys
20

<210> SEQ ID NO 16
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 16
Arg Lys Cys Leu Glu Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1  5  10  15
Lys

<210> SEQ ID NO 17
<211> LENGTH: 27
<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)...(20)
<223> OTHER INFORMATION: a, c, t or g
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)...(22)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 17
nnnnnnnnnnnnnnnn nnengaaw
27

<210> SEQ ID NO 18
<211> LENGTH: 19
<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)...(12)
<223> OTHER INFORMATION: a, c, t or g
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)...(14)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 18
nnnnnnnnnnnnnnnn nnengaaw
19

<210> SEQ ID NO 19
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

FEATURE:
NAME/KEY: modified_base
LOCATION: (1) . . (20)
OTHER INFORMATION: a, c, t or g

FEATURE:
NAME/KEY: modified_base
LOCATION: (21) . . (22)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 19

nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

FEATURE:
NAME/KEY: modified_base
LOCATION: (1) . . (11)
OTHER INFORMATION: a, c, t or g

FEATURE:
NAME/KEY: modified_base
LOCATION: (12) . . (13)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 20

nnnnnnnnnnnnnnnn

FEATURE:
NAME/KEY: modified_base
LOCATION: (1) . . (20)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 21

nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

gttttgtgc tcctcaagtt tagaataaa tcctgcaas

gtttataaatg ttttttttta attttttttt

tttt

FEATURE:
NAME/KEY: modified_base
LOCATION: (1) . . (20)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 22

nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

gttttgtgc tcctcagaaat gaacaagcata caagataag

gtttcatgcc gaatctcaca ccoctgtcatt ttattgcagg gttttttoct tatttatttt
<210> SEQ ID NO 23
<211> LENGTH: 110
<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
nnnnnnnnnnnnnnn nn ttttggta c tttgagaa at gcaaga gta caagataag  60
gtttactgc gaatacaaca cccctgtcatt ttatggcagg gttttttttttt 110

<210> SEQ ID NO 24
<211> LENGTH: 162
<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: 24
nnnnnnnnnnnnnnn nnnn ttttggagc tagaaatagc aagtttaa at aaggttagtc 60
cggtataac ttgaaaaagtt ggcacagagtt cgggtctttt tt 102

<210> SEQ ID NO 25
<211> LENGTH: 86
<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)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 25
nnnnnnnnnnnnnnnnnnn ttttggagc tagaaatagc aagtttaa at aaggttagtc 60
cggtataac ttgaaaaagtt gttttttttttt 88

<210> SEQ ID NO 26
<211> LENGTH: 76
<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)...(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 26
nnnnnnnnnnnnnnnnnnnnn ttttggagc tagaaatagc aagtttaa at aaggttagtc 60
cggtatacatt tttttttttttttt 76

<210> SEQ ID NO 27
<211> LENGTH: 12
<212> TYPE: DNA
ORGANISM: Artificial Sequence

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

SEQUENCE: 27
gttttagacg ta

SEQ ID NO 28
LENGTH: 31
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 28
tagcaagtttaaaataagctagtcggtttt t

SEQ ID NO 29
LENGTH: 27
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
FEATURE:
NAME/KEY: modified_base
LOCATION: (1), (22)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 29
nnnnnnnnnnnnnnnnnn nnagaw

SEQ ID NO 30
LENGTH: 12
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 30
guuuaagc ua

SEQ ID NO 31
LENGTH: 33
TYPE: DNA
ORGANISM: Homo sapiens

SEQUENCE: 31
ggacagctgtcacatcactaggttgg

SEQ ID NO 32
LENGTH: 33
TYPE: DNA
ORGANISM: Homo sapiens

SEQUENCE: 32
cattgaaggtgacatcgatgtcctcccat tgg

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

SEQUENCE: 33
ggaagggcct gagtcgcgac agaagaagaa ggg

<210> SEQ ID NO: 34
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 34

ggtgccgaga ggaggccgaga ttggtgttc agg

<210> SEQ ID NO: 35
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 35

atgcagaggg gtgcgcagag ggaggccgagat tgg

<210> SEQ ID NO: 36
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 36

aaaacccacct ttcctctgg c

<210> SEQ ID NO: 37
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 37

ggagattgga gacaagagaa g

<210> SEQ ID NO: 38
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 38

cggagaagcc aatgcgtgac

<210> SEQ ID NO: 39
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 39

ggcagcaaac tccttgctct

<210> SEQ ID NO: 40
<211> LENGTH: 335
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 40

gaggggcat gtccctgat tcctctatat ttgcataac accaaggg tagttagag 60
taatggaa ttaatttgc acgttaaacac aagatattag tacaacaaat acgtacgtga 120
aatgtaatat ttctgtggtta acgttaagtt ttttaattat gtttcacatt gaggtctatct 180
atgcttaccc taaatgagga gatttctgtc taatctgttgct ttatatatctt gttgaagag 240
cgaaaccacc cgacacttta accaagcg atcgattata ataagctgtag ttctatattac 300
acagattgaa gattcagcgg accttgcgtt tttttt 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

gaggggcat gtccctgat tcctctatat ttgcataac accaaggg tagttagag 60
ntaatggaa ttaatttgc acgttaaacac aagatattag tacaacaaat acgtacgtga 120
aatgtaatat ttctgtggtta acgttaagtt ttttaattat gtttcacatt gaggtctatct 180
atgcttaccc taaatgagga gatttctgtc taatctgttgct ttatatatctt gttgaagag 240
cgaaaccacc cgacacttta accaagcg atcgattata ataagctgtag ttctatattac 300
acagattgaa gattcagcgg accttgcgtt tttttt 423

<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

gaggggcat gtccctgat tcctctatat ttgcataac accaaggg tagttagag 60
ntaatggaa ttaatttgc acgttaaacac aagatattag tacaacaaat acgtacgtga 120
aatgtaatat ttctgtggtta acgttaagtt ttttaattat gtttcacatt gaggtctatct 180
atgcttaccc taaatgagga gatttctgtc taatctgttgct ttatatatctt gttgaagag 240
cgaaaccacc cgacacttta accaagcg atcgattata ataagctgtag ttctatattac 300
acagattgaa gattcagcgg accttgcgtt tttttt 339

<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
gagggccat ttcctcagat tctttcatat ttcgatatat gataacaaggg tggtagagag 60
atataggaat ttaattttgac tgaasacac caagataattag tacaattataac gtgacgttaga 120
aagtaaaat ttccttggta gtttgcagt ttaaaattat gttttaaaat ggaatcatctat 180
atgcttaccc taacttggaa gtattctcgt tctttgcct tatatatott gtgggaaagga 240
cgaasaccc aggctcctgag aagacctgttt tttagagctag aasatacaag tttaaatag 300
gctagcgcg 309

<210> SEQ ID NO 44
<211> LENGTH: 1648
<220> TYPE: PRT
<230> ORGANISM: Artificial Sequence
<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 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 Phe Asp Ser Gly Glu
85 90
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
115 120 125
Met Ala Lys Val Asp Asp Ser Phe His Arg Leu Glu Glu Ser Phe
130 135 140
Leu Val Glu Asp Lys 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 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 Leu Phe
210 215 220
Ile Gin Leu Val Gel Thr Tyr Asn Gin Leu Phe Glu Gin Gin 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 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
275 280 285
Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gin
<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Start Position</th>
<th>End Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly</td>
<td>1130</td>
<td>1140</td>
</tr>
<tr>
<td>Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu</td>
<td>1145</td>
<td>1155</td>
</tr>
<tr>
<td>Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe</td>
<td>1160</td>
<td>1170</td>
</tr>
<tr>
<td>Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val</td>
<td>1175</td>
<td>1185</td>
</tr>
<tr>
<td>Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu</td>
<td>1190</td>
<td>1200</td>
</tr>
<tr>
<td>Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile</td>
<td>1205</td>
<td>1215</td>
</tr>
<tr>
<td>Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu</td>
<td>1220</td>
<td>1230</td>
</tr>
<tr>
<td>Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly</td>
<td>1235</td>
<td>1245</td>
</tr>
<tr>
<td>Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn</td>
<td>1250</td>
<td>1260</td>
</tr>
<tr>
<td>Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala</td>
<td>1265</td>
<td>1275</td>
</tr>
<tr>
<td>Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln</td>
<td>1280</td>
<td>1290</td>
</tr>
<tr>
<td>Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile</td>
<td>1295</td>
<td>1305</td>
</tr>
<tr>
<td>Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp</td>
<td>1310</td>
<td>1320</td>
</tr>
<tr>
<td>Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp</td>
<td>1325</td>
<td>1335</td>
</tr>
<tr>
<td>Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr</td>
<td>1340</td>
<td>1350</td>
</tr>
<tr>
<td>Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr</td>
<td>1355</td>
<td>1365</td>
</tr>
<tr>
<td>Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp</td>
<td>1370</td>
<td>1380</td>
</tr>
<tr>
<td>Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg</td>
<td>1385</td>
<td>1395</td>
</tr>
<tr>
<td>Ile Asp Leu Ser Glu Leu Gly Gly Asp Ala Ala Val Ser Lys</td>
<td>1400</td>
<td>1410</td>
</tr>
<tr>
<td>Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu</td>
<td>1415</td>
<td>1425</td>
</tr>
<tr>
<td>Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly</td>
<td>1430</td>
<td>1440</td>
</tr>
<tr>
<td>Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys</td>
<td>1445</td>
<td>1455</td>
</tr>
<tr>
<td>Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr</td>
<td>1460</td>
<td>1470</td>
</tr>
<tr>
<td>Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met</td>
<td>1475</td>
<td>1485</td>
</tr>
<tr>
<td>Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val</td>
<td>1490</td>
<td>1500</td>
</tr>
<tr>
<td>Gin Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr</td>
<td>1505</td>
<td>1515</td>
</tr>
<tr>
<td>Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile</td>
<td>1520</td>
<td>1530</td>
</tr>
</tbody>
</table>
169

Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
1520 1525 1530

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met
1535 1540 1545

Ala Asp Lys Gln Lys Asp Gly Ile Lys Val Asn Phe Lys Ile Arg
1550 1555 1560

His Asn Ile Glu Asp Gly Ser Val Glu Leu Ala Asp His Tyr Gln
1565 1570 1575

Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn
1580 1585 1590

His Tyr Leu Ser Thr Glu Ser Ala Leu Ser Lys Asp Pro Asn Glu
1595 1600 1605

Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly
1610 1615 1620

Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
1625 1630 1635

<210> SEQ ID NO 45
<211> LENGTH: 1625
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 45

Met Asp Lys Tyr Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Thr Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40

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 Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Glu Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
95 100 105 110

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
120 125

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
1 5 10 15

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Leu Val Asp
135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Ala His
145 150 155 160

Met Ile Lys Phe Arg G1y 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 Glu Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn 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
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>645</td>
<td>Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp</td>
<td>665</td>
<td>670</td>
</tr>
<tr>
<td>650</td>
<td>Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Arg Gly Phe</td>
<td>680</td>
<td>695</td>
</tr>
<tr>
<td></td>
<td>Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Ser Leu Thr Phe</td>
<td>690</td>
<td>705</td>
</tr>
<tr>
<td>655</td>
<td>Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu</td>
<td>710</td>
<td>725</td>
</tr>
<tr>
<td></td>
<td>His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly</td>
<td>730</td>
<td>735</td>
</tr>
<tr>
<td></td>
<td>Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly</td>
<td>740</td>
<td>745</td>
</tr>
<tr>
<td></td>
<td>Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln</td>
<td>750</td>
<td>755</td>
</tr>
<tr>
<td></td>
<td>Thr Thr Glu Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile</td>
<td>760</td>
<td>765</td>
</tr>
<tr>
<td>670</td>
<td>Glu Glu Gly Ile Lys Glu Leu Gly Ser Gin Ile Leu Lys Glu His Pro</td>
<td>770</td>
<td>780</td>
</tr>
<tr>
<td></td>
<td>Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu</td>
<td>790</td>
<td>805</td>
</tr>
<tr>
<td></td>
<td>Gln Asn Gly Arg Asp Met Tyr Val Asp Gin Glu Leu Asp Ile Asn Arg</td>
<td>810</td>
<td>815</td>
</tr>
<tr>
<td></td>
<td>Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gin Ser Phe Leu Lys</td>
<td>820</td>
<td>825</td>
</tr>
<tr>
<td></td>
<td>Asp Asp Ser Ile Asp Gln Val Leu Thr Arg Ser Asp Lys Asn Arg</td>
<td>830</td>
<td>835</td>
</tr>
<tr>
<td></td>
<td>Gly Lys Ser Asp Met Gin Val Pro Ser Glu Gln Val Leu Lys Met Lys</td>
<td>840</td>
<td>845</td>
</tr>
<tr>
<td></td>
<td>Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys</td>
<td>850</td>
<td>855</td>
</tr>
<tr>
<td></td>
<td>Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp</td>
<td>860</td>
<td>865</td>
</tr>
<tr>
<td></td>
<td>Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr</td>
<td>870</td>
<td>875</td>
</tr>
<tr>
<td></td>
<td>Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp</td>
<td>880</td>
<td>885</td>
</tr>
<tr>
<td></td>
<td>Glu Asn Asp Lys Leu Ile Arg Glu Val Leu Ile Thr Leu Lys Ser</td>
<td>890</td>
<td>895</td>
</tr>
<tr>
<td></td>
<td>Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg</td>
<td>900</td>
<td>905</td>
</tr>
<tr>
<td></td>
<td>Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val</td>
<td>910</td>
<td>915</td>
</tr>
<tr>
<td>680</td>
<td>Val Gly Thr Ala Leu Ile Lys Tyr Pro Lys Leu Glu Ser Glu Phe</td>
<td>920</td>
<td>925</td>
</tr>
<tr>
<td>685</td>
<td>Lys Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala</td>
<td>930</td>
<td>935</td>
</tr>
<tr>
<td>690</td>
<td>Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe</td>
<td>940</td>
<td>945</td>
</tr>
<tr>
<td>695</td>
<td>Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala</td>
<td>950</td>
<td>955</td>
</tr>
<tr>
<td>700</td>
<td>Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu</td>
<td>960</td>
<td>965</td>
</tr>
<tr>
<td>Thr</td>
<td>Gly</td>
<td>Glu</td>
<td>Ile</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>1070</td>
<td>1075</td>
<td>1080</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>1089</td>
<td>1090</td>
<td>1095</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Gln</td>
<td>Thr</td>
</tr>
<tr>
<td>1100</td>
<td>1105</td>
<td>1110</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Asn</td>
<td>Ser</td>
<td>Asp</td>
</tr>
<tr>
<td>1115</td>
<td>1120</td>
<td>1125</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Lys</td>
<td>Tyr</td>
<td>Gly</td>
</tr>
<tr>
<td>1130</td>
<td>1135</td>
<td>1140</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>1145</td>
<td>1150</td>
<td>1155</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td>1160</td>
<td>1165</td>
<td>1170</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Glu</td>
<td>Lys</td>
<td>Asn</td>
</tr>
<tr>
<td>1175</td>
<td>1180</td>
<td>1185</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Lys</td>
<td>Lys</td>
</tr>
<tr>
<td>1190</td>
<td>1195</td>
<td>1200</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Glu</td>
<td>Leu</td>
<td>Glu</td>
</tr>
<tr>
<td>1205</td>
<td>1210</td>
<td>1215</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Gin</td>
<td>Lys</td>
</tr>
<tr>
<td>1220</td>
<td>1225</td>
<td>1230</td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Phe</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td>1235</td>
<td>1240</td>
<td>1245</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Glu</td>
<td>Asp</td>
<td>Asn</td>
</tr>
<tr>
<td>1250</td>
<td>1255</td>
<td>1260</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Tyr</td>
<td>Leu</td>
<td>Asp</td>
</tr>
<tr>
<td>1265</td>
<td>1270</td>
<td>1275</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
<td>Ile</td>
<td>Leu</td>
</tr>
<tr>
<td>1280</td>
<td>1285</td>
<td>1290</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Asn</td>
<td>Lys</td>
<td>His</td>
</tr>
<tr>
<td>1295</td>
<td>1300</td>
<td>1305</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Ile</td>
<td>His</td>
<td>Leu</td>
</tr>
<tr>
<td>1310</td>
<td>1315</td>
<td>1320</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Lys</td>
<td>Tyr</td>
<td>Phe</td>
</tr>
<tr>
<td>1325</td>
<td>1330</td>
<td>1335</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Lys</td>
<td>Glu</td>
<td>Val</td>
</tr>
<tr>
<td>1340</td>
<td>1345</td>
<td>1350</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Leu</td>
<td>Tyr</td>
<td>Glu</td>
</tr>
<tr>
<td>1355</td>
<td>1360</td>
<td>1365</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Ala</td>
<td>Val</td>
</tr>
<tr>
<td>1370</td>
<td>1375</td>
<td>1380</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Ile</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td>1385</td>
<td>1390</td>
<td>1395</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Ser</td>
<td>Gly</td>
</tr>
<tr>
<td>Thr</td>
<td>Leu</td>
<td>Lys</td>
<td>Phe</td>
</tr>
<tr>
<td>1415</td>
<td>1420</td>
<td>1425</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Thr</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td>Arg</td>
<td>Tyr</td>
<td>Pro</td>
<td>Asp</td>
</tr>
</tbody>
</table>
Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp
1460 1465 1470
Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp
1475 1480 1485
Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu
1490 1495 1500
Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser
1505 1510 1515
His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
1520 1525 1530
Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Glu
1535 1540 1545
Leu Ala Asp His Tyr Glu Asn Thr Pro Ile Gly Asp Gly Pro
1550 1555 1560
Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Glu Ser Ala Leu
1565 1570 1575
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu
1580 1585 1590
Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr
1595 1600 1605
Lys Lys Arg Pro Ala Ala Thr Lys Ala Gly Gln Ala Lys Lys
1610 1615 1620
Lys Lys
1625

<210> SEQ ID NO 46
<211> LENGTH: 1664
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<400> 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 Asp Lys Met Ala Pro Lys Lys Arg Lys Val
20 25 30
Gly Ile His Gly Val Pro 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 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 Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
130 135 140
Leu Val 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 Tyr Pro Thr Ile Tyr His
165 170 175
<p>| Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu | 180 | 190 |
| Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu | 195 | 205 |
| Ile Glu Gly Asp Leu Asn Pro Asp Ser Asp Val Asp Lys Leu Phe | 210 | 220 |
| Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile | 225 | 235 |
| Asn Ala Ser Gly Val Ala Asn Ala Ile Leu Ser Ala Arg Leu Ser | 245 | 255 |
| Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys | 260 | 270 |
| Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr | 275 | 285 |
| Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln | 290 | 300 |
| Leu Ser Lys Asp Thr Tyr Asp Asp Leu Asn Leu Leu Ala Gln | 305 | 320 |
| Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser | 325 | 335 |
| Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr | 340 | 350 |
| Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His | 355 | 365 |
| Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Glu Leu Pro Glu | 370 | 380 |
| Lys Tyr Lys Glu Ile Phe Phe Glu Glu Ser Lys Asn Gly Tyr Ala Gly | 385 | 400 |
| Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys | 405 | 415 |
| Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu | 420 | 430 |
| Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser | 435 | 445 |
| Ile Pro His Gln Ile His Leu Gly Leu His Ala Ile Leu Arg Arg | 450 | 460 |
| Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu | 465 | 480 |
| Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg | 485 | 495 |
| Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile | 500 | 510 |
| Thr Pro Trp Asn Phe Glu Val Val Asp Lys Gly Ala Ser Ala Gin | 515 | 525 |
| Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu | 530 | 540 |
| Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr | 545 | 560 |
| Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro | 565 | 575 |
| Ala Phe Leu Ser Gly Glu Gin Lys Lys Ala Ile Val Asp Leu Leu Phe | 580 | 590 |</p>
<table>
<thead>
<tr>
<th>1010</th>
<th>1015</th>
<th>1020</th>
</tr>
</thead>
<tbody>
<tr>
<td>His</td>
<td>Asp</td>
<td>Ala</td>
</tr>
<tr>
<td>1025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Tyr</td>
<td>Leu</td>
</tr>
<tr>
<td>1030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Asn</td>
<td>Ala</td>
</tr>
<tr>
<td>1035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Val</td>
<td>Gly</td>
</tr>
<tr>
<td>1040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Pro</td>
<td>Leu</td>
</tr>
<tr>
<td>1045</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td>1050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Phe</td>
<td>Val</td>
</tr>
<tr>
<td>1055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td>1060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Lys</td>
<td>Met</td>
</tr>
<tr>
<td>1065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Ala</td>
<td>Lys</td>
</tr>
<tr>
<td>1070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td>1075</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Pro</td>
<td>Tyr</td>
</tr>
<tr>
<td>1080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Lys</td>
<td>Phe</td>
</tr>
<tr>
<td>1085</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Lys</td>
<td>Arg</td>
</tr>
<tr>
<td>1090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Thr</td>
<td>Ala</td>
</tr>
<tr>
<td>1095</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Val</td>
<td>Arg</td>
</tr>
<tr>
<td>1100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Ile</td>
</tr>
<tr>
<td>1105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Gly</td>
<td>Thr</td>
</tr>
<tr>
<td>1110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Thr</td>
</tr>
<tr>
<td>1115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Gln</td>
<td>Val</td>
</tr>
<tr>
<td>1120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>1125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Pro</td>
<td>Lys</td>
</tr>
<tr>
<td>1130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Gly</td>
<td>Val</td>
</tr>
<tr>
<td>1135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ile</td>
<td>Val</td>
</tr>
<tr>
<td>1140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td>1145</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Glu</td>
<td>Ser</td>
</tr>
<tr>
<td>1150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td>1155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Arg</td>
<td>Asn</td>
</tr>
<tr>
<td>1160</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Thr</td>
<td>Ala</td>
</tr>
<tr>
<td>1170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Lys</td>
<td>Ser</td>
</tr>
<tr>
<td>1180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>1185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>1190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td>1195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td>1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td>1205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td>1210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Met</td>
<td>Glu</td>
</tr>
<tr>
<td>1215</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Ser</td>
</tr>
<tr>
<td>1220</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Glu</td>
</tr>
<tr>
<td>1225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Lys</td>
<td>Gly</td>
</tr>
<tr>
<td>1230</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td>1235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Pro</td>
<td>Lys</td>
</tr>
<tr>
<td>1240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Ser</td>
<td>Leu</td>
</tr>
<tr>
<td>1245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
<td>Glu</td>
</tr>
<tr>
<td>1250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td>1255</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Gly</td>
<td>Asn</td>
</tr>
<tr>
<td>1260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Met</td>
<td>Ala</td>
</tr>
<tr>
<td>1265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>1270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>Ser</td>
</tr>
<tr>
<td>1275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>His</td>
<td>Tyr</td>
</tr>
<tr>
<td>1280</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>1285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Ser</td>
<td>Pro</td>
</tr>
<tr>
<td>1290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Asp</td>
<td>Ala</td>
</tr>
<tr>
<td>1295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Tyr</td>
<td>Leu</td>
</tr>
<tr>
<td>1300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td>1305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>1310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Lys</td>
</tr>
<tr>
<td>1315</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
<td>Ile</td>
</tr>
<tr>
<td>1320</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Asn</td>
<td>Leu</td>
</tr>
<tr>
<td>1325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Lys</td>
<td>Val</td>
</tr>
<tr>
<td>1330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Tyr</td>
<td>Asn</td>
</tr>
<tr>
<td>1335</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ser</td>
<td>Ala</td>
</tr>
<tr>
<td>1340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>His</td>
<td>Leu</td>
</tr>
<tr>
<td>1345</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Phe</td>
<td>Thr</td>
</tr>
<tr>
<td>1350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Ile</td>
<td>Arg</td>
</tr>
<tr>
<td>1355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Glu</td>
<td>Ala</td>
</tr>
<tr>
<td>1360</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Ala</td>
<td>Lys</td>
</tr>
<tr>
<td>1365</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Phe</td>
</tr>
<tr>
<td>1370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Arg</td>
<td>Asp</td>
</tr>
<tr>
<td>1375</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Thr</td>
<td>Ser</td>
</tr>
<tr>
<td>1380</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Thr</td>
<td>Lys</td>
</tr>
<tr>
<td>1385</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>His</td>
</tr>
<tr>
<td>1390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td>1395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Ala</td>
<td>Ile</td>
</tr>
<tr>
<td>1400</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Gln</td>
<td>Leu</td>
</tr>
<tr>
<td>1405</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Asp</td>
<td>Ala</td>
</tr>
<tr>
<td>1410</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gin Glu Ile Phe Ser Asn Glu</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Met Ala Lys Val Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>Thr Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Leu Arg Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu</td>
<td>195</td>
<td></td>
</tr>
<tr>
<td>Ile Glu Gly Asp Leu Asn Pro Asp Ser Asp Val Asp Lys Leu Phe</td>
<td>210</td>
<td></td>
</tr>
<tr>
<td>Lys Gin Leu Val Gin Thr Tyr Asn Gin Leu Phe Glu Gin Asn Pro Ile</td>
<td>225</td>
<td></td>
</tr>
<tr>
<td>Asn Ala Ser Gin Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser</td>
<td>245</td>
<td></td>
</tr>
<tr>
<td>Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gin Leu Pro Gly Glu Lys</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr</td>
<td>275</td>
<td></td>
</tr>
<tr>
<td>Leu Ser Lys Asp Thr Tyr Asp Asp Leu Asn Leu Leu Ala Gin</td>
<td>305</td>
<td></td>
</tr>
<tr>
<td>Ile Gly Asp Gin Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser</td>
<td>325</td>
<td></td>
</tr>
<tr>
<td>Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr</td>
<td>340</td>
<td></td>
</tr>
<tr>
<td>Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His</td>
<td>355</td>
<td></td>
</tr>
<tr>
<td>Gin Asp Leu Thr Leu Lys Ala Leu Val Arg Gin Gin Leu Pro Glu</td>
<td>370</td>
<td></td>
</tr>
<tr>
<td>Lys Tyr Lys Glu Ile Phe Phe Asp Gin Ser Lys Asn Gly Tyr Ala Gly</td>
<td>395</td>
<td></td>
</tr>
<tr>
<td>Tyr Ile Asp Gly Gly Ala Ser Gin Glu Glu Phe Tyr Lys Phe Ile Lys</td>
<td>405</td>
<td></td>
</tr>
<tr>
<td>Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu</td>
<td>420</td>
<td></td>
</tr>
<tr>
<td>Arg Arg Glu Asp Leu Leu Arg Lys Gin Arg Thr Phe Asp Asn Gly Ser</td>
<td>435</td>
<td></td>
</tr>
<tr>
<td>Ile Pro His Gin Ile His Leu Gly Leu His Ala Ile Leu Arg Arg</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>Gin Glu Asp Phe Tyr Pro Phe Leu Lys Asn Arg Glu Lys Ile Glu</td>
<td>465</td>
<td></td>
</tr>
<tr>
<td>Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg</td>
<td>485</td>
<td></td>
</tr>
<tr>
<td>Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Pro</td>
<td>Trp</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Phe</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Thr</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Lys</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Phe</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asp</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Ile</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Arg</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gin</td>
<td>Leu</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ile</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Asp</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Gly</td>
<td>Gin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Pro</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Met</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Glu</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gin</td>
<td>Ile</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gin</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Pro</td>
<td>Gin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Arg</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg</td>
<td>930</td>
<td>940</td>
</tr>
<tr>
<td>Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu</td>
<td>945</td>
<td>950</td>
</tr>
<tr>
<td>Val Glu Thr Arg Gln Ile Thr His Val Ala Gln Ile Leu Asp Ser</td>
<td>965</td>
<td>970</td>
</tr>
<tr>
<td>Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val</td>
<td>980</td>
<td>985</td>
</tr>
<tr>
<td>Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp</td>
<td>995</td>
<td>1000</td>
</tr>
<tr>
<td>Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Arg Tyr His His Ala</td>
<td>1010</td>
<td>1015</td>
</tr>
<tr>
<td>His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys</td>
<td>1025</td>
<td>1030</td>
</tr>
<tr>
<td>Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys</td>
<td>1040</td>
<td>1045</td>
</tr>
<tr>
<td>Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Glu Glu Ile</td>
<td>1050</td>
<td>1060</td>
</tr>
<tr>
<td>Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn</td>
<td>1070</td>
<td>1075</td>
</tr>
<tr>
<td>Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys</td>
<td>1095</td>
<td>1090</td>
</tr>
<tr>
<td>Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp</td>
<td>1100</td>
<td>1105</td>
</tr>
<tr>
<td>Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met</td>
<td>1115</td>
<td>1120</td>
</tr>
<tr>
<td>Pro Gin Val Asn Ile Val Lys Lys Thr Glu Val Gin Thr Gly Gln</td>
<td>1130</td>
<td>1135</td>
</tr>
<tr>
<td>Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu</td>
<td>1145</td>
<td>1150</td>
</tr>
<tr>
<td>Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Tyr Gly Gly Phe</td>
<td>1160</td>
<td>1165</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>
</tr>
<tr>
<td>Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu</td>
<td>1190</td>
<td>1195</td>
</tr>
<tr>
<td>Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile</td>
<td>1205</td>
<td>1210</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>
</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>
</tr>
<tr>
<td>Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gin Lys Gly Asn</td>
<td>1250</td>
<td>1255</td>
</tr>
<tr>
<td>Glu Leu Ala Leu Pro Ser Lys Tyr Val Ann Phe Leu Tyr Leu Ala</td>
<td>1265</td>
<td>1270</td>
</tr>
<tr>
<td>Ser His Tyr Glu Leu Leu Gly Ser Pro Glu Asp Asn Glu Gin</td>
<td>1280</td>
<td>1285</td>
</tr>
<tr>
<td>Lys Glu Leu Phe Val Glu Gin His Lys His Tyr Leu Asp Glu Ile</td>
<td>1295</td>
<td>1300</td>
</tr>
<tr>
<td>Ile Glu Gin Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp</td>
<td>1310</td>
<td>1315</td>
</tr>
<tr>
<td>Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp</td>
<td>1320</td>
<td>1325</td>
</tr>
</tbody>
</table>
Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr
Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg
Ile Asp Leu Ser Gln Leu Gly Gly Asp Lys Arg Pro Ala Ala Thr
Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

<210> SEQ ID NO 48
<211> LENGTH: 483
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATREC:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 48

Met Phe Leu Phe Leu Ser Leu Ser Phe Leu Ser Ser Ser Arg Thr
1       5       10       15
Leu Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Lys Glu Phe
20      25      30
Met Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe
35      40      45
Glu Ile Glu Gly Gly Glu Gly Gly Arg Pro Tyr Glu Gly Thr Gln Thr
50      55      60
Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asep
65      70      75      80
Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His
85      90      95
Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe
100     105     110
Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val
115     120     125
Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys
130     135     140
Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys
145     150     155     160
Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly
165     170     175
Ala Leu Gly Gly Ile Lys Gln Arg Leu Lys Lys Asp Gly Gly
180     185     190
His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val
195     200     205
Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser
210     215     220
His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly
225     230     235     240
Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Ser Lys Gln
245     250     255
Leu Glu Glu Leu Leu Ser Thr Ser Phe Asp Ile Gln Phe Asn Asp Leu 260 265 270
Thr Leu Leu Glu Thr Ala Phe Thr His Thr Ser Tyr Ala Asn Glu His 275 280 285
Arg Leu Leu Asn Val Ser His Asn Glu Arg Leu Glu Phe Leu Gly Asp 290 295 300
Ala Val Leu Glu Leu Ile Ile Ser Glu Tyr Leu Phe Ala Lys Tyr Pro 305 310 315 320
Lys Lys Thr Glu Gly Asp Met Ser Lys Leu Arg Ser Met Ile Val Arg 325 330 335
Glu Glu Ser Leu Ala Gly Phe Ser Arg Phe Cys Ser Phe Asp Ala Tyr 340 345 350
Ile Lys Leu Gly Lys Gly Glu Lys Ser Gly Gly Arg Arg Arg Asp 355 360 365
Thr Ile Leu Gly Asp Leu Phe Ala Phe Leu Gly Ala Leu Leu Leu 370 375 380
Asp Lys Gly Ile Asp Ala Val Arg Arg Phe Leu Lys Gln Val Met Ile 385 390 395 400
Pro Gln Val Glu Lys Gly Asn Phe Glu Arg Val Lys Aep Tyr Lys Thr 405 410 415
Cys Leu Gln Glu Phe Leu Gln Thr Lys Gly Asp Val Ala Ile Asp Tyr 420 425 430
Gln Val Ile Ser Glu Lys Gly Pro Ala His Ala Lys Glu Phe Glu Val 435 440 445
Ser Ile Val Val Asn Gly Ala Val Leu Ser Lys Gly Leu Gly Lys Ser 450 455 460
Lys Lys Leu Ala Glu Gin Asp Ala Ala Lys Asn Ala Leu Ala Gin Leu 465 470 475 480
Ser Glu Val

<210> SEQ ID NO 49
<211> LENGTH: 483
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 49
Met Lys Gln Leu Glu Glu Leu Leu Ser Thr Ser Phe Asp Ile Gln Phe 1 5 10 15
Asn Asp Leu Thr Leu Leu Glu Thr Ala Phe Thr His Thr Ser Tyr Ala 20 25 30
Asn Glu His Arg Leu Leu Asn Val Ser His Asn Glu Arg Leu Glu Phe 35 40 45
Leu Gly Asp Ala Val Leu Gin Leu Ile Ile Ser Glu 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 85 90 95
Asp Ala Tyr Ile Lys Leu Gly Lys Gly Glu Gly Lys Ser Gly Gly Arg 100 105 110
Arg Arg Asp Thr Ile Leu Gly Asp Leu Phe Glu Ala Phe Leu Gly Ala 115 120 125
Leu Leu Leu Asp Lys Gly Ile Asp Ala Val Arg Arg Phe Leu Lys Gln 130 135 140
Val Met Ile Pro Gln Val Glu Lys Gly Asn Phe Glu Arg Val Lys Asp 145 150 155 160
Tyr Lys Thr Cys Leu Gln Glu Phe Leu Gin Thr Lys Gly Asp Val Ala 165 170 175
Ile Asp Tyr Gin Val Ile Ser Glu Lys Gly Pro Ala His Ala Lys Gin 180 185 190
Phe Glu Val Ser Ile Val Val Asn Gly Ala Val Leu Ser Lys Gly Leu 195 200 205
Gly Lys Ser Lys Lys Leu Ala Glu Gin Asp Ala Ala Lys Asn Ala Leu 210 215 220
Ala Gin Leu Ser Glu Val Gly Ser Val Ser Lys Gly Glu Glu Asp Asn 225 230 235 240
Met Ala Ile Ile Lys Glu Phe Met Arg Phe Lys Val His Met Glu Gly 245 250 255
Ser Val Asn Gly His Glu Phe Glu Ile Glu Gly Glu Gly Gly Arg 260 265 270
Pro Tyr Glu Gly Thr Gln Thr Ala Lys Leu Lys Val Thr Lys Gly Gly 275 280 285
Pro Leu Pro Phe Ala Trp Asp Ile Leu Ser Pro Gin Phe Met Tyr Gly 290 295 300
Ser Lys Ala Tyr Val Lys His Pro Ala Asp Ile Pro Asp Tyr Leu Lys 305 310 315 320
Leu Ser Phe Pro Glu Gly Phe Trp Glu Arg Val Met Asn Phe Glu 325 330 335
Asp Gly Gly Val Val Thr Val Thr Gln Asp Ser Ser Leu Gin Asp Gly 340 345 350
Glu Phe Ile Tyr Lys Val Lys Leu Arg Gly Thr Asn Phe Pro Ser Asp 355 360 365
Gly Pro Val Met Gin Lys Thr Met Gly Thr Gin Ala Ser Ser Glu 370 375 380
Arg Met Tyr Pro Glu Asp Gly Ala Leu Lys Gly Glu Ile Lys Gln Arg 395 390 395 400
Leu Lys Leu Lys Asp Gly Gly His Tyr Asp Ala Glu Val Lys Thr Thr 405 410 415
Tyr Lys Ala Lys Lys Pro Val Gin Leu Pro Gly Ala Tyr Asn Val Asn 420 425 430
Ile Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr Thr Ile Val Glu 435 440 445
Gln Tyr Glu Arg Ala Glu Gly Arg His Ser Thr Gly Gly Met Asp Glu 450 455 460
Leu Tyr Lys Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gin Ala Lys 465 470 475 480
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 polypeptide
<400> SEQUENCE: 50
<table>
<thead>
<tr>
<th>Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp</th>
<th>1 5 10 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr Lys Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val</td>
<td>20 25 30</td>
</tr>
<tr>
<td>Gly Ile His Gly Val Pro Ala Ala Asp Lys Tyr Ser Ile Gly Leu</td>
<td>35 40 45</td>
</tr>
<tr>
<td>Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr</td>
<td>50 55 60</td>
</tr>
<tr>
<td>Lys Val Pro Ser Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His</td>
<td>65 70 75 80</td>
</tr>
<tr>
<td>Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu</td>
<td>85 90 95</td>
</tr>
<tr>
<td>Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr</td>
<td>100 105 110</td>
</tr>
<tr>
<td>Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gin Glu Ile Phe Ser Asn Glu</td>
<td>115 120 125</td>
</tr>
<tr>
<td>Met Ala Lys Val Asp Asp Ser Phe His Arg Leu Glu Glu Ser Phe</td>
<td>130 135 140</td>
</tr>
<tr>
<td>Leu Val Glu Glu Arg Lys His Glu Arg His Pro Ile Phe Gly Asn</td>
<td>145 150 155 160</td>
</tr>
<tr>
<td>Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His</td>
<td>165 170 175</td>
</tr>
<tr>
<td>Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu</td>
<td>180 185 190</td>
</tr>
<tr>
<td>Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu</td>
<td>195 200 205</td>
</tr>
<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 Gin Leu Val Gin Thr Tyr Asn Gin Leu Phe Glu Gin Asn Pro Ile</td>
<td>225 230 235 240</td>
</tr>
<tr>
<td>Asn Ala Ser Gin Val Gin Ala Ile Leu Ser Ala Arg Leu Ser</td>
<td>245 250 255</td>
</tr>
<tr>
<td>Lys Ser Arg Gin Leu Gin Asn Leu Ile Ala Gin Leu Pro Gin Glu Lys</td>
<td>260 265 270</td>
</tr>
<tr>
<td>Lys Asn Gin Leu Phe Gin Asn Leu Ile Ala Ser Leu Gin Leu Thr</td>
<td>275 280 285</td>
</tr>
<tr>
<td>Pro Gin Phe Gin Ser Gin Asp Gin Leu Asp Ala Gin Leu Gin</td>
<td>290 295 300</td>
</tr>
<tr>
<td>Leu Ser Lys Asp Thr Tyr Asp Asp Leu Asn Leu Leu Ala Gin</td>
<td>305 310 315 320</td>
</tr>
<tr>
<td>Ile Gin Leu Val Gin Thr Tyr Ala Asp Leu Ala Ala Gin Leu Ser</td>
<td>325 330 335</td>
</tr>
<tr>
<td>Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Gin Ile Thr</td>
<td>340 345 350</td>
</tr>
<tr>
<td>Lys Ala Gin Leu Ser Ala Ser Gin Thr Gin Arg Tyr Gin His His</td>
<td>355 360 365</td>
</tr>
<tr>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>370 375 380</td>
</tr>
<tr>
<td>Lys Tyr Lys Gin Ile Phe Gin Ser Lys Gin Asp Tyr Ala Gin</td>
<td>385 390 395 400</td>
</tr>
<tr>
<td>Tyr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>405 410 415</td>
</tr>
<tr>
<td>Pro Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
</tbody>
</table>
Ams Arg Glu Asp Leu Leu Arg Lys Glu Arg Thr Phe Asp Asn Gly Ser
420 425 430
Ile Pro His Gin Ile His Leu Gly Leu His Ala Ile Leu Arg Arg
435 440 445
Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu
450 455 460
Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg
465 470 475 480
Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
485 490 495 499
Thr Pro Trp Asn Phe Glu Val Val Asp Lys Gly Ala Ser Ala Gin
500 505 510
Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
515 520 525
Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr
530 535 540
Asn Glu Leu Thr Lys Val Tyr Val Thr Gly Met Arg Lys Pro
545 550 555 560
Ala Phe Leu Ser Gly Glu Gin Lys Lys Ala Ile Val Asp Leu Leu Phe
565 570 575 580
Lys Thr Asn Arg Lys Val Thr Val Lys Gin Leu Lys Glu Asp Tyr Phe
585 590 595 600
Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile
610 615 620
Lys Arg Asp Lys Gin Arg Asp Glu Gin Asp Gin Ile Leu Glu
625 630 635 640
Asp Pro Ile Val Leu Thr Leu Thr Leu Asp Arg Glu Met Ile Glu
645 650 655 660
Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
665 670 675 680
Gln Leu Lys Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
685 690 695 700
Leu Ile Asn Gly Ile Arg Asp Gin Ser Gly Lys Thr Ile Leu Asp
705 710 715 720
Phe Leu Lys Ser Asp Gin Phe Gin Asp Met Gin Leu Ile
725 730 735
His Asp Gin Ser Leu Thr Phe Lys Gin Gin Ile Gin Lys Ala Gin Val
740 745 750
Ser Gly Gin Gly Asp Ser Leu His Glu His Ile Ala Gin Leu Ala Gly
755 760 765
Ser Pro Ala Ile Lys Gly 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 Gin Thr Thr Gin Lys Gin Lys Asn Ser
805 810 815
Arg Gin Arg Met Lys Arg Ile Gin Gin Ile Lys Leu Gin Gin Gin Gin
820 825 830
Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gin Leu Gin Gin Glu
835 840 845
<table>
<thead>
<tr>
<th>Residue</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp</td>
<td>850 855 860</td>
</tr>
<tr>
<td>Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile</td>
<td>865 870 875 880</td>
</tr>
<tr>
<td>Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu</td>
<td>885 890 895</td>
</tr>
<tr>
<td>Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Glu</td>
<td>900 905 910</td>
</tr>
<tr>
<td>Glu Val Val Lys Lys Met Lys Asp Tyr Trp Arg Glu Leu Leu Asn Ala</td>
<td>915 920 925</td>
</tr>
<tr>
<td>Lys Leu Ile Thr Glu Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg</td>
<td>930 935 940</td>
</tr>
<tr>
<td>Gly Gly Leu Ser Glu Leu Arg Gly Phe Ile Lys Arg Glu Leu</td>
<td>945 950 955 960</td>
</tr>
<tr>
<td>Val Glu Thr Arg Glu Ile Thr Lys His Val Ala Glu Leu Asp Ser</td>
<td>965 970 975</td>
</tr>
<tr>
<td>Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val</td>
<td>980 985 990</td>
</tr>
<tr>
<td>Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp</td>
<td>995 1000 1005</td>
</tr>
<tr>
<td>Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Arg Tyr His His Ala</td>
<td>1010 1015 1020</td>
</tr>
<tr>
<td>His Asp Ala Tyr Leu Asn Ala Val Val Glu Thr Ala Leu Ile Lys</td>
<td>1025 1030 1035</td>
</tr>
<tr>
<td>Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys</td>
<td>1040 1045 1050</td>
</tr>
<tr>
<td>Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Glu Glu Ile</td>
<td>1050 1060 1065</td>
</tr>
<tr>
<td>Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn</td>
<td>1070 1075 1080</td>
</tr>
<tr>
<td>Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys</td>
<td>1085 1090 1095</td>
</tr>
<tr>
<td>Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp</td>
<td>1100 1105 1110</td>
</tr>
<tr>
<td>Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met</td>
<td>1115 1120 1125</td>
</tr>
<tr>
<td>Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Glu Thr Gly Gly</td>
<td>1130 1135 1140</td>
</tr>
<tr>
<td>Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu</td>
<td>1145 1150 1155</td>
</tr>
<tr>
<td>Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Tyr Gly Gly Phe</td>
<td>1160 1165 1170</td>
</tr>
<tr>
<td>Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Ala Lys Val</td>
<td>1175 1180 1185</td>
</tr>
<tr>
<td>Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu</td>
<td>1190 1195 1200</td>
</tr>
<tr>
<td>Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Gly Lys Asn Pro Ile</td>
<td>1205 1210 1215</td>
</tr>
<tr>
<td>Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu</td>
<td>1220 1225 1230</td>
</tr>
<tr>
<td>Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly</td>
<td>1235 1240 1245</td>
</tr>
</tbody>
</table>
Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Arg
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
<210> SEQ ID NO 51
<211> LENGTH: 205
<212> TYPE: DNA
<220> ORGANISM: Artificial Sequence
<222> FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<400> SEQUENCE: 51
gaatgctgcc ctgaacaccc gttctctcct gttcctgtct gttccaaagg aatgagggtct
60
catggggga ttccgactcc ccttgaggag ctgaggctct gcggacacgg ccccccacct
120
cccggctgct gatggctgtg ccaggcaac tcttctgtat aacacaccac tctagtat
180
gaaaccatgc caactctgcc tctgtatgt gaaagagca tgggctgccc gctggtggtg
240
gttccatcc tagggctgtg gggagatcat gggacacaac cagctggggtc ataggcttcc
300
tcattctct ctcacactcc ctctacctag agggtttat gatctctctt ctagaaacctc
360
gtagagtcc ccgtctccag cctgcacact cccacctctgg ctttgcttttg
420
cttgggctcag agagctctgg atgcacacga ctttgtgac cttttggttt agagacagc
480
gaaacacacc ctctctgtg gcaccacttg tcctttcttt gccttactat ccccttctgt
540
gatgtaga cccatgggag cagctggtcga ggagggacct cgggcttgcc cccotaacc
600
tagttagcag ccacattgca aatggtggtc gggaaaccag cgggcttgcc cccotaacc
660
tatgtagct cagcattcct atcgggctct cagcgtctcc tactgtgtaa ggccccacagc
720
gtagctcttg ggccccctcg atgtgtcctt ccctggtcct cccctctgct gcacaggggtg
780
gtagcgctct cgagccacag tgaacagctt ccctggtcct cccctctgct gcacagggggc
840
tgaggtgcgg aaggcgctag cagcgtctctt cgcaggagcc cgggcttgcc cccotaacc
900
tgaggtgcgg aaggcgctag cagcgtctctt cgcaggagcc cgggcttgcc cccotaacc
960
gaggcggtgac gggagagccc tggctatgca cttaaatgca agaaggctta cgggtgggca 1020
accacaaacc cagcagagcc gatactgctc tgcggcgctg caggcccttg ctggcccca 1080
agctggcact gggcgccact ctgccgacgc tttgggaggg ccggaggtca tgcggccaca 1140
gggtctcgag ccggcctggc ccgtcgacag aggacacagc atgggcgtgg ctggagcgctg 1200
ggcacacttgg gcccctcctgg cggagacttt ggctgcctgg gcgggcctgg cggcagcggc 1260
agccctcccg ccggctctcgg ttgtagcgac ttttagcgac tttctgtttt 1320
aattatatct cagggcacca ctggtagtgta tggactccca ggtgcctccc tctctatggi 1380
aatataaaaa gctctcttct taatgacaccgc gctatccgac tccagccccag gagcctgggg 1440
tggttagattt cggctctggag ggcagagggg ggtagcgtaca gcaacagcgt tcagggcttg 1500
ggagctgggg gttggtgtact ggtgaggggg tcgaagggta atctattacaa tctctctttt 1560
ttgagggga ccctgttcttc tacctccgac tccacagcag gacaaacagc gtagagctag 1620
ggagggcaca ccctgtatcct tggaggggga cggccagcag tttctctttaa gtagtgagta 1680
gagtggaaact gggccagagt atgtcaatgg gagagggaga gtctggcctct gtgccttag 1740
acgttctgggcc gtttctacag tggagagaaa cccagggagag gggagagtgg gggtcttggg 1800
gaggggagac ccaacctcagc aggtgagagc ttcagctcgg aagctgctggg ccacacagga 1860
tgctcagcttg tcctggagac ccctgctgggg aggttgagac tgcagagaca gggcattagc 1920
cggacctgc aaccagctcc caaggtcgact gggtcttttc agcacaagaag aagcagcttg 1980
gccagccgc cgtagcgtct cggcttaccc tg 2040

SEQ ID NO: 52
LENGTH: 1153
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
SEQUENCE: 52

Met Lys Arg Pro Ala Ala Thr Lys Ala Gly Gin Ala Lys Lys Lye
1  2  3  4  5  6  7  8  9  10  11  12  13
Lys Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
14 15 16 17 18 19 20 21 22 23 24 25 26
Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
27 28 29 30 31 32 33 34 35 36 37 38 39
Arg Ile Phe Pro Ala Ala Gin Ala Gin Asn Asn Leu Val Arg Arg Thr
40 41 42 43 44 45 46 47 48 49 50 51 52
Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys His Arg Arg Val
53 54 55 56 57 58 59 60 61 62 63 64 65
Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
66 67 68 69 70 71 72 73 74 75 76 77 78
Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gin Leu Arg Val Lys Gly Leu
79 80 81 82 83 84 85 86 87 88 89 90 91
Thr Asp Glu Leu Ser Asn Glu Leu Phe Ile Ala Leu Lys Asn Met
92 93 94 95 96 97 98 99 100 101 102 103 104
Val Lys His Arg Gly Ile Ser Tyr Leu Asp Ala Ser Asp Asp Gya
105 106 107 108 109 110 111 112 113 114 115 116 117
Asn Ser Ser Val Gly Asp Tyr Ala Gin Ile Val Lys Glu Asn Ser Lye
118 119 120 121 122 123 124 125 126 127 128 129 130
Gln Leu Glu Thr Lye Thr Pro Gly Gin Ile Gin Leu Glu Arg Tyr Gln
131 132 133 134 135 136 137 138 139 140 141 142 143
Thr Tyr Gly Glu Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys 180 185 190
Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu 195 200 205
Ala Leu Arg Ile Leu Glu Thr Glu Gln Glu Phe Asn Pro Gln Ile Thr 210 215 220
Asp Glu Phe Ile Asn Arg Tyr Leu Glu Ile Leu Thr Gly 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 Glu Glu Phe Asn 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 Leu 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 Trp His Asn Phe Ser Val Lys Leu Met 435 440 445
Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr 450 455 460
Ile Leu Thr Arg Leu Gly Lys Thr Thr Thr Ser Ser Ser Asn Lys 465 470 475 480
Thr Lys Tyr Ile Asp Glu Leu Leu Thr 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 Asp Asn Ile 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
Glu Leu Ala Thr Lys Ile Arg Leu Trp His Gin Gin 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
625  630  635  640
Lys  Gly  Gln  Arg  Thr  Pro  Tyr  Gln  Ala  Leu  Asp  Ser  Met  Asp  Asp  Ala
645  650  655
Trp  Ser  Phe  Arg  Glu  Leu  Lys  Ala  Phe  Val  Arg  Glu  Ser  Lys  Thr  Leu
660  665  670
Ser  Asn  Lys  Lys  Lys  Glu  Tyr  Leu  Thr  Glu  Asp  Ile  Ser  Lys
675  680  685
Phe  Asp  Val  Arg  Lys  Lys  Phe  Ile  Glu  Arg  Asn  Leu  Val  Asp  Thr  Arg
690  695  700
Tyr  Ala  Ser  Arg  Val  Leu  Asn  Ala  Leu  Gln  Gln  His  Phe  Arg  Ala
705  710  715  720
His  Lys  Ile  Asp  Thr  Lys  Val  Ser  Val  Arg  Gly  Gln  Phe  Thr  Ser
725  730  735
Gln  Leu  Arg  Arg  His  Trp  Gly  Ile  Glu  Lys  Thr  Arg  Asp  Thr  Tyr  His
740  745  750
His  His  Ala  Val  Asp  Ala  Leu  Ile  Ala  Ala  Ser  Ser  Gln  Leu  Asn
755  760  765
Leu  Trp  Lys  Gln  Lys  Asn  Thr  Leu  Val  Ser  Tyr  Ser  Glu  Asp  Gln
770  775  780
Leu  Leu  Asp  Ile  Glu  Thr  Gly  Leu  Ile  Ser  Asp  Asp  Glu  Tyr  Lys
785  790  795  800
Glu  Ser  Val  Phe  Lys  Ala  Pro  Tyr  Gln  His  Phe  Val  Asp  Thr  Leu  Lys
805  810  815
Ser  Lys  Glu  Phe  Glu  Asp  Ser  Ile  Leu  Phe  Ser  Tyr  Gln  Val  Asp  Ser
820  825  830
Lys  Phe  Asn  Arg  Lys  Ile  Ser  Asp  Ala  Thr  Ile  Tyr  Ala  Thr  Arg  Gln
835  840  845
Ala  Lys  Val  Gly  Lys  Asp  Ala  Asp  Glu  Thr  Tyr  Val  Leu  Gly  Lys
850  855  860
Ile  Lys  Asp  Ile  Tyr  Thr  Gln  Asp  Gly  Tyr  Asp  Ala  Phe  Met  Lys  Ile
865  870  875  880
Tyr  Lys  Lys  Lys  Ser  Lys  Phe  Leu  Met  Tyr  Arg  His  Asp  Pro  Gln
885  890  895
Thr  Phe  Glu  Lys  Val  Ile  Glu  Pro  Ile  Leu  Glu  Asn  Tyr  Pro  Asn  Lys
900  905  910
Gln  Ile  Asn  Glu  Lys  Gly  Lys  Glu  Val  Pro  Cys  Asn  Pro  Phe  Leu  Lys
915  920  925
Tyr  Lys  Glu  Glu  His  Gly  Tyr  Ile  Arg  Lys  Tyr  Ser  Lys  Lys  Gly  Asn
930  935  940
Gly  Pro  Glu  Ile  Lys  Ser  Leu  Lys  Tyr  Asp  Ser  Lys  Leu  Gly  Asn
945  950  955  960
His  Ile  Asp  Ile  Thr  Pro  Lys  Asp  Ser  Asn  Asn  Lys  Val  Val  Leu  Gln
965  970  975
Ser  Val  Ser  Pro  Trp  Arg  Ala  Asp  Val  Tyr  Phe  Asn  Lys  Thr  Thr  Gly
980  985  990
Lys  Tyr  Glu  Ile  Leu  Gly  Leu  Lys  Tyr  Ala  Asp  Leu  Gln  Phe  Glu  Lys
995  1000  1005
Gly  Thr  Gly  Thr  Tyr  Lys  Ile  Ser  Gln  Glu  Lys  Tyr  Asn  Asp  Ile
<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 ttcgccatgat tccttcctat ttgcatactac gatacaaggc tggtagagag 60
ataattgga tttaattgac tgtgaacaca aagatattag tacaattat gtaatcggctgg 120
agaattatat ttcttgggtta gttttgcaatt tttaaatatt gttttaaat ggaatcattgg 180
atgctttcacc ccattttctgat tccctctctct ttaatttttatt tttaattttaa 240
cgaaacaccc tttaaacatt cctgcaagca ctaaacgatt aagatcctataat ggggaaacg 300
acacccgtc tttatagggc agggtttttt cgtttatgg 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

<400> SEQUENCE: 54

gagggcctat ttcgccatgat tccttcctat ttgcatactac gatacaaggc tggtagagag 60
ataattgga tttaattgac tgtgaacaca aagatattag tacaattat gtaatcggctgg 120
agaattatat ttcttgggtta gttttgcaatt tttaaatatt gttttaaat ggaatcattgg 180
atgctttcacc ccattttctgat tccctctctct ttaatttttatt tttaattttaa 240
cgaaacaccc tttaaacatt cctgcaagca ctaaacgatt aagatcctataat ggggaaacg 300
acacccgtc tttatagggc agggtttttt cgtttatgg 340
<210> SEQ ID NO 55
<211> LENGTH: 318
<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: (256)...(269)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 55

gagggcctat tcctaatat ttgctataac gatacag gcttgagag 60
ataattggaa ttaatttgac tgtaaaca aagatattag tacaataatctgtgacgtaga 120
agatgtaat ttcttgggtt gttggcagtt tttaaattat gttttaaaat ggcattatcat 180
atgcttcag taacttggaaa gatatcggat ttcttggttc tataaatctt gttggaagga 240
cgaaacacn nnnnnnnnn nnnnnnnnngtttagagct agaataagca agttaaata 300
agctgatgct gt ttttttttt 318

<210> SEQ ID NO 56
<211> LENGTH: 325
<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: (256)...(269)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 56

gagggcctat tcctaatat ttgctataac gatacag gcttgagag 60
ataattggaa ttaatttgac tgtaaaca aagatattag tacaataatctgtgacgtaga 120
agatgtaat ttcttgggtt gttggcagtt tttaaattat gttttaaaat ggcattatcat 180
atgcttcag taacttggaaa gatatcggat ttcttggttc tataaatctt gttggaagga 240
cgaaacacn nnnnnnnnn nnnnnnnnngtttagagct agaataagca agttaaata 300
agctgatgct gt ttttttttt 325

<210> SEQ ID NO 57
<211> LENGTH: 337
<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: (256)...(269)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 57

gagggcctat tcctaatat ttgctataac gatacag gcttgagag 60
ataattggaa ttaatttgac tgtaaaca aagatattag tacaataatctgtgacgtaga 120
agatgtaat ttcttgggtt gttggcagtt tttaaattat gttttaaaat ggcattatcat 180
atgcttcag taacttggaaa gatatcggat ttcttggttc tataaatctt gttggaagga 240
cgaaacacn nnnnnnnnn nnnnnnnnngtttagagct agaataagca agttaaata 300
agctagtcc gttatcaact taaaaggt tttttttt

<210> SEQ ID NO 58
<211> LENGTH: 382
<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: 58

gagggcctat ttcctcatg tcttttatg tttctatata ctaaa gataggttg tgcaggttag

<210> SEQ ID NO 59
<211> LENGTH: 5201
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<400> SEQUENCE: 59

cgtaataact cttacggtaa atggccggcc tggctgaccc cccaaagacc cccgcccatt

gagctcaata atagcatgat ttcctcatag aagcccaata gggaccccttc atggacgta

<420> SEQ ID NO 60
<421> LENGTH: 2481
<422> TYPE: DNA
<423> ORGANISM: Artificial Sequence
<440> SEQUENCE: 60

cgtaataact cttacggtaa atggccggcc tggctgaccc cccaaagacc cccgcccatt

gagctcaata atagcatgat ttcctcatag aagcccaata gggaccccttc atggacgta
cttcagcaac gagatggcga aggtggacga cagcttcttc cacagactgg gaggctcttt 1260
cctgtggaaa gaggataaga agcagagacg gcaccccatac ttggccacaac ttcggcaga 1320
ggtccctac acagagaagt accoccaact ttaaacaagtg aagaaxaaac tggatacag 1380
caccaacaag gcgcagcctgg ggcgtaaccttc tccggccctgg gcacacatga tcaagttccg 1440
gggccacct ttgtagcaggg gcgcctgcgaacc gggccagaca acggacagtg gaaagctg 1500
catccagcagt gtgagcaacct cacaacgact ggtccaggaaa aaccctccata accggacg 1560
cgtagcgcc aagggcattac tgtgtcgcag actgagcaag aggagacaggg ttgagaaatct 1620
gatccggccag cttgcccggcg aagaaagaaaga tggtgctgcttc ggacatctga ttgcctctgg 1680
cctgagggcctg aoccccaact tcaagagccca cttggaacgtg gcggagaagtcc caaactgc 1740
gcttgagcaag gacacccactc agcagacgct ggcgcarccct gcggccagca tggcgcgac 1800
gtacgcgac agttgcctcg gcgcacagaac cccggtctgcg aggcttgcagcag cgggcagc 1860
cctgagatgg aaacacagag acaacccagc tccggccaga cccggtgago gcctcttata tcaagagata 1920
cgactagcgc cacccaccggc gcagtctgcg gaagagtccc gctcgccagac agctctgtga 1980
gagtacacaa gagattttcttc tggccaagagc aacaacgccac ataagcgcgc 2040
cgagagcagc caggaagtaact acataaagtt tgaagggacccc atcgctggaaa agatggacgc 2100
cacccggaggt cttgcctgtga agtcggacta ctggcgaacctc acgggacgcc tggcggac 2160
cgacaacgagcgccagcttccc aggatacctcct gcgacgacag cggacagcttgcgcggc 2220
goaggagagg tttaccatct cttgaagagc caccgggggg aagatccagag agactctgac 2280
cctccgacct cctctacacg tgggacctgct ggcagggggga aacagcagact tgggtcctgaat 2340
gacccagaaag gcggagggaa ccacagccttc cttgggaacctcc gggagatttg gggcggaggg 2400
cggtcgcgc cagagttcctc tcggcgctgct gattccactt ggcacacgata acgggtgtag 2460
gaatgtgaag ccacgcacac agctctgtct agatctcact gaagtctgata acgcgcgcgc 2520
caaagttcag aatcgtgacg agggaatgag aaaccccgcct ttcgctgacgc gcgcgagcag 2580
aaggccattgc tgggactctgg ttgctcaagac ccaacggggg aaaaagctgg aagcagcgga 2640
agagcaactac tggagaaaaa tggagtcttt ggactctggg gaaatctccgc ggtgagaaag 2700
tggtcagagc cttccctgcttg gcgcataacc ggatctgcgtt aaattaatcta aggacgaaga 2760
cttctcgagct aatgaggagaa acagaggacct ttcggaagat atcggtgtgct gcttgaacact 2820
gttctgagggc agagagatga tcggaggaac ggatgaaaccct ttcgctgacga 2880
caaaatagtg aagctcgtga aggctggagg attacocaggc tggggccagg ggacagcag 2940
gctgaacac gcggatacgg acaagcgatgc gggcagaaaa acatctgggatt ttcggaagct 3000
cgcctgtcct gcacagcaggaaacctctgcgcgagtcagcgcagactccttga 3060
agaggactac cagaaagccc agggtctcgg ccagggcgag actgctgacg acgcacactgc 3120
cacactgtgcgg gcgcacccgac cattaagaga gcgcaccttgc gaggacgtcg gttggtggtga 3180
cgagcagcactc aaggggattgg ggggggacaa ggcgggagg aagttgctgc ctgcttgcaac 3240
agagcaacta cagaaagccc agggtctcgg ccagggcgag actgctgacg acgcacactgc 3300
agaggactac aaggggattgg ggggggacaa ggcgggagg aagttgctgc ctgcttgcaac 3360
getcgcagac gagagactgtc acattgtacta cttgcaagaat gggggtgata tgtaagttaa 3420
ccaggttcgc gcagcagcctg ggtctgcttg aattcgtatt gcaccatatcg tgccctcgag 3480
ctttctgagg gcgagctcaca ttcgacccgaa ggtgctgacc aagacgacaga aacccggggg 3540
caagagcag acacgtgccct ccgaagaggt cgtggaagaag atgagaatc actggcggca 3600
gctgcagca gcaacagcgc ttaaccagag aaaaattgac cactggaacc agggccagac 3660
agggcggcttg agcgaactgtg atagggcggc tgttcatcaag agaacagctgg tggaaacccg 3720
gcagatcaca aacagcgttgc cagacagtcct ggtactcggcg atgaacacota agtacagcaga 3780
gaatgcaacag cttgacggccg aagaggaagt gatacccttg aagtcocagc tgggttcgga 3840
tttcgagag ctaacgcagcct tttcacaaggt gcgccagatg caccacatgcc accacgcccca 3900
cagcgcctac cttgacgcgcct tgtggggaac gcgcctgcatc aaaaaatcacc ctaagctgga 3960
aagcagagttc gtgtcagccc gacacaaaggt gtcagacgttg cggagagatgag tgcacagag 4020
cagagcggcaa atcgccgaag ctcacgccaa gtaactcttc tccagacaca ataatgaaatt 4080
tttcagacc cagagcctcc tgtgcccaagc cggacatcccg aagccgctct cgtgagcagac 4140
aaacgcgggaa acggcgagaac tggcgtgaaac tggcgtgagc cggatgggca cgtggcggga 4200
agtttcgagc atcgcctacg gcaaatatctg gaaacacgac gcgggtcggc agggcggttt 4260
cagcgaacag tgtatctctc gcaacagggga gagcagcatag ccgtcagccc gaaacagaga 4320
tccggacccct cagagatcgac ggcagcttcg cagccccacg tggcctatt cgtgctggtt 4380
gtgccccaa gtcggggagaa gcagctcaca gaaacagagc agttgcgaag agctgctggg 4440
gatccacaact atggagaagag cagcgtcagc gaagatccccc atcagacctt gtaaagacca 4500
ggcctcaaca gaagatgaaaa agagacctgg ctcagacgtg cttatacgtct cctggtctgca 4560
gtgaaaaac ggcaggagaa agatgctgcg cttgagccg gcagctcacc tttgccagct 4620
actggtctctg cccctccaat atgtgaacct tgtctacagc tggccaccac tttgcagact 4680
gaagaggtctcc ccgagagataagcagctgttt gtggaacagc acaacatc 4740
cctgacgacg atcagtccag cagtccagcgt gttctccaaac agagttggac tggccagcgc 4800
tactgtgcag cagcttcggt ctgctctcgc ctgcagccag cacaagctggt gtcagacgtc 4860
ggcggagata atccactcacc tgtttacccg gcaacactct cggacgcctg cggactctcag 4920
gtacttgagc accacctagcg cggagagagcg tccacccgac accacagagc tgtgacgaa 4980
cacccctgac caccagacga ttcacgcccgt ttcagacagc cggacgccgc tgttctcagct 5040
gggaggccag ttttttttccc ttagcttgac cagcttttctttt aatgtccagc ggaagcgtta 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 nnnnnnnnn nttattgtac tttcaagatt tagaatattaaat cttggcgsa 60
gctacaaga aagggcttca tgtggccacacc aacacctgtg cattttatgg cagggggtttt 120
tcgttatatt aaaaaa 137
<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 ttcgagaaactgcagagtc taagataag 60
gtttcatgc gaaatacaaca cccccctcatt ttaagcaggg gtttggctgg tattaatttt 120

<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 polynucleotide
<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 ttcgagaaactgcagagtc taagataag 60
gtttcatgc gaaatacaaca cccccctcatt ttaagcaggg gtttggctgg tattaattttttttt 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 polynucleotide
<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 ttcagatt tagaataaactgcagagtcattctggcag 60
gttttactga taaggtcta tgccgaaaa aacacctgt cattttatatg cagggttgtttttttt 120
tctgttatattttttt 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 polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1) .. (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 64

nnnnnnnnnn nnnnnnnnn gttattgtac ttcgagaaactgcagagtc taagataag 60
gttttcatgc gaaatacaaca cccccctcatt ttaagcaggg gtttggctgg tattaattttttt 120
tttt 123
ttccttgacgg tgaacgacaa cgagacagag gcgccacgta gcagcgcctg gattaagcggag 1020
tacaagacdg acacaagaggg tctggtctcg ctgaagaggt acatccggaaaa cactcagcttg 1080
aaaaacatca atggaggttgct caaacggcag cacaagagcaca ctaagcgcctg 1140
lggaacgacc accagcagaa tttctatgtg tacgctgaagag cagttgctgc caggttcgag 1200
gggcccagact agtttccgaca aaaaatcagcag cagggactattctggtgc gacgcgcacggacc 1260
ttcgacacag cgcaagcctcg ctacccagacct ccgtgcagag aaatgcgaccc catccgcttgac 1320
aagcagcggc aagtcattacca atctctgccctgc gaaacacaaag aagcgtcgcg cagatccctg 1380
aactctcggc tcccttacta cgggagcccct cttggccagag gcaacacgaaga ttmtgctctgg 1440
tccatccgga aagcagactca cccctggcact ctcgaggaact ttcgagaaagc gatctgacaaa 1500
agatctcgcag ccagggcctt ctcagaacccg atgcagcctgc tcagccgcgaca agagcttgaag 1560
gaaagagtcgcc ggcccaagcag cagccctgcgt ctcagagcact tcatagtgtag tacagagct 1620
aagcagagcg cggtttagcccg ccgggtgcgct ccggcagaga acgtctgcag cttcaagcag 1680
aagggacaca gctgggctgcag cagccgatggat gtaacccgca gcccagcagaa cagccggccctc 1740
atcagagtcgcc ccacggccacag cagccggcag cagcgcgcag acagctcgagc caaatgcaag 1800
cagtcacact tccacccgctg cagagacacac cagctggcagcac cagctgacgaa 1860
ctttggacag cacatcagcag gagaaggtgct gaaaacctgaa cagccggccag 1920
ctttgagggc gcatcaagctg caagcgacagg cttgaccagag ttcgagaaatcg ctggtgacaa 1980
agcgtgcctgg cagagccagcc ccagccggtgc ccggtgcgctgc cttcaagcagac 2040
tctgaacacgg ccttgggcag cggagagtcc gcggcggccgc cttcggagctc ctcctgagctc 2100
gcggcctgct gcacccggcag ctcgctgac ctagccgccc gaaagtgatcag ggtgaaatggag 2160
aagagatctg gccagacccccc ccagctctcag cagctgacgacg caaatgcaagc 2220
gtgagacagcg ctgctgacatt gatgggggctg agagaagcctgcc cagcactggcag 2280
gctagttacag cattccggcc aagaccaaacgc ccagcagacag aagctggaaga 2340
ctgtggaaaat ctcctcaggg ccggtgcgcat ccggtgcgcat ccggtgcgcat 2400
tctggtgaccggac cgggtgcgcaac cgggtgcgcaac cgggtgcgcaac 2460
atcagagtcgcc ccagccggcag ctcgctgac ctagccgccc gaaagtgatcag ggtgaaatggag 2520
aatggtggccag cacatcagcag gagaaggtgct gaaaacctgaa cagccggccag 2580
atcagaccaac cctcctgccg gcaaacagcag ctcgctgac ctagccgccc gaaagtgatcag ggtgaaatggag 2640
atcagaccaac cctcctgccg gcaaacagcag ctcgctgac ctagccgccc gaaagtgatcag ggtgaaatggag 2700
aatgctggaga agagagccagcc ccagccggcag ctcgctgac ctagccgccc gaaagtgatcag ggtgaaatggag 2760
aagctggagca ccagacccccc ccagctctcag cagctgacgacg caaatgcaagc 2820
ctgtggaaaat ctcctcaggg ccggtgcgcat ccggtgcgcat ccggtgcgcat 2880
aatggtggccag cacatcagcag gagaaggtgct gaaaacctgaa cagccggccag 2940
ctgtggaaaat ctcctcaggg ccggtgcgcat ccggtgcgcat ccggtgcgcat 3000
aatggtggccag cacatcagcag gagaaggtgct gaaaacctgaa cagccggccag 3060
ctgtggaaaat ctcctcaggg ccggtgcgcat ccggtgcgcat ccggtgcgcat 3120
aatggtggccag cacatcagcag gagaaggtgct gaaaacctgaa cagccggccag 3180
aatggtggccag cacatcagcag gagaaggtgct gaaaacctgaa cagccggccag 3240
atcagaccaac cctcctgccg gcaaacagcag ctcgctgac ctagccgccc gaaagtgatcag ggtgaaatggag 3300
cgggtgcga tttatctca agtgastgtc gtgaagaagc tgggaagaaca gaaccagggc 3360
cgtgatcgg gcacaccacaag gcggctgtct acaccccaacc ttgctccagaa cccaatgccc 3420
aactccacag gagatctctg gggggcacaag gatgtaccttg acocctaagaat gtaacgagg 3480
taccgcccag tccctcaatag tttccacgctg ccgtgtaagg gcacatcaga gaagggcccct 3540
aagaaaaaga taacaaacttg gtctgaatatt caagggatct ctatcctgga cggactacac 3600
taccgcaaggg ataagcttggag ctctcgcttg gaaagaaggct acaaagcacat tgaagctatt 3660
atccgacgctg ctaaggactct ccttgttcaaa gctctgacag ccgcaggacgc gatcgctggc 3720
tccatatcttg ccaccaacaa caagcgggggc gacatccaca agggaaaccct gcatttctctg 3780
aggcagaat atttgaactt ccggagcgc gcgaacgccg occttcacacacta actaatag 3840
aacccaccag aatatgctgca aaaccccaacg aagagcttgcctg gaaaaagcctgct acctacatc 3900
ctgggatcga acacgacaactga ttggggagccc aaagaagcagc ccaacattgct gaaactcgcoc 3960
tccgagagct ggcacccacac cagcattcgac gacgctgtgca gctcttctact cggccctaccc 4020
ggacagcggc ggaagggctct gtctggactc gctctcagag ccacaatgccc ccttctagctg cttggaagac 4080
tctcttggtg gtagaatcctgg ccgctagccgg ccagctattgg ccctttctttt ctgagctactg 4140
gcaccacctga cccacccagct gctggagcgc gctgtcagac ccctatgtcag ccctgttgaag 4200
cgctggcagagc gaaagctgccgcag cttctgtctact aagaaatcgcg tctcaacta gaaaaagaaa 4260
taa 4263

<210> SEQ ID NO: 68
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: length
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 68
toctagcagatttctgatattctgctcagttttagacagttagctgttttaga 53

<210> SEQ ID NO: 69
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: length
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 69
tggcacatctatctcagatctgtcagga gtttgaggctccatccacacg 53

<210> SEQ ID NO: 70
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: length
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 70
ggggttcagctttgtgacatag 25

<210> SEQ ID NO: 71
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: length
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 71
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 71

gcgaatgaac ggaaaccttt ggtc

SEQ ID NO: 72
LENGTH: 25
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
FEATURE:
NAME/KEY: modified base
LOCATION: (1) (4)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 72

nnngacagc gcaatggcctg aatac

SEQ ID NO: 73
LENGTH: 25
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
FEATURE:
NAME/KEY: modified base
LOCATION: (1) (4)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 73

nnnttatatt ggctcatatt tgctg

SEQ ID NO: 74
LENGTH: 25
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 74

cttacacca atgcrgceas cagac

SEQ ID NO: 75
LENGTH: 46
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 75

caaaatctt agctctctttt gctttttccc ataaaacct cttta

SEQ ID NO: 76
LENGTH: 45
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 76

aggttttat gggaagaagc aaagaagact aaaaatttg atacc
<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
cttacgtgc ataaagtcas ttccc 25

<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
tggtcgatt tcagccatttg c 21

<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
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (33)...(33)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 79
ccttgacgag gcatagtgctg aatcgcgcc aaaaaagcg aag 43

<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
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (34)...(34)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 80
ccttgacgag gcatagtgctg aatcgcgcc aaaaaagcg aag 43

<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
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (35)...(35)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 81
ccttgacgag gcatagtgctg aatcgcgcc aaaaaagcg aag 43
ctttgacgag gcaatggctg aatctgagcc aaaaaagtgc aag

ctttgacgag gcaatggctg aatctgagcc aaaaaagnc aag

ctttgacgag gcaatggctg aatctgagcc aaaaaagtgc aag

ctttgacgag gcaatggctg aatctgagcc aaaaaagnc aag

ctttgacgag gcaatggctg aatctgagcc aaaaaagtgc aag

ctttgacgag gcaatggctg aatctgagcc aaaaaagnc aag
<400> SEQUENCE: 86
ctttgacgag gcataagctg aatcagac cc aaaaaagcn aagaa 46

<210> SEQ ID NO 87
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<222> LOCATION: (41)...(41)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 87
cctttgacgag gcataagctg aatcagac cc aaaaaagcn aagaa 46

<210> SEQ ID NO 88
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 88
gcgctttttt gcgtcgatttc cag 23

<210> SEQ ID NO 89
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<222> LOCATION: (31)...(31)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 89
caatgctgca aatcagacca aaaaaagcga ngaagaatc 40

<210> SEQ ID NO 90
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<222> LOCATION: (32)...(32)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 90
caatgctgca aatcagacca aaaaaagcga anagaatc 40

<210> SEQ ID NO 91
<211> LENGTH: 40
<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: (33). .(33)
<222> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 91
caggtgctga aatcgagcga aaaaaacta ahaagaactc

<210> SEQ ID NO 92
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<221> NAME/KEY: modified_base
<222> LOCATION: (34). .(34)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 92
caggtgctga aatcgagcga aaaaaacta ahaagaactc

<210> SEQ ID NO 93
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<221> NAME/KEY: modified_base
<222> LOCATION: (35). .(35)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 93
caggtgctga aatcgagcga aaaaaacta ahaagaactc

<210> SEQ ID NO 94
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<221> NAME/KEY: modified_base
<222> LOCATION: (36). .(36)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 94
caggtgctga aatcgagcga aaaaaacta ahaagaactc

<210> SEQ ID NO 95
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<221> NAME/KEY: modified_base
<222> LOCATION: (37). .(37)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 95
caggtgctga aatcgagcga aaaaaacta ahaagaactc

<210> SEQ ID NO 96
<211> LENGTH: 44
<212> TYPE: DNA
ORGANISM: Artificial Sequence

FEATURES:

- OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

- NAME/KEY: modified_base
- LOCATION: (36)...(38)
- OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 96

catggctga aatcagocca aaaaagcgcga aagaacnt acca  

SEQ ID NO 97

LENGTH: 44

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURES:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

- NAME/KEY: modified_base
- LOCATION: (39)...(39)
- OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 97

catggctga aatcagocca aaaaagcgcga aagaacnc acca  

SEQ ID NO 98

LENGTH: 47

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURES:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

- NAME/KEY: modified_base
- LOCATION: (40)...(40)
- OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 98

catgggtca aatcagocca aaaaagcgcga aagaatncn accagc  

SEQ ID NO 99

LENGTH: 47

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURES:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

- NAME/KEY: modified_base
- LOCATION: (41)...(41)
- OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 99

catgggtca aatcagocca aaaaagcgcga aagaatnc naccagc  

SEQ ID NO 100

LENGTH: 31

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURES:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 100

gatctctcat cgctacaacc cacaacccctg g  

SEQ ID NO 101
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 101
aattccaggg ttgtgggttg taggatatgga 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
catgatcct attccttaat aactaaat 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 actcaacaag ttcagtttg 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 ttcacattta ggaaaaagga tgctg

<210> SEQ ID NO 105
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 105
aaacagcat cttttttctt aatggagaa aaat

<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
aaacctttaa tcaagtcaaa atacagcaaa aattg

<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

aaaacaatt tgtgtgtatt tgtgaagat ttaag

35

<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

aaaactttta tcatcagacc aatctgcttt atttg

35

<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

aaaacaata aagcagattg gtcgtatgat gaaaa

35

<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

aaaactccgtc aaggaaggtc gtaaagaaa tcag

35

<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

aaaactcgcgt ataacttcag ataacgtctg gacga

35

<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

aaaactcctcg tccaggttt ccttaaaaat cctcg

35

<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

aaaaacagga ttttaagga aaccttggag agatt

<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

aaaaagccatc gtcaggaaga agcatatgctt gagtg

<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

aaaaacactca agcatatgctttctgacg atggc

<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

aaacatctct atacattatgg aaaaaaaaaa gattg

<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

aaaaatcata aagaaaattc aataagtata gagat

<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

aaactagctg tgaatatgctg caaacaccgc otctcg

<210> SEQ ID NO 119
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
aaacgaggg cttgttttgc ggactacag acctag

35

aaaacatcggc agtcgacatc agtaattatc tttag

35

aaaacaaatc ataattacct gctcgcctt cgcct

35

aaacagatg gatcgcacca gtaagctca atag

35

aaacactgact caattgtac ccactcttgat cct

35

gagaccccttg aacctccgag acttgctctca gttttggac cattcaaac ag

52
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 125
tgagacagt ctggaagct caaagggctc gttttagagc tatgtgttct tg

<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 gatgtgctt 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
aaaaaaacaa acggaacttc gcgctcgta 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: 129
atgcggctac tgcgcgggcct ctgcgggat taacgaaatca tcctg

<210> SEQ ID NO 129
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 129
gtgaactggcg atgctgtgcg aatggacagt cagactactc ttcct

<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
ttaagaata atccttcatct aasataact tcaagtacactt cctagctgac
<400> SEQUENCE: 131
atggatttgagtcagctaggagtttgactgtagatatatttagtgaag 48

<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
gagacctttgagctcagacttgtctcaatatgggacacctaacagttgacatggct 60
taaasctctgtagatattttgtgc 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
gagaccaacgcgtaacctcaggaagttctcgtagattcgttctgtgatggtcttttgaatggctcc 60
aaasctcttgccacttagagctggctg 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
agttcttcaagctcaatgcaaataaatg 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
cggtgtaaatggataactcttcagtgaaag 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
tgcctctctttcacaacaaggg 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

aagccaaagt tggcaccac c 21

<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

gtagcctatt cagtcctagt gg 22

<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

cgttttgtga aactatggt gcaaattac g aatctttcc t g ac g 45

<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

cgctcaggaga agattcgttaa tttgcaccca ttggctcaac a a ac g 45

<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

gatattaggg agccttttttt t g t g g g t t t t t t g a t c a a a a a c a t a t g 48

<210> SEQ ID NO 142
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 142

catatagt t t t t t t t t t t t t t t t t g a t c a a a a c a c a a a a a a g g t c c a a a t a t a t g 48

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

SEQUENCE: 143
atatttcct aataactaa aatag

SEQ ID NO 144
LENGTH: 24
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 144
cgtgtaaatt gcgtagctct gcgc

SEQ ID NO 145
LENGTH: 24
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 145
gacccgctgta ctaagctcctcc tagg

SEQ ID NO 146
LENGTH: 47
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 146
cctagggact gtagttcaacg gtgcaaatag taggcagaataa aatat

SEQ ID NO 147
LENGTH: 44
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 147
gccgtagcatg acsaattgta cagtgttgtgt gaataagttggt gcgc

SEQ ID NO 148
LENGTH: 25
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

SEQUENCE: 149
tcsaatttt ccaattgtgct ctcctc

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

<400> SEQUENCE: 149

cattttttt agttttaaag aataataacc agcatacagt ccaccc

47

<210> SEQ ID NO 150
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 150

agacgattca atagacaata agg

23

<210> SEQ ID NO 151
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 151

gttttgggac cattcaaaaac agcatagtc taaaacctcg tagac

45

<210> SEQ ID NO 152
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 152

gtatgtgtgt tttgaatgtg cccaaaacca ttatatagaac aacagaggtg

50

<210> SEQ ID NO 153
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 153

gtatgtgtgt tttgaatgtg cccaaaacgc acccattagt tcactaaaacg

50

<210> SEQ ID NO 154
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 154

aattcttttc tcatacatcg gtc

23

<210> SEQ ID NO 155
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of ArtificialSequence: Synthetic
primer

<400> SEQUENCE: 155

aagaagata gaagattggtt cattg

<210> SEQ ID NO 156
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 156

ggtactaattc aaaaatggtgagag

<210> SEQ ID NO 157
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 157

gtttttcaatc atggtggtt gcg

<210> SEQ ID NO 158
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 158

aaaatgtgga aaaaatggtgac aacac

<210> SEQ ID NO 159
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 159

atctcgaaaa ctctatcgttt tttttttttt gcggttgagcg cattcaaaaac agc

<210> SEQ ID NO 160
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 160

tttaaaaaaca accattccttt tttaggttct gttitagagc tattggttt tgttga

<210> SEQ ID NO 161
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
aaacggtatc ggttcttttt aaatgaatg gtttggac cattcaaaac agc

<210> 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
aaatgaatg aaagagaac gatacgttt gttttagac tattgtttt tga

<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
gttcccttaaa ccagagctgt atcggtttct tttaaatc

<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
gaagcggatatc gctggtttggt ttaaggaaca ggtaaagggc attaac

<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
cagattgagc cattgcctcg tc

<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: (29)–(33)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 166
gcctttgacg aggccatggc tgtaatcyn nnnaaaagc gcagaagaa atcaac

<210> SEQ ID NO 167
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 167
tcgtcaac ccacacccct gcattggagc gttttggagc cattaacac ac 53

<210> SEQ ID NO 169
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 168
gtcctagtac agggttggat gtttagcag atttttagagc tatgtgttt tga 53

<210> SEQ ID NO 169
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 169
tgggtgccac cttccctct ttc 23

<210> SEQ ID NO 170
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 170
caggggttag ggtgtttgctg attgtagtaa ctcccatctc c 41

<210> SEQ ID NO 171
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 171
ggggatggta acctatggcct ccacacccac caacctgtgt g 41

<210> SEQ ID NO 172
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 172
gtggatcctta tgggtctgtgc ac 22

<210> SEQ ID NO 173
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 173

ttacgaaac ggaatttatac 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

aagctagag ttccgcaatt gg 22

<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

gtgggttgta cggattgag taacctccat otcttc 37

<210> SEQ ID NO 176
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 176

gatggaggtt aacctcaatcc gtacaaccca caacocctg 38

<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

gttaacacta ttgcaagaco atgggccaca atgaagatag 40

<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

gtggctcaatt ggtgctgcag tagtagaagc taatggtgat g 41

<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
primer
<400> SEQUENCE: 179
tgtattgta ttaatgtaa gacattatgc ttcaccttc
  39

<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 ctaaataacta tacaatcag tgaatcatg
  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
gttttggac cattcaaact agcatagctc taaaagctga 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
gtttttaggc tagtctgttt tgaatgctc ctaaagctga actacgcggg tgt
  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
atacttaac cagacgctgg ttcggttttg taggagtgttg agtatataca cagctcat
  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
gctcaatag ccgggtgtgg gttgtacgga tgtg
  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

tcttcagcgg atttctgata ttactgtcac tgg

<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: 186

ttttaaaagaa accgataccg tttacgaaat tgg

<210> SEQ ID NO 186
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 187

ggaacccttc ataacagcag acaaggtatat aataaggtctat cgtccgttac t aac ttgaaaa

<210> SEQ ID NO 187
<211> LENGTH: 64
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 188

gtggcaccg agtcgggtct tttt

<210> SEQ ID NO 188
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 189

gttatagac tctcgtgta tgaatgtcc caaaac

<210> SEQ ID NO 189
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 190

ggaacccttc ataacagcag acaaggtatat aataaggtctat cgtccgttac t aac ttgaaaa

<210> SEQ ID NO 190
<211> LENGTH: 64
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 191

gttatagac tctcgtgta tgaatgtcc caaaac

<210> SEQ ID NO 191
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220>  FEATURE:
<222>  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: 191

nnnnnnnn nnnnnnnn gttttagac tagaatagc aagttaaaaat aaggtagtc  60
cgttataac tggaaaaagt gcgccagagt cggtgccttt ttt  103

<210>  SEQ ID NO 192
<211>  LENGTH: 103
<212>  TYPE: DNA
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<222>  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 nnnnnnnn gttttagac tagaatagc aagttaatat aaggtagtc  60
cgttataac tggaaaaagt gcgccagagt cggtgccttt ttt  103

<210>  SEQ ID NO 193
<211>  LENGTH: 123
<212>  TYPE: DNA
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<222>  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 nnnnnnnn gttttagac ttagttgta ttggaaacaa acacatagc  60
aagttaaat aaggtagtc cggttataac tggaaaaagt gcgccagagt cggtgccttt  120
ttt  123

<210>  SEQ ID NO 194
<211>  LENGTH: 123
<212>  TYPE: DNA
<213>  ORGANISM: Artificial Sequence
<220>  FEATURE:
<222>  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 nnnnnnnn gttttagac ttagttgta ttggaaacaa acacatagc  60
aagttaaat aaggtagtc cggttataac tggaaaaagt gcgccagagt cggtgccttt  120
ttt  123

<210>  SEQ ID NO 195
<211>  LENGTH: 20
<212>  TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 195

gtcctcctca atgactagg  
20

<210> SEQ ID NO 196
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 196

gactcgtg ttcctcctca tgg  
23

<210> SEQ ID NO 197
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 197

gatccgagc agaagaagaa ggg  
23

<210> SEQ ID NO 198
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 198

gcgcacccg tgtatgtcgt ggg  
23

<210> SEQ ID NO 199
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 199

ggggcacaga tgagaactc agg  
23

<210> SEQ ID NO 200
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 200

gtacaacgg cagaagctgg agg  
23

<210> SEQ ID NO 201
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 201

ggcagaagct ggaggagaaa ggg  
23

<210> SEQ ID NO 202
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 202

gggccttc ttcttctgct cgg  
23

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

<400> SEQUENCE: 203

gggcaaccac aaacccagaga ggg

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

<400> SEQUENCE: 204

gccccatca cactaaccggt ggg

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

<400> SEQUENCE: 205

gtgcgcgcat gccacgaga agg

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

<400> SEQUENCE: 206

ggcagatgct gtgtgctgc tgg

<210> SEQ ID NO: 207
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 207

gcctctgctgt ggcccaagc tgg

<210> SEQ ID NO: 208
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 208

gatggcccag agtccagcctt ggg

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

<400> SEQUENCE: 209

ggctccccca aagcttggcc agg

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

<400> SEQUENCE: 210

ggggccccga ttggtgttc agg

<210> SEQ ID NO: 211
gtggccgagag ggcgcgagat tgg

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

<400> SEQUENCE: 210

gaagttcgcc gaggcggaggg cgg

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

<400> SEQUENCE: 213

ggagttcgcc cgagccggaggg cgg

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

<400> SEQUENCE: 214

ggagcgaggt ggcgcgaggg cgg

<210> SEQ ID NO 215
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

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

<400> SEQUENCE: 215

coccccttcg tgtgaatgt

<210> SEQ ID NO 216
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

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

<400> SEQUENCE: 216

ggagattgga gacacggaga

<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

aagcaccgac tcggtgcac

<210> SEQ ID NO 218
**US 8,871,445 B2**

---continued---

<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 218

tacctccas tgactagggg

<210> SEQ ID NO 219
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 219

cagttgatat ccgactagc ct

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

<400> SEQUENCE: 220

agttccgagca gaagaagaag ttt

<210> SEQ ID NO 221
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 221

tttcaagttg ataacggact agcct

<210> SEQ ID NO 222
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 222

aacaccgaga ttcgcttgga

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

<400> SEQUENCE: 223

tcatacgtc gatgaagctc

<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

tocaaacctc agtggggcctg

<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
tgatgacctt ttgggttttccc

<210> SEQ ID NO 226
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 226
gaggaattct tttttgttg gtaaatgttg gaggtttttt gggag

<210> SEQ ID NO 227
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 227
gagaagctta aataaaaaac raacaatctc aacccaaaaa cc

<210> SEQ ID NO 228
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 228
caggaacag ctatgacctgag cccctatctcc ctaggttc

<210> SEQ ID NO 229
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 229
gcctctagag gtaacctggag gcctattttccc catgattccc

<210> SEQ ID NO 230
<211> LENGTH: 133
<212> TYPE: DNA
acctctagaa aaaaagcacc gactcggtgc caacttttcta aagtgtacac ggacagctcct 60
tatatattaact tgtatatattct gactctaaaa cnnnnnnnnnnnnnnnnnggtgtgctg 120
ctcttttccac aag 133

acctctagaa aaaaagcacc gactcggtgc caacttttcta aagtgtacac ggacagctcct 60
tatatattaact tgtatatattct gactctaaaa cnnnnnnnnnnnnnnnnnggtgtgctg 120
tcttttccac aag 133

acctctagaa aaaaagcacc gactcggtgc caacttttcta aagtgtacac ggacagctcct 60
tatatattaact tgtatatattct gactctaaaa cnnnnnnnnnnnnnnnnnggtgtgctg 120
ctcttttccac aag 133

acctctagaa aaaaagcacc gactcggtgc caacttttcta aagtgtacac ggacagctcct 60
tatatattaact tgtatatattct gactctaaaa cnnnnnnnnnnnnnnnnnggtgtgctg 120
ctcttttccac aag 133

acctctagaa aaaaagcacc gactcggtgc caacttttcta aagtgtacac ggacagctcct 60
tatatattaact tgtatatattct gactctaaaa cnnnnnnnnnnnnnnnnnggtgtgctg 120
ctcttttccac aag 133
<210> SEQ ID NO 234
<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: 234

agggcctagt gtggctcttn aa 22

<210> SEQ ID NO 235
<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: 235

acrcaacg gtggctcttn at 22

<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

aaggtgtgtgt tccagaac gn ac 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

ccacacatc accggtgtgn ag 22

<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

aaccggcaga agctggaggn ta 22

<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)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 239

ggcagagcggagaggtttt

SEQ ID NO: 240
LENGTH: 22
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: modified_base
LOCATION: (20)...(20)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 240

ggttgtgtctcagacggcgcgttc

SEQ ID NO: 241
LENGTH: 22
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: modified_base
LOCATION: (20)...(20)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 241

aaccggagcaagtacagtagg

SEQ ID NO: 242
LENGTH: 22
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: modified_base
LOCATION: (20)...(20)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 242

ttcagaaccggagaccaanc

SEQ ID NO: 243
LENGTH: 22
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: modified_base
LOCATION: (20)...(20)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 243

gtggtggtcagaacgggcat

SEQ ID NO: 244
LENGTH: 22
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: modified_base
LOCATION: (20)...(20)
OTHER INFORMATION: a, c, t, g, unknown or other

SEQUENCE: 244

tccagacagggaccaancc

SEQ ID NO: 245
LENGTH: 22
cgsaagctgg aggaggaagn cg

catcaacogg tgccgcattgn ga

gcgsaagctg gaggaggaan gt

cctccccccc tgcgcagagc gc

tcatctgtgc cctcccctcn aa
ggagagcat cgatgtcagcn at
<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

caacggcgag aagctgagcn ac
<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

ggagagccaa cacaacaccn ag
<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

ggctgccatt cactcaacccn ta
<210> SEQ ID NO 254
<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: 254

ggagagccctg asgctgccgagcn tc
<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
<222> LOCATION: (20)...(20)
<222> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 256

caacggtgg cgacattgcn tg

<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

agaggaagggcgctgagtcn ca

<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

agctggagga ggaagggccn ct

<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

gcaattgccac gaagcaggcn cc

<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

atgccccag aagcagggccn cg

<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

agaacccgag gacaagtcn ga

<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

tcaacggtg ggcacagtgc

22

<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

gaacgtggag gagaacagcgc

22

<210> SEQ ID NO: 264
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 264

ccatgagggg gacattctag gtcacctccca atgacttaggg tgtgcaacca caaaccccgag 60
agggcagag gtagttgctg ggtggccagg ccctggtcgtg gcggcagcgt gcacgtggc 120
cac 123

<210> SEQ ID NO: 265
<211> LENGTH: 121
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 265

cgacacagc agagtggtgg ccctacacta caacggtcg gctgtgaga cgaacgcagg 60
cagtgagggg gacattctag tcaacctccca tgaactaggg ggggaaccac aaaccacga 120
g 121

<210> SEQ ID NO: 266
<211> LENGTH: 128
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 266

gagacacag gtcacaacagc cagaaagtgg aggaggaggg gctgtagtcg cagacagag 60
agagggctgc ccatcacata aacggttgcc gcatggccac gaggcagggc aatggggagg 120
acattctag 128

<210> SEQ ID NO: 267
US 8,871,445 B2

299

-continued

<211> LENGTH: 130
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 267

agaagctgga gagggaaggg cctgagtcg aagcagaagaa gasgggctcc catccacatca 60
acggtgtgcg catgcccacg aagcagggca atggyggaga catcgagtc acctcaccatg 120
actaggttg 130

<210> SEQ ID NO 267
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 269

cctcagttt cccatcaggc tccagctca gctgagttgt tgaggccoca gttggcgcctc 60
tgggggcctc ctgagttttc catctgagcc cctccctccc tggcccaggt gaaggtggttg 120
ttcc 125

<210> SEQ ID NO 269
<211> LENGTH: 129
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 270

tcactcgtgc cccctcctcc ctggcoccag tgaaggtgtg gttccagac ccgaggccac 60
agtacaaag cgacagctgt gagggaag ggcctgagtc cgacagaaag aagaagggct 120
cocatcaca 129

<210> SEQ ID NO 270
<211> LENGTH: 129
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 271

cctcaagac tagggggtggg aaccacaaaa ccagaggaggc agaatgtctgc ttgtgtgtgg 60
cagggcccct gctgtggggcc aagctgagct ctcgcccactc cctgcaccag ctgctgggag 120

gccctggagt 129

<210> SEQ ID NO 271
<211> LENGTH: 127
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 271

cgyctgtccg ctggccaggg cctggctgtg gcccagcgtg gactctggcc actccctggc 60
caggtctttgg gggggtctggt agtcagtgccc ccacagggct tgaagccggy ggcggcctatt 120
gacagag 127

<210> SEQ ID NO 272
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 272
gaatatta ctgaacctctag gagg

<210> SEQ ID NO 273
<211> LENGTH: 126
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide

<400> SEQUENCE: 273
aaaaaagac ggactcgttg ccactttttc aagttgtttac ccggactagcc ttatatctttac 60
ttgctacctt tagctttoaa acaaagcagc gcgtagcacc acccatatgt gacgctatt 120
aatttc 126

<210> SEQ ID NO 274
<211> LENGTH: 126
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide

<400> SEQUENCE: 274
aaaaaagac ggactcgttg ccactttttc aagttgtttac ccggactagcc ttatatctttac 60
ttgctacctt tagctttoaa acaaagcagc gcgtagcacc acccatatgt gacgctatt 120
aatttc 126

<210> SEQ ID NO 275
<211> LENGTH: 4677
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymucleotide

<400> SEQUENCE: 275
tttttctgtg gctatgacac ttccacagaaa agtatgggcct gtctgacgaga cggttcccgg 60
gggtgtgcgt ggcacctgag ccaccctgcag tctcttgggg gcgtgtatgg 120
gcctccagtt gcgtgcctgg gcgggtctgg ttttatagc cagggccccg atttgcaagc 180
cattttcagcgtactaca gacccactaca acaccaactgt ctatcactca tcttacacag 240
gccctacgct cttttgtatcc cactgtcgc tgggggctcc ttttttcttgg cgtttttgtc 300
cacacccgac actatgtacc cactatgtt ctaagattac gttcgcacag aggaaagac 360
cgggtggtggc gtctgacgaga agaagctgac ggcctggcc gacccactaca ggatgcttggg 420
cacacccgac ccgacaagcc ccagccggtgg gcctggtcgt cttcagaggg 480

caacacctcc tcaagagaga cctgatcgc ggcgtgtcgt gcagcagcag 540
cgaacacgcc gaggccaccc ggctggaagag aacgcacaga agaagataca ccagacggaa 600
gaacccgatc tgctatcgc aagagatctc cagcaacacag atggcccaagg tggacacagc 660
cctcttccac aagctgtgac agtcctctctg gtggtaagag gataaagaag aagagccgga 720
ccccctcttt gcgcacactg tggcagcaggt gcggccaccc gagaagtaacc cccacactta 780
cocctgaga aagaaacttcg tggacacgac ccacacacgc gacgcctggcgt gatgtctact 840
ggcctctggcc cacatgattac atggctcggggc ccactttcct tgtcaggggct acctgaaccc 900
cgacacacgc gacggtgacac agctgtctctg ccagcttggtg ccagcttcaca accagtgttt 960
cagagaaac cccacactacgc ccagccggtg gcacagccag gcctctccgtct tgtccagact 1020
gagcaacagcc agacgcctgg gaaaatctgtac ccgccaacctg cccgctggac aagaaagatgg 1080
cctctgtgagc aacctgtcttc ccctgccgct ggcggctggcc cccacactca agacgaacct 1140
cgacccctggcc gcagctggatat gcccagcgct cggcctggcc ccaacactca cggccgacct 1200
cacactgtct gcggagctgt tcggacacgt cggcagccct gcggcaacgct tttctggtgc gcgaacaacct 1260
gtgcacgagcc atccctgttg gcacactcct gcaggtgtaac ccagcacacgta ccagaggccccc 1320
cotgaggggcc ttatggtcgg ccagagataca gcggcaacac cggacacctgta ccctgtggaas 1380
agctctgtgt gcggcagagc tgtccttgcggag atctatttttcc aaccagcgca 1440
gagcggctgcac ccgggcttcg tgtcggcagga gtcacaaagc attttcccctc accagcgca 1500
cagccctgcctc ggagggccgc tggagggcag acagaggtcag atctctacgc tgtcggagaga 1560
ggacacgctg cggaagccag gcacgctgga ccagggcgac atcccccacc gcagaacacc 1620
gggagacgct gcacagtccgc tggccggcgc gagaagtttt taccacatcc tggagggcacc 1680
cggggaaaag atcgagaaag tccctgttct ccgcatcacc ttacatgtgg gcgccttgcgg 1740
cgggagggac gcagacatctg ctcgctgtac gcagagacacg gacgacacac gacgacacac 1800
ggacgctgcac gcacagacatg tccgttgagc gcagagacacg gacgacacac gcagacagcag 1860
ccacacagt gcacagctgg ccaggcgaagag gcacagcagc gcgctgtgccc aagacccgac gcgggtgagca 1920
gcgccctgcct gcgcctgccgag aagaggtggcc atggctcggg aacagacacc gcacagcagc 1980
gggagagctg gcgcctgccgag aagaggtggcc atggctcggg aacagacacc gcacagcagc 2040
cggggggct gcgcctgccgag aagaggtggcc atggctcggg aacagacacc gcacagcagc 2100
tccctgtgga atccctgcttg gtcagatcag ccctgttctgt ccctgtggcgc aataaccaaca 2160	tcctgtgga attatacgag ccaggacact ttcggtgacat gcgggaacct gcgggatcttt 2220
gggagaggtgc gcacagacatg tccgttgagc gcagagacacg gacgacagcag gcagacagcag 2280
gaacacccgt gcgcctgccgag aagaggtggcc atggctcggg aacagacacc gcacagcagc 2340
cgcctgcttc gcgcctgccgag aagaggtggcc atggctcggg aacagacacc gcacagcagc 2400
cacacagact ctggagctcc gcagcctcgc gcgcctgtcgc ccagagacact acagagacact 2460
gatcacaacgcc gcagacgctgc cccttaaag ccgcacactgc ccagagacact gcgctgtgcca 2520
gggcgtgaag ctcgagagcc acctggccag tgtgacccgc ccagagacact gcgctgtgcca 2580
catcctgtcg acaggtgaag ccggtgagcc gcgctgtgga gtcgctgtgca gcgctgtgca 2640
cgagacactc ggtgacacag accagagaca ttcggtgagc tcgggctgagc ccagagacact 2700
cgcgcgagcg aagagacagc gcagagctgcc gcgagagagc gcgctgtgca gcgctgtgca 2760
gagacacgcc cccctgcgcc aacagacgtgc gcgcacactgc ccagagacact gcgctgtgca 2820
gcagagctgg ccgcatgtct gcgcacactgc ccagagacact gcgctgtgca gcgctgtgca 2880
cagatggtgac cattaacctgg ctctagagct ttctgaaggac gactccacac gaaacaaggt
2940
gctgacagca agacgaaaga accgggcaaa gacaagcaca cgtgcctcctac acaggctagt
3000
gaaagaaatcg aagactaactgc ggcgacagct gctgcaagcc aagctgtagta ccacagaaaa
3060
ctgccgacat ctgacacagg ccagagagag ccgcttcagc gacactggtta aacaggcagcgt
3120
catcagaga cagctgtgag aacaagcgca gatcacaagag cagctgggcag aagctcctga
3180
cctcggcgagg aacactaagtg aacagcagaga ttgcaagcttg ttgagggagag cagaaagct
3240
cacccctgag tctcagtcgg tctccagattc ccggaaggt tacccagttttt ccacggtcg
3300
cgagatacaac aacatcaacac acgcctcagca cgcttaacctg aacaggctcgtc tgggaacccc
3360
cctggataaa aagtacatca agctggaagag cgatattgtg taacggctacct aacagtctta
3420
cgatcgtcgg aagatgctacg ccaagagcga gcggaaatct gcagaggtta cggcaagttc
3480
ccttctctac aagcaacatac tgaactctttt ccagacgtag ataccccttg gaaaccggagca
3540
gatcggccag ggcctcctgta tccagaacaa ccggcggacacc ggagsgatgtc tgggtgattaa
3600
gggcggcggt ttgcccacccg tcgggaaggt gctgacatttg ccaccaagctga atacggtgaatt
3660
aaagacgagc gtgcagagcgc cggcatccag caaagagctct atocggcoca agagggacag
3720
cgaataagct atcgcagcaga aagagaagggt gacaccaactg gtcctcgacag
3780
ccccctctcg gctatccttg tctgctgatt gcggcagaggt gaaagagggc aatgccaagct
3840
actgaaagatg tggaaacgagc tggcctggaatt cccagctatcg ccaaacacgat gtaagggaa
3900
aatcgcctac gacctatctgg aacccagggac ctcagaaagaa cttgaaaaag acctgcatat
3960
caagctcgct aagttctccct ttgugcagct ggaaaaacggc cggagggcaaa tggcggctcct
4020
tgcggcgagc ctgacgaaagc gaccagacatgc ctggctgcgc cctaaatatg tgaactcttc
4080
gtacctgcgg cgagcaactgtg agaagctgaa ggggctccccgc ggagaaaaat agcagggaca
4140
gctgtttgct gacaaccatgc agcataactgc gcggaggttac atcggcagcata tcagctgagtt
4200
cctcaagaga cttgacgctc cccgacgctcct ctgtgacgaaaa tggctgcgct gcctcaacaa
4260
gcacggggtg aagatcctgac ggcagcagcgc cagagatatac atccacctgt ttcaccctcg
4320
caatcctggga gcocctggcg cccttcagta ctctgaccacc accarccaco gggagaggtta
4380
ccacccagcc aaagggctgc tggagcccaac ccgtgaccca cagaccatca cggccttcagta
4440
cggacccgtg atgacctgtct ctcagctgag ggcggcaagc cccagagaga gagaagaggt
4500
ggaggcccgc taaggctacgc gcaagctggg ccctggctgg gaaacgaaca ctggagctcccct
4560
ccctagtctg tttggtgcctg tgaacttgta ttcgctgatt gcggccaggg tcacccgaaga
4620
cgatttgata ccgccgagtcc tattctctgt ccagatgttaa cgtggtgattg tgggtact
4677

<210> SEQ ID NO: 276
<211> LENGTH: 3150
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 276

ttttctctgc gtatgcaacac ttccagcacaag aagagggccc ggtgcggagag cggcttcctccg
60
ggctgtgcagc aacacaagatc gatgctcctga ccccccagaag ctccttccgggt gctgcctgggg
120
cgctgtgtgcagc gcgcgccctgc tttatatcgc cggcccaccgc attgcaagag
180
cattatagc gcgctacacaag cgcctataca cacaaccttga tcttacacgc ttcctacacg
240
| GCACTGGAG CTTGATCG CACCTGCTA AGGGGGCGCC TTTCCCTTT CTTTCACTG | 307 |
|---------------|-----------------|---------------|
| ACAACCGCG AAGACCCAA AGAAGAACG GAGGGTTAAC ACGTTAAAC TGCTGAAGA | 360 |
| CGACTCTCT CACCATGAGA TGGTGATCG CGGTTTCACT CACCTGCGT ACCATCAGG | 420 |
| TGAACAGGTC AGCTGCAGA ACCTGGCCCT TGCATCAGG TGCTGAGG CTGGTGACG | 480 |
| AGCTGCACG ACATGTTGTC AGCCTAATG TAAAGTCGT GAGGTTGGG ATACGCTG | 540 |
| CGACTCGTAC CCGACCCACT CCGTACCTAC TGAAGATTG CGAGTTGCGA AGCACTGCT | 600 |
| TGAGGAAGTG AAAGCTAAGT CGCGCAAGCG CCACAGGCA TACCAAGCAC TGAAGAAGG | 660 |
| CAACCGCGAA GCGTCTAGGT ACACTGCCAC TACCAAGCAG AACAGAAGC TGAGGAACG | 720 |
| ACTCAAAAC AGCTGACCG AGCTGCTACAA GAAGACATTG ATGCGATTG TGAAGGTCG | 780 |
| CATTCGCTCT ACGCTAGCGG TCGTGGCGAC GCCATTGGT CTCGACGGG ACCAGGCG | 840 |
| TATCATGTCG GATGACG ATGCTGAGAT GCTCAGATG ATCAGCGGA TACGACGCTGT | 900 |
| AACCGCGTCA ATCGGACCG TACTTGTGAG ACTACCAAGC ACAGAGACG TACGACGCTG | 960 |
| ATACGTGCTG CATGCTGCA GCGCTGCGGG TGGCTGAGGT GCTGACTTCG GCATGTCGCA | 1020 |
| AGCTGCTGCTG TCGTGCTGAC TACGCTGACG GAAGCTGACG AGCTGCTGACG TACGCTGACG | 1080 |
| CGCTGCTGCTG CTCGCTGACG GGCTGCTGACG GCCGGGCTG GGGTGTTCG GCTGACGCGG | 1140 |
| AGCTGCTGCTG TCGTGCTGAC TACGCTGACG GGCTGCTGACG GCCGGGCTG GGGTGTTCG | 1200 |
| AGACTCTAGT TCTGCTGGCG TGGTCTGAC TCTACAGTAGA AAGACACTAGA TGGCTACAGA | 1260 |
| AACACGCGCTC ATCGGCTGACG TTACGAACG GATAACATG CCGGAACACG AACAGGCGAC | 1320 |
| ACTCCTCAGT AAGTTGCTAC CGCCCGCGTG CGTAAATGCA AAGTTGACAG GTGTCCGCGT | 1380 |
| CGAAGCCGAC CCAGTGACCT AGCTGGGCA ACGTTCAGT AACGGGGAA CAGACGACAT | 1440 |
| GAACTCGCTG TGGCAGCAG TGGGCCAGG TGGCAGCAG TGGCAGCAG TGGCAGCAG | 1500 |
| GGTGCTGCTG TCGTGCTGAC TACGCTGACG GCCGGGCTG GGGTGTTCG GCTGACGCGG | 1560 |
| TAACCATGTC GCTGCTGCTG CTCCGATGCA CATGCGCTG CGGGTGTTCG TTACGCTG | 1620 |
| GTGCTGCTGAC ACGCGGACG TAAAGCTGAT CAGCAAGCAG TGGGGAAGG TGGGGAAGG | 1680 |
| TAACCATGTC GTCTGCTGCTG CTCCGATGCA CATGCGCTG CGGGTGTTCG TTACGCTG | 1740 |
| TGGCCAGAAG GGCTGCTGCC CTGGGCTCCAA AGTCTGCTGCTG CTGGGCTCCAA AGTCTGCTGCTG | 1800 |
| CATGGCTGAC GTACGTTCAG GACTGGGAGA CAGTTGGGAG GATGCAAGG ATGTCCGGT | 1860 |
| CGCCTCTCGT AAGTTGCTAC CGCCCGCGTG CGTAAATGCA AAGTTGACAG GTGTCCGCGT | 1920 |
| CTGCTGCTGAC TGGCAGCAG TGGGCCAGG TGGCAGCAG TGGCAGCAG TGGCAGCAG | 1980 |
| GGCCTGAGAT GGGTGGTCTG GTCGGCAGG GTCTGCTGCTG CTCCGATGCA CATGCGCTG | 2040 |
| CATCTGCTGAC TGTGCTGACG AGAAGATCAAA AGAGCTCTGGA AGAAGATCAAA AGAGCTCTGGA | 2100 |
| GCCGAGAAGG AGACGACG TGGGCGTGACCC GGGGCTGACG GCCGGGCTG GGGTGTTCG | 2160 |
| CAGCTGCGCG AATGCTGACG TGGCAGCAG TGGCAGCAG TGGCAGCAG TGGCAGCAG | 2220 |
| GCTGCTGCTG ACAGTTGGCA GAGGTTGGG GCTGCTGCTG CTCCGATGCA CATGCGCTG | 2280 |
| AGTGTGAGG GATACAGGATG TGACCAGGAC AGGAGCTGGA TGGTCCGCGT GCTGCTGCTG | 2340 |
| GCTGCTGCTG ACAGTTGGCA GAGGTTGGG GCTGCTGCTG CTCCGATGCA CATGCGCTG | 2400 |
| GGTGCTGAGA GATGCTGACG TGGGCGTGACCC GGGGCTGACG GCCGGGCTG GGGTGTTCG | 2460 |
| GGTGCTGAGA GATGCTGACG TGGGCGTGACCC GGGGCTGACG GCCGGGCTG GGGTGTTCG | 2520 |
| TCTGCTGACG GCCGAGAAGG AGAAGATCAAA AGAGCTCTGGA AGAAGATCAAA AGAGCTCTGGA | 2580 |
| GATGCTGCTG ACAGTTGGCA GAGGTTGGG GCTGCTGCTG CTCCGATGCA CATGCGCTG | 2640 |
tgatgcaac aacaggagt ctggtaatcg tcttaacttt gtcaacacgc aagacggtag 2700
ccacctctgtg aagacttgtag tggggcaca cgagaagttc gpaatcgaaat ctttgcact 2760
gatccagct tcccttggta cggatccggc tgaagctgtag aacctgtagc aagacggtcgg 2820
cgaacactaagtgcaactat atgactccttg tgaatgactg tctgaatcttt cagacacgtt 2880
gctgacagct tggcaactgg ctcaatgagca cacaattgcaac gcctctcggg ctaaagtagtta 2940
catttaacgctgtagacaccttagatcggtcc ctctctctcagc tctctctcgtctctctcgtct 3000
tggcaggtct tggcaacagca acacagtgagc cctctctcggc tgggtgatgc gatgtgactat 3060
gtatctggtcttgccgcatc gcggcccaac gcagagattg atacccggtc tggcacttcc 3120	tgctgcaagat taaacgtaact cctgtagtct tggagttcagt 3150

<210> SEQ ID NO: 277
<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: 277

gaattaata gcaagttatct taaaatatctt tagaaahaaa ctaagtcgagc gttctgataata 60
gcagaagtttagatcttggagctgatcgtcctttgctgaacttggtctta 120
ttttt 125

<210> SEQ ID NO: 278
<211> LENGTH: 8452
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 278

tgcggtaatgt cacaccgcaat caggtgcccg ttttccggga aatgtgcgcg agaaccctat 60
	tgtgttatttt tcttatatac atcataatct tggcgcccct cggagggatat taaaagatatt 120
cctccacccag aatatcatttt cctctctttt tcaatttctaa tcatatattga 180
gtaaacgtg ttcggcgagt aacagtctt aatcgctag gcacactat cagpagatctg 240
tcttagctg tttttatttt tgcggcccgg cccccccccgt cggaggttgg ggccccgctt 300
gagtttaaa ccaataaccc acgagcggg agggcgccg cagcagagg gcttcgaccc 360
tttattcctc tctccacgt ttaatttta ctgccccccc gttattgtaga ttaatgtcctc 420
gtttagtgg attcagttg tggccccccc ccaagagacg catcagacgc catcagcctcc 480
cagtttagt tggcccggag tggammagat tttggtgatgcaatcttcttgcttcgct 540
gttcggtctc gttggtagt ggttcctcag cccccccctt gtcggtttgt gcacgagctc 600
catcgtgtgc aaaaaaaaaag ttcgctctc gtcgtgcttt cagttgtag ttaatgtcctc 660
gcgagctgc ttaaaatccttt ctggttggct gcggcgcctt tggggcaagt gtaataactgc 720
tacgtagat gttgttgtgct gtaactggta atacccgcaat ctaattttttt cctctctctc 780	tagtcgtgtggt ctcgcccgg gtaataaggc cgcagcctag 840
<table>
<thead>
<tr>
<th>Site</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>900</td>
<td>cagaaccttta aaagtgtca tcattgaaa aagttctctg ggcccacaaac tctcaagggat</td>
</tr>
<tr>
<td>960</td>
<td>ctaccgtct ggagatacga gttggatgta acaacgtctgt gcacaccaact gatottcagoct</td>
</tr>
<tr>
<td>1020</td>
<td>atcttttaact ttcacacgcgt tttctgggtg aagcaaacaaa gcagagccaaa atgcggcaca</td>
</tr>
<tr>
<td>1080</td>
<td>aaagggaaaag gccagcaac gccagatttg aatactcata ctctctcttct tcaattattata</td>
</tr>
<tr>
<td>1140</td>
<td>ttcgaacatt ttcacaggttt atttgtccat gcaccaaatc ctttaacttg aagtttttgtt</td>
</tr>
<tr>
<td>1200</td>
<td>ccactgacg ctcaccagccg tagaaagat ctaaaggtct tgttggagatct tttttttttctt</td>
</tr>
<tr>
<td>1260</td>
<td>gcgcgtaact tgcgtggctgg aacaaaaat gcacagcaata ccagcggttg cttgggttgc</td>
</tr>
<tr>
<td>1320</td>
<td>ggatcaagag ctcaccaaccc tttctccgaag gttacgtggc tcacagcagag cgacagataacc</td>
</tr>
<tr>
<td>1380</td>
<td>aaatctegtt cttctaggtt agcgcagttt agcgcacacc ttcagaaact ctgtgacacc</td>
</tr>
<tr>
<td>1440</td>
<td>ggtctctacc gccggtgtgcat caagcaagata gttctçaggg aagggcagctgg ggtgggtcctg</td>
</tr>
<tr>
<td>1500</td>
<td>aacgaggggt gtcgctgcac agccacgctt gcagcgacagc acctacaccc aactcaagata</td>
</tr>
<tr>
<td>1560</td>
<td>ccctacaggt gcgctagtag aagagccaccc gttcctgca aagagagaa aggcacagata</td>
</tr>
<tr>
<td>1620</td>
<td>ctcggtaagc gcggaggggt gcacagagca aggctcagag gcgagctcagag ggggaacaagt</td>
</tr>
<tr>
<td>1680</td>
<td>ctcgggtaccct ttcggtgtgcc ccaccctgta gcaggggtgtag gtttgggttg</td>
</tr>
<tr>
<td>1740</td>
<td>attgcgatccg ggcggctagag gttatggatag cctgagcag ptcacagcagag ggggaacagag</td>
</tr>
<tr>
<td>1800</td>
<td>cttggacttt tgcggccttt tgcgtcagcat gttctctctg tgcgtcatcgc gctaattacgt</td>
</tr>
<tr>
<td>1860</td>
<td>ggataacgtt attaccagct ttcagtgatg tgcagctgct gcgcgccagcc gacgcacagca</td>
</tr>
<tr>
<td>1920</td>
<td>gcgcgagcgcg ttcggtagcc aagggcaggtt ccagctgtgg tgcagctgatag gttgctgatag</td>
</tr>
<tr>
<td>1990</td>
<td>tttgtgtcgt ttcagtcacag ctgtgccgtg cggcagcttg ggggagaagtg cttggaataag</td>
</tr>
<tr>
<td>2040</td>
<td>gccgccagcg gccggcgatgt gcacacaggag ctcacagcgg cccgtcggcc gcaccctcaca</td>
</tr>
<tr>
<td>2100</td>
<td>ggttaaaag cccgagcacc gcagagtgta ccctcacttc cgcgcaacaac gcacgacatg</td>
</tr>
<tr>
<td>2160</td>
<td>tacatactg gtacctctgc gcgcacacat aacactcaac tgcattcaag tgaacccaca</td>
</tr>
<tr>
<td>2220</td>
<td>gttgctggcc gcggcgccgg aagagagggg cgcacacacc aggatagtgt tttgatggtg</td>
</tr>
<tr>
<td>2280</td>
<td>atttacgag ttcgcaaccttatcgcggga ccctggtgggt gacaagaggtt gcggcacaac</td>
</tr>
<tr>
<td>2340</td>
<td>gcacagacgt ccgtagcaac cccgtgagcc tgcgccctct gataaaacgg cgcaggggccc</td>
</tr>
<tr>
<td>2400</td>
<td>tatggtttct tttttttttca aagagagatgc actcacaact taaatggtg acagtgatgc</td>
</tr>
<tr>
<td>2460</td>
<td>ggcgaagcgg cccccggagtg cacggcagct gtctggccag tggccctgca aogcccgcct</td>
</tr>
<tr>
<td>2520</td>
<td>tgtgcgatgc aagttcttttg gtcacctttgt gcgcgcgtcga acgcagctat aerctgctg</td>
</tr>
<tr>
<td>2580</td>
<td>cgccttcttc atttccaggg gattcgaggg actatgtact cctgcagtgc tccagacattc</td>
</tr>
<tr>
<td>2640</td>
<td>ggtgccagca agaaagacgg caagggcgaa ggtgctacca aagagcagct catogcctgc</td>
</tr>
<tr>
<td>2700</td>
<td>gacactgcgtg ccaactggt ggcgtcggct gcgtatcaggg cgagagcacaag ccagttgctg</td>
</tr>
<tr>
<td>2760</td>
<td>aagagacactc acgggtcgtg gcagccagac cgccgacaccc gcgtgaagag acgcggcaca</td>
</tr>
</tbody>
</table>
| 2820 | gcgcctgcgt gcctgagccgg caagagcagc gcgtacagcg gcgcgacccc cggtggag 
| 2880 | aagagacactc acgggtcgtg gcagccagac gcgtacagcg gcgcgacccc gcgtggag |
| 2940 | atgcgcgagg tgtgacgagc cttctcctccac agactgtggag acgttgttcct gcgtggagag |
| 3000 | gataagacgc gcagagcgcc ccccatcttc gcagaacatgc tggagagag tcgcagtcac |
| 3060 | gagagacactc acgggtcgtg gcagccagac gcgtacagcg gcgcgacccc gcgtggag |
| 3120 | gagagacactc acgggtcgtg gcagccagac gcgtacagcg gcgcgacccc gcgtggag |
| 3180 | gacgctggggct gcgcgtctct gcgcgacccc gcgtggagag ccctgccctc tgctgagag |
| 3240 | atcggaggggcc gcacacagcc gcgtgagggg gcgcgtctct gcgcgacccc gcgtggag |

This page continues from the previous page.
cagacccgaca accagctgtt cgaggaaaac ccaccaagcc ccaagggcgc ggacgcacag 3300
gccatcctct gtgcagccag cacgaacagc agacaccgag ggcaatgcct gaaacactcg 3360
cgccgaaga agaagaatgc cctggtcagc aacgctagtt cctggtcagc ggaccggacc 3420
cacaatcctga agaaggtcaga cagcgtgagc gagaagtgaa aacctgcaag cagcaagcag 3480
acccatcagc acagctgccg cgaagcagcc cgaagtgagc cagcgtgacc cagcgtgacc 3540
cttctgcggcc ccaagaaact ctcgcactgc aacctgctga gagaagcatt gagaaggtac 3600
accagatag ccacagtccg cccagacgcc cctggtcagc tctggtcagc cctggtcagc 3660
cagcacggcc accatgtcag cagcgtgagc cagcgtgacc cagcgtgacc cagcgtgacc 3720
acttatcttg accagcgcagc gaagctgtac cccaggtcct cctggtcagc cctggtcagc 3780
gagaagttc ccaaggccact ctggaanagc tggagcggcc caggagaatg 3840
cctgcgaaa gcagacgcac cggacgcac gcagacgcac gcagacgcac gcagacgcac 3900
atccacagcc aagatcagcgt cagcgtgacc cagcgtgacc cagcgtgacc cagcgtgacc 3960	taaccctcct cagagacccag cggagacgcag gcagacgcag gcagacgcag gcagacgcag 4020
actcagtcg gcccctgtgc cccagacgcc cccagacgcc cccagacgcc cccagacgcc 4080
aggacacagg taccgccggtgc ccaagacgcc ccaagacgcc ccaagacgcc ccaagacgcc 4140
agtctctacgc agaggacgcac ccaagagcttc cccagacgcc cccagacgcc cccagacgcc 4200
aagcacgacgc acctcccttggctc cctggtcagc cctggtcagc cctggtcagc cctggtcagc 4260
gttggagccct caagagcgct cccaggtcct cctggtcagc cctggtcagc cctggtcagc 4320
gaatggtgtgc cccagacgcc cccagacgcc cccagacgcc cccagacgcc cccagacgcc 4380
aagagaatcg atctggtgc cccaggtcct cctggtcagc cctggtcagc cctggtcagc 4440
tccctgggac catatcagaag cctggtcagc cctggtcagc cctggtcagc cctggtcagc 4500
gagagacacgc cccagacgcc cccagacgcc cccagacgcc cccagacgcc cccagacgcc 4560
aggtggtgctg cccagacgcc cccagacgcc cccagacgcc cccagacgcc cccagacgcc 4620
cccgaggtgc cccagacgcc cccagacgcc cccagacgcc cccagacgcc cccagacgcc 4680
acctgggaag ccaagcggggc cccaggtcct cctggtcagc cctggtcagc cctggtcagc 4740
aagcagacgg cctggtcagc cctggtcagc cctggtcagc cctggtcagc cctggtcagc 4800
ccacgggacta cccaggtcct cctggtcagc cctggtcagc cctggtcagc cctggtcagc 4860
agcagacgg ggttcgggac cccaggtcct cctggtcagc cctggtcagc cctggtcagc 4920
gtcgagagc ggttcgggac cccaggtcct cctggtcagc cctggtcagc cctggtcagc 4980
accagagactc caagacagct cccaggtcct cctggtcagc cctggtcagc cctggtcagc 5040
gggccagggag ggcagagactc caagacagct cccaggtcct cctggtcagc cctggtcagc 5100
aagctgatc cccaggtcct cctggtcagc cctggtcagc cctggtcagc cctggtcagc 5160
ataccaccgc cccaggtcct cctggtcagc cctggtcagc cctggtcagc cctggtcagc 5220
gtcgtccggc cccaggtcct cctggtcagc cctggtcagc cctggtcagc cctggtcagc 5280
acccaccagc cccaggtcct cctggtcagc cctggtcagc cctggtcagc cctggtcagc 5340
aagccgtttcgccccgtttc cctggtcagc cctggtcagc cctggtcagc cctggtcagc cctggtcagc 5400
gacgccgag cccaggtcct cctggtcagc cctggtcagc cctggtcagc cctggtcagc 5460
cgaccgggac ccagacagct cccaggtcct cctggtcagc cctggtcagc cctggtcagc 5520
ctcggatcc cccaggtcct cctggtcagc cctggtcagc cctggtcagc cctggtcagc 5580
tccccagt
aacaagtgcc
cggagcaac
aacaccaaac
acgcaccaag
cgctacaact
5640
aagcagctgg
tgggaaacgg
ccctgatcaa
aagaactcta
gactggaag
acaggtcttg
5700
tacggaagct
acagaatgta
cgagctggag
aagatgacag
ccaagacgca
gcgaaatcc
5760
ggcagcgtta
cggacacta
tctacattgt
caaacagcag
5820
attacccctg
ccacagggcg
atcggcgacg
agcagcagctg
tccgaggtct
5880
gggagagctcg
tggcgtgaaac
tgctggggtat
tccgacaccg
tgcggaagggt
gctggacctg
5940
cccacagtag
attgctggaac
aaagacggag
gtcgagacag
gccgcttcaag
caaagatct
6000
atcctgcccc
agaggaacga
cgataagctg
atccgacgag
gaaagacttg
6060
aagtaagcggc
gttgcaacag
cccacagctg
gctatcttgct
tggtgcttct
gccaaaagt
6120
gaaagggcaca
tcagagagtt
gtaagaggct
tgtcgtgctgg
tcaccattag
caccactatcg
6180
gaaagaagcct
gttgacagga
gaatcctcct
gaatctcttg
aagccaaagg
ctcaaagaa
6240
gtggaaagat
ccagcatcat
aaagtctccc
tgtgagcagct
gaaagacggc
6300
cggagagagac
tgcgcgtggtca
tgagcagagac
tgggcttccg
6360
tccaaatagct
tgacacccct
tgacacagctg
gagaattgga
6420
ggagcataag
cgcagcaacc
gtgctttttgg
caacagcaca
ggcacactct
gacagaggtc
6480
attcgcgaga
tcaagctgtt
tccacagaga
gtacatctgg
ccgagcctaa
tctgcaaca
6540
gtgcgttcgc
ccctcacaaca
gacccgagat
aagccctctca
gagagcgagc
gagatactc
6600
atccacctgt
ttcctcgtag
aatgtggtgga
ccgcgcgctgg
tctcataagta
tctgcaacac
6660
acccagccg
ggagaggttg
cacccagctc
aaagaggtgct
tgacgaccac
tctgatccac
6720
cgagacacta
cgccgcgtga
gagagacgg
atgacgctgt
tcagcgctgg
aggggacagc
6780
ccacagagac
gagaaagctg
ggagccgcag
taactagata
ctcgatgtggc
6840
tatgacaecc
tcagcacaag
tgagggcgctg
tgcgagacgct
gctgccgctg
6900
acagcagata
tgctggccgg
cccaaacagct
tggccggggc
tgcctagggc
tgctggacgc
6960
gttccagggc
gacgcgtcttg
taaataagcca
gccgccgctat
tcgaaacaca
7020
tcaccacagct
tatccagcct
tacaactctt
tcactacaggg
tactgcagct
7080
tgtgatgcct
tcctgcagat
ggcggctcct
tctctctctg
tctacgatca
taccgcaca
7140
catgcacaca
gacaacctgt
tccttctctga
cgttgggtagtt
ctccagcgtag
7200
cctcgagcag
gacagggagt
tgctgggggt
gagtgcagca
gcaagcccg
cggagacgg
7260
agcgtgtcgg
caggttgacc
tctgcaacgc
gcaggctgtg
tgggtggttcc
7320
tctgtgtgcg
tctccagcag
catacctccc
tgcgtgctcg
tctctctcttg
tgcggtgcctg
7380
ccggcgaggc
tctggagagc
tgctgtggagc
tgggtgccctg
tggcggcggc
tggtgctggc
7440
agcggcacaac
cctgctactg
tgctggcgagc
cgaccgcggt
tgacagcagct
tcgggctagac
7500
ctattgctcg
tcgggaggac
tcccgtgaga
tgctggctggc
tgggtgcttg
7560
cccgggtgcg
gggggtcggct
tcctctgcggc
gcgggacgct
ggccgctggc
acggaggtct
7620
gccggtggcc
tctacctggtg
tagacgcca
tgacgaccac
tacccgagg
tcggcagatgg
7680
acgcgcagac
gacggcagcag
gcgtctctggc
tctggtcgcctg
gaaagctgga
7740
gcggggtgcg
tcaggggagg
ctacgagcgcc
acaagcgtcct
cacccgcaact
tcagagatgt
7800
tccgggacaat
tccggggaga
gccgcggcgg
cagccgcgtggacacgacg
gggtggtggt
7860
acccgctcgg
cggggtgctt
gacgctcttg
gcgaaagctg
tgcgagcagc
7920
tccggggcgg
gaagaGGcctg
2ccgtctcaag
gcagagctgca
ggggaccaac
atccatcttg
7980
aacctggccgc gaccgaggtc acgggatcg tcgaatctcct cacgaogtmtat gcgggagact 8040
cccgcgtcag ccctgtgcaaa tctccatctca acggccttccg ggggagacgc gagatctcttg 9100
cgacgctgct gcacggcggc gacctgaaac gcacgccgaa ctctggcgcgc gcagctgctcg 9160
ccttccaacctt ccctgcaaac tctgatcgttc tcggagagac cccctgtaggt ctctccggtc 8220
tccgacattcg gcacaggtc cctgcgggctt cctggggtccc gcggagaccc gcggcgcggc 9280
cctgataagg atcggggaact gctgcccccg ctctgcaacc ccaacagtgg caccccctctctc 8340
gctgtttgttg gatgtcagct atgtatctcg gtctttggcct aacgggcacac cgaagcagat 8400
tgataccgag ctcgggacctt cctggtaaga tgtacgctca gttgtaggta ctgat 9452

<210> SEQ ID NO 279
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 279

gttttagaac tattgctttg tataatgtc ccaaaacggcag gggctgtagt cgagcagcaa 60
gaaagaattt tagaatagtt gtttttgtgc tggtcctaa aac 102

<210> SEQ ID NO 280
<211> LENGTH: 100
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 280

cgagggcagc agtcaanacgc gcacagcctg gaggaggaag ggcctgtagtc cgagcagaa 60
aagaaagggct ccataccat caacggttg gcacgctgca 100

<210> SEQ ID NO 281
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 281

agctggagga ggaaggggtct gatctcggagc agaagaagaa gggctcccac 50

<210> SEQ ID NO 282
<211> LENGTH: 30
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 282

gagucagcagc agaagaaga guuuuagagc 30

<210> SEQ ID NO 283
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 283

agctggagga ggaaggggctct gatctcggagc agaagaagaa gggctcccat 49
<210> SEQ ID NO 284
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 284

cggagaggg aagggctgag actggagacag aagaagaggg gactccatca cat

<210> SEQ ID NO 285
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 285

cggagaggg aagggctgag actggagacag aagaagaggg gctccatcac at

<210> SEQ ID NO 286
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 286

cggagaggg aagggctgag actggagacag aagaagaggg ggtcccct acat

<210> SEQ ID NO 287
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 287

cggagaggg aagggctgag actggagacag aagaagggc ccatcacat

<210> SEQ ID NO 288
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 288

cggagaggg cccgagacg aagggctccc atcacat

<210> SEQ ID NO 289
<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
<220> FEATURE:
<221> NAME/KEY: modified base
<222> LOCATION: (1) ... (20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 289

nnnnnnnnnnnnnnn guuuuuagac uaquaaagc aaguauaau aaggtgtc 60

cguuuu

<210> SEQ ID NO 290
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

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

SEQUENCE: 290
gaguccgac agaagaagas

SEQ ID NO 291
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 291
gacaucaug uccuccocau

SEQ ID NO 292
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 292
guccuccca augaucaggg

SEQ ID NO 293
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 293
auugggugu caggccagag

SEQ ID NO 294
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 294
guggagcagag ggcccgagau

SEQ ID NO 295
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 295
gggccgaga uuggguguu

SEQ ID NO 296
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 296

gugccaua guaaugccau

20

<210> SEQ ID NO 297
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 297

guacaccoca cagugccacag

20

<210> SEQ ID NO 298
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 298

gaagccucu ggcgccagaa

20

<210> SEQ ID NO 299
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 299

cgagaggg aagggcctga gtcgacagc aagaagagg gtcgccact

48

<210> SEQ ID NO 300
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 300

gaguccgagc agaagaagau

20

<210> SEQ ID NO 301
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 301

gaguccgagc agaagaagau

20

<210> SEQ ID NO 302
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 302
gaguccgagc agaagaacaa

<210> SEQ ID NO 303
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 303
gaguccgagc agaagaagaa

<210> SEQ ID NO 304
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 304
gaguccgagc agaaguagaa

<210> SEQ ID NO 305
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 305
gaguccgagc agaauagaa

<210> SEQ ID NO 306
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 306
gaguccgagc acasaagaas

<210> SEQ ID NO 307
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 307
gaguccgagg agaagaagaas

<210> SEQ ID NO 308
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 308

```
gaguccgagc agaagaagaa
```

<210> SEQ ID NO: 309
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 309

```
gagucggagc agaagaagaa
```

<210> SEQ ID NO: 310
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 310

```
gagaccggagc agaagaagaa
```

<210> SEQ ID NO: 311
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 311

```
aagtacaagc ttgtcagcgg tggg
```

<210> SEQ ID NO: 312
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 312

```
aaaacggagc ggccgagtc cggacgaga agaagtttt
```

<210> SEQ ID NO: 313
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 313

```
aaacaggaggc cgagatgggg tgttcagggc agaagtttt
```

<210> SEQ ID NO: 314
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 314
aaacggaag ggcctgagtc cgacgacaag aagaagttt

<210> SEQ ID NO: 315
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 315
aacggaggg gggcgacaga tgagaaactc agggttttag

<210> SEQ ID NO: 316
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 316
agcccttct ctctgtctcg gactcagggc ctctctcc

<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
cagggaggg gggcgacaga tgagaaactc aggaggcccc

<210> SEQ ID NO: 318
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 318
ggcagctgac caccggttga tgtgatggga ggccttctag gagggccccca gacggaccc

<210> SEQ ID NO: 319
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 319
tggggcttca acacctaggg

<210> SEQ ID NO: 320
<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: 320
gtcaccttca atgactaggg tgg
<210> SEQ ID NO 321
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<226> 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: 321
aaccnnnnnn nnnnnnnnn nnnnn 25

<210> SEQ ID NO 322
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 322
catcgatgct ctocccacttg gctgttctcg tgg 33

<210> SEQ ID NO 323
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 323
ttcgtggcaat tgccccacccg gttgatgtga tgg 33

<210> SEQ ID NO 324
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 324
tggtgcaat ggcgccacccg ttgatgtgat ggg 33

<210> SEQ ID NO 325
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 325
tcagctttct gcgtttgtga ctgtgtcttc cgg 33

<210> SEQ ID NO 326
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 326
ggggaggg gcacagatga gsaactcagg agg 33

<210> SEQ ID NO 327
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 327
taggggcag atgggttgtt cagggcagag agg 33
<210> SEQ ID NO 328
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 328
aacacgggctttgagaag acctgtttta gagctagaaa tagcaagtt aaat 54

<210> SEQ ID NO 329
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 329
caaaacgggctttgagaag acctgtttta gctatgtgt tgtgaatggg coca 54

<210> SEQ ID NO 330
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<400> SEQUENCE: 330
cagcacgtgagtgcaacatcactgtcat agg 33

<210> SEQ ID NO 331
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<400> SEQUENCE: 331
aatgcattggtaccaccaca caggtgccag ggg 33

<210> SEQ ID NO 332
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<400> SEQUENCE: 332
acacacatg gaaagcctct gggccagga agg 33

<210> SEQ ID NO 333
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 333
gagggagtga gtatacagaa acacagagaa gtgaaat 37

<210> SEQ ID NO 334
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 334
agaatagta ggagtcagaa aaactcagca ctagaa 37

<210> SEQ ID NO 335
<211> LENGTH: 96
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 335

ggacgaaaca cgggaaact tcaaacaac atagcaagtta aaaaaaaggg ttagtccgta
60
tcaacggtgg aaagtggcac cgatcgcgtg gtttttttt
98

<210> SEQ ID NO 336
<211> LENGTH: 186
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polyonucleotide

<400> SEQUENCE: 336

ggacgaaaca cgggtagtat taagtattgt ttatggtct ataataatctt ttgaattttct
60
cctgtgatat tgtgtataaa agttataaaa taactctgtgt ggaaccattc aaacacgcat
120
tagcagtttaa aataaaggct gaatacgttc aacttgaaaa aagtggcacc aagttgggtct
180

tttttt
186

<210> SEQ ID NO 337
<211> LENGTH: 95
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 337

gggtttttaga gctatgcgtt ttgaattgt ccgcasecgg gctttcggaga aagctttta
60
gagctagcgt gttttgaatt gttccccaaaac ttttt
95

<210> SEQ ID NO 338
<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: modified_base
<222> LOCATION: (5) (34)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 338

aasnnnnnn nnnnnnnnn nnnnnnnnn nnngt
36

<210> SEQ ID NO 339
<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: modified_base
<222> LOCATION: (7) (36)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 339

taaaacnnnn nnnnnnnnn nnnnnnnnn nnnnnn
36
<210> SEQ ID NO 340
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 340

gtgsgaagga cgaaacaccg ggtcttcag aagacctgtt ttagagctag aataagcaag

60
ttaaaataag gctagctcgt tttt

84

<210> SEQ ID NO 341
<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: (6)...(24)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 341

cacgnnnnn nnnnnnnnn nnn

24

<210> SEQ ID NO 342
<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: 342

aacnnnnnn nnnnnnnnn nnc

24

<210> SEQ ID NO 343
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 343

gtttttagag tatgcgttgt tgaatggtcc caaaaactag caccaggtc tgcctttaga

60
gctagctgt tttcatagtt ccacaaaac

88

<210> SEQ ID NO 344
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 344

aagcggagg gctgaggtcc gacgagaaga agaag

35
<210> SEQ ID NO 345
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 345

aaaccttcctctcactgctgccc ctccc

<210> SEQ ID NO 346
<211> LENGTH: 46
<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) .. (19)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<400> SEQUENCE: 346

nnnnnnnnnnnnnnnnnnn nnnnnnnnnn uuaauggacc cuacgaauuu uuuuu 46

<210> SEQ ID NO 347
<211> LENGTH: 91
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 347

guaucuaa ucuuucac ucuuacaaa uacggcuuauc ugcgcgaauuc aacaccccguu 60
cauuaauagg caggguguau ucguuaauua a 91

<210> SEQ ID NO 348
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 348

ttttctagtg ctgagttttct gtgacctccc tacattctac ttcctgtgt ttcctgtatac 60
tacccctcctcc

<210> SEQ ID NO 349
<211> LENGTH: 122
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 349

ggaggaggg cctgatgtcog agcagaagaag agggtgctcc catcacatc aaacgtggcgc 60
cattgcccag aagcaggcca attgggagga catcgatgc acctccaaag actagggtgg 120
gc 122

<210> SEQ ID NO 350
<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: (3) to (32)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 350

acnnnnnn nnnnnnnn nnnnnnnn nguuuuaga gcuagcu

<210> SEQ ID NO 351
<211> LENGTH: 67
<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

<400> SEQUENCE: 351

agcauagcau guuuuuuuu ggtagucgc uuaucacuu gaaasagugg caccgagucg
uguuuu

<210> SEQ ID NO 352
<211> LENGTH: 62
<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) to (20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 352

nnnnnnnn nnnnnnnn nnnnnnnn guuuuagac uagaauagc aaguuuuuu aagguuuug

cg

<210> SEQ ID NO 353
<211> LENGTH: 73
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 353

tgaatggtc caaaaagcc gggctctgagt cctagcagaagagattg tagatctcgag
ctgttttgcagatc

<210> SEQ ID NO 354
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 354

tgccctttc cactctctgc cctgaaccc cattctcgc cccttcgc cactctctgc
atccctggt

<210> SEQ ID NO 355
<211> LENGTH: 139
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 355

tgccctttc cactctctgc cctgaaccc cattctcgc cccttcgc cactctctgc
acccagcag tgaagtcgct tagctaaatg catagggta caccacacag gtccagggg
60
ttttccccaa gtccccggcg cctttcctca cctttccttg gcccagggct tttccatgtg
120
tgtgcttggc cccctttga
138

<210> SEQ ID NO 356
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 356

gtgccttggc gaggcctacc
20

<210> SEQ ID NO 357
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 357

ccttgagcgc atgcagtagt
20

<210> SEQ ID NO 358
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 359

acctctggtg ttctccacat tc
22

<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

tggggactg cacagacctc
20

<210> SEQ ID NO 360
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 360

ggctcccttg gtccaaagtta
20

<210> SEQ ID NO 361
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
primer

agagggtct ggtatgctgta a 21

SEQ ID NO 362
LENGTH: 30
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic probe

SEQUENCE: 362
tagctctaa actctttttg ctgcctgagac 30

SEQ ID NO 363
LENGTH: 30
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic probe

SEQUENCE: 363
tatcttttat ttaactttgc tatgctgttt 30

SEQ ID NO 364
LENGTH: 99
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
NAME/KEY: modified_base
LOCATION: (1)...(20)
OTHER INFORMATION: a, c, u, g, unknown or other

SEQUENCE: 364
nnnnnnnnnnnnnnnnnnnnnnnnguauagac uagaaauagc aaguanauaaau aagcuacaguc 60
guauacac uagaaauagu gcacccagag cggugcuuu 99

SEQ ID NO 365
LENGTH: 12
TYPE: DNA
ORGANISM: Homo sapiens

SEQUENCE: 365
tagcggstaa gc 12

SEQ ID NO 366
LENGTH: 12
TYPE: DNA
ORGANISM: Homo sapiens

SEQUENCE: 366
tcggtgcata cat gt 12

SEQ ID NO 367
LENGTH: 12
TYPE: DNA
ORGANISM: Homo sapiens

SEQUENCE: 367
actcccgta gg 12
<210> SEQ ID NO 368
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 368

agtgcgtgat 12
<210> SEQ ID NO 369
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 369

agtcgcctg at 12
<210> SEQ ID NO 370
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 370

tagtcgacc ag 12
<210> SEQ ID NO 371
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 371

ggcgtaatgt at 12
<210> SEQ ID NO 372
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 372

tgtgcagtgta 12
<210> SEQ ID NO 373
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 373

atggaaaccgc at 12
<210> SEQ ID NO 374
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 374

gccgaattcc tc 12
<210> SEQ ID NO 375
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 375
gcatggtacg ga 12

<210> SEQ ID NO 376
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 376
cggtaccttt ac 12

<210> SEQ ID NO 377
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 377
gcctgtcccg ta 12

<210> SEQ ID NO 378
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 378	tacggttaagt cg 12

<210> SEQ ID NO 379
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 379
cgcaaattta cc 12

<210> SEQ ID NO 380
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 380	aaccaagata cg 12

<210> SEQ ID NO 381
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 381
gattgcgtac gc 12

<210> SEQ ID NO 382
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 382
gtctcagat cg 12

<210> SEQ ID NO 383
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
tggtcggtgca

aactccgtagtga

caggacgtcctgc

tcgatatcccatc

tttcaagggcggg

cggcgggtggaat

gaacgcgtctcata

gatctcacgcg

gttcatcagcg
<400> SEQUENCE: 391

acaccggtct tc 12

<210> SEQ ID NO 392
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 392

atcgtgccct aa 12

<210> SEQ ID NO 393
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 393

gcgtcaatgt tc 12

<210> SEQ ID NO 394
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 394

cctccgatatct cg 12

<210> SEQ ID NO 395
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 395

cggatcttcct cg 12

<210> SEQ ID NO 396
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 396

tgcgctctca gt 12

<210> SEQ ID NO 397
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 397

taacgtcgga gc 12

<210> SEQ ID NO 398
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 398

aaggtcgccc at 12

<210> SEQ ID NO 399
<211> LENGTH: 12
<212> TYPE: DNA

<400> SEQUENCE: 399
<210> SEQ ID NO 400  
<210> SEQ ID NO 401  
<210> SEQ ID NO 402  
<210> SEQ ID NO 403  
<210> SEQ ID NO 404  
<210> SEQ ID NO 405  
<210> SEQ ID NO 406  
<210> SEQ ID NO 407  
<211> LENGTH: 12  
<211> LENGTH: 12  
<211> LENGTH: 12  
<211> LENGTH: 12  
<211> LENGTH: 12  
<211> LENGTH: 12  
<211> LENGTH: 12  
<211> LENGTH: 12  
<212> TYPE: DNA  
<212> TYPE: DNA  
<212> TYPE: DNA  
<212> TYPE: DNA  
<212> TYPE: DNA  
<212> TYPE: DNA  
<212> TYPE: DNA  
<212> TYPE: DNA  
<213> ORGANISM: Homo sapiens  
<213> ORGANISM: Homo sapiens  
<213> ORGANISM: Homo sapiens  
<213> ORGANISM: Homo sapiens  
<213> ORGANISM: Homo sapiens  
<213> ORGANISM: Homo sapiens  
<213> ORGANISM: Homo sapiens  
<213> ORGANISM: Homo sapiens  
<400> SEQUENCE: 399

GTCGGGACT ATT

TTCGAGCAGT TT

TGAGTCGCG AG

TTAGCAGAGG

AGAAGTATCG GC

AUCTGATAACC AT

CAGCTACATAG CA

TTCATACCG GC

AGAAGTATCG GC
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 407

cgaacagtt aa 12

<210> SEQ ID NO 408
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 408
cgatctccttc gt 12

<210> SEQ ID NO 409
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 409
cgtctgtgat aa 12

<210> SEQ ID NO 410
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 410
agtggcagtg ac 12

<210> SEQ ID NO 411
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 411
ccctacggc ac 12

<210> SEQ ID NO 412
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 412
gccacccgc ac 12

<210> SEQ ID NO 413
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 413
tggacacccgt 12

<210> SEQ ID NO 414
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 414
ttgactgcgg cg 12

<210> SEQ ID NO 415
actatgcgta gg 12

tcaccccaag cg 12

goaggacgtc cg 12
	
acaccgassa cg 12

cggtgatttg ag 12

cacgagrat gc 12

taaagcggcc cg 12

cctagcgcc ca 12
<210> SEQ ID NO: 423
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 423
cgaaacgtg gc

<210> SEQ ID NO: 424
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 424
cgtgcctga ac

<210> SEQ ID NO: 425
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 425
tttaccatcg aa

<210> SEQ ID NO: 426
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 426
cgtgcctgt tt

<210> SEQ ID NO: 427
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 427
cocaaacgtt ta

<210> SEQ ID NO: 428
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 428
ggttatcag aa

<210> SEQ ID NO: 429
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 429
tcgatgstaac ac

<210> SEQ ID NO: 430
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 430
cgactttttg ca
<210> SEQ ID NO 431
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 431
togacgactc ac 12

<210> SEQ ID NO 432
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 432
acgctcaga ta 12

<210> SEQ ID NO 433
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 433
cgtaacgcac ag 12

<210> SEQ ID NO 434
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 434
catatccttg ca 12

<210> SEQ ID NO 435
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 435
cgctcagct at 12

<210> SEQ ID NO 436
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 436
aasatcggta gc 12

<210> SEQ ID NO 437
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 437
ccttcgaagg ag 12

<210> SEQ ID NO 438
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 438
gtcgtgacct ac 12
<210> SEQ ID NO 439
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 439
ggtgtctactc aa 12

<210> SEQ ID NO 440
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 440
gttaacaggc tg 12

<210> SEQ ID NO 441
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 441
tagctaaccg tt 12

<210> SEQ ID NO 442
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 442
agtaaagggc ct 12

<210> SEQ ID NO 443
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 443
ggtaattnccg tg 12

<210> SEQ ID NO 444
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 444
guucaccucca augacuagggg guuuuagagc uagaaauagc aaguuaaaaau aagguuaguc 60
cguuuaaaau 69

<210> SEQ ID NO 445
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 445
gcaacucaug uccucccccau guuuuagagc uagaaauagc aaguuaaaaau aagguuaguc 60
<210> SEQ ID NO 446
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 446
gagucgcaag aagaaasagc uagaaauagc aaguaaaau aaggcusaguc
60
cguuuuuu
69

<210> SEQ ID NO 447
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 447
gggccgaga uaggguguc guuuaagagc uagaaauagc aaguaaaau aaggcusaguc
60
cguuuuuu
69

<210> SEQ ID NO 448
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 448
gugscgcagcg gggccgagau guuuaagagc uagaaauagc aaguaaaau aaggcusaguc
60
cguuuuuu
69

<210> SEQ ID NO 449
<211> LENGTH: 76
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 449
gucaccucsc aaugacuaggc guuuaagagc uagaaauagc aaguaaaau aaggcusaguc
60
cguuauucau uuuuu
76

<210> SEQ ID NO 450
<211> LENGTH: 76
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 450
gacauagcaug uccucccau guuuaagagc uagaaauagc aaguaaaau aaggcusaguc
60
cguuauucau uuuuu
76

<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

gaguccgagc agaagagagc uauuuagac cagauuuauu uagguuuuuu
60
cguaucauu uuuuu
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

gggccgaga uuggyguguc guuuagagc uagaaauagc aaguuuaauu aagccaguc
60
cguaucauu uuuuu
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

gggccgagc gggcgagau guuuagagc uagaaauagc aaguuuaauu aagccaguc
60
cguaucauu uuuuu
76

<210> SEQ ID NO 454
<211> LENGTH: 88
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 454

guacucccua augacuaggg guuuagagc uagaaauagc aaguuuaauu aagccaguc
60
cguauacauc uugaaaaggu uuuuuu
88

<210> SEQ ID NO 455
<211> LENGTH: 88
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 455

gacaucaug ucucoccuau guuuagagc uagaaauagc aaguuuaauu aagccaguc
60
cguauacasc uugaaaaaggu uuuuuu
88

<210> SEQ ID NO 456
<211> LENGTH: 86
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 456

gaguccgagc agaagaagaa guuuaagagc uagaaaaugc aaguuuaauu aagguacaguc

cgususcac uugaaagaagu guuuuuuu

SEQ ID NO 457
LENGTH: 88
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 457

gggcccgaga uaggguguc guuuaagagc uagaaauagc aaguuuaauu aagguacaguc

cgususcac uugaaagaagu guuuuuuu

SEQ ID NO 458
LENGTH: 88
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 458

gugsgcgagc gggccagagau guuuaagagc uagaaauagc aaguuuaauu aagguacaguc

cgususcac uugaaagaagu guuuuuuu

SEQ ID NO 459
LENGTH: 103
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 459

gucaccuca augaucagg gguuaagagc uagaaauagc aaguuuaauu aagguacaguc

cgususcac uugaaagaagu ggcacgagc cgugcuuuc uu

SEQ ID NO 460
LENGTH: 103
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 460

gacaucagug ucucuocacau guuuaagagc uagaaauagc aaguuuaauu aagguacaguc

cgususcac uugaaagaagu ggcacgagc cgugcuuuc uu

SEQ ID NO 461
LENGTH: 103
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 461
gaguccgagc aagaagaagc guuuuagagc uagaaauagc aaguuuaaaau aagguuaguc 60
cguuauacac uguuuaagu uggcaccgagc cggugcuuuu uuu 103

SEQ ID NO 462
LENGTH: 103
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 462

ggggccgaga ugggguguuc guuuuagagc uagaaauagc aaguuuaaaau aagguuaguc 60
cguuauacac uguuuaagu uggcaccgagc cggugcuuuu uuu 103

SEQ ID NO 463
LENGTH: 103
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 463

ggggccgagc gggccgagau guuuuagagc uagaaauagc aaguuuaaaau aagguuaguc 60
cguuauacac uguuuaagu uggcaccgagc cggugcuuuu uuu 103

SEQ ID NO 464
LENGTH: 120
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

SEQUENCE: 464

ttggsaagga cggaaacccg ggttttcag aagactgttt ttagagctag aaatagcaag 60
ttttaataa gcgtgacgct tatcaacttg aaaaagttgc accgatgcca tggggtttttt 120

SEQ ID NO 465
LENGTH: 60
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 465

tcggcgct gtgtgatcttc tctttgctct ttggtttttt 40

SEQ ID NO 466
LENGTH: 26
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

SEQUENCE: 466

gauuuuucac uugccuuuuu guuuua 26

SEQ ID NO 467
LENGTH: 26
<210> SEQ ID NO 468
<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: (22)...(26)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 468

tgatttcttc ttcgcttttt tnnnn

<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: (21)...(21)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 469

tgatttcttc ttcgcttttt tggct

<210> SEQ ID NO 470
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
<400> SEQUENCE: 470

gatttcttct ttcgcttttt tgg

<210> SEQ ID NO 471
<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: 471

tcc atc ctc aca acc cac cct gct agt gag c
Ser Ile Arg Thr Thr His Asn Pro Ala Ser Glu
1 5 10
<210> SEQ ID NO 472
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 472
Ser Ile Arg Thr Thr His Asn Pro Ala Ser Glu
1  5  10

tcc atc gca acc acc cac aac cct gct gat gag c
Ser Ile Ala Thr Thr His Asn Pro Ala Ser Glu
1  5  10

<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:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(33)

<400> SEQUENCE: 473
tcc atc gca acc acc cac aac cct gct gat gag c
Ser Ile Ala 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 Ala 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 gct aca acc cac aac cct gct gat gag c
Ser Ile Arg Thr Thr His Asn Pro Ala Ser Glu
1  5  10

<210> SEQ ID NO 476
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(36)

<400> SEQUENCE: 476
cac tgg asl tta aaa gas acc acc gtt tgg gtt tagga
Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Leu Val
1  5  10
<210> SEQ ID NO 477
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 477

Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Leu Val
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) .. (42)

<400> SEQUENCE: 478

caa ttg sat tta aaa gag acc gat acc gtt tac gas att gga
42
Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> SEQ ID NO 479
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 479

Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> SEQ ID NO 480
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (2) .. (34)

<400> SEQUENCE: 480

t cct aaa aaa ccc gag acc tcc gog cgt aaa gta
34
Pro Lys Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> SEQ ID NO 481
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 481

Pro Lys Lys Pro Asn Ser Ala Leu Arg Lys Val
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 aaa ccc gag tcc gog cgt aaa gta
34
Pro Thr Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10
<210> SEQ ID NO: 483
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 483

Pro Thr Lys Pro Arg Ser Ala Leu Arg Lys Val
1  5  10

<210> SEQ ID NO: 484
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 484
tgcgttgtt gattttctct tgcgccttttt tgg

<210> SEQ ID NO: 485
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 485
tacgcttgtt gatctctttct tgcgccttttt ttg

<210> SEQ ID NO: 486
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 486
gagaggcttttt atggggaag gcccattg

<210> SEQ ID NO: 487
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 487
gaaaaaga agactagaa tttgatac

<210> SEQ ID NO: 488
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 488
gagaggcttttt atggggaag gccaagaaga ctagaaatttt tgcatac

<210> SEQ ID NO: 489
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 489
agttgaagca taatgtctca aaaaata

<210> SEQ ID NO: 490
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 490
atittatat tacaatcag tgaatcat
<210> SEQ ID NO 491
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 491

aggtgaagca taatgtctca aaatattac aaatcagtga aatcat 46

<210> SEQ ID NO 492
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (36)

<400> SEQUENCE: 492

aat tta aaa gaa acc gat acc gtt tac gaa att gga 36
Aam Leu Lye Glu Thr Aep Thr Val Tyr Glu Ile Gly
1 5 10

<210> SEQ ID NO 493
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 493

Aam Leu Lye Glu Thr Aep Thr Val Tyr Glu Ile Gly
1 5 10

<210> SEQ ID NO 494
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (30)

<400> SEQUENCE: 494

aat tta aaa gaa acc gat acc gtt gtt taagga 36
Aam Leu Lye Glu Thr Aep Thr Val Leu Val
1 5 10

<210> SEQ ID NO 495
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 495

Aam Leu Lye Glu Thr Aep Thr Val Leu Val
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) .. (36)

<400> SEQUENCE: 496

ttg gat cca aaa aaa tat ggt gtt ttt gat gtt cca 36
Tip Aep Pro Lye Lye Tyr Gly Gly Phe Aep Ser Pro
1 5 10
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1  5  10

Trp Asp Pro Lys Lys Tyr Cys Gly Phe Asp Ser Pro
1  5  10

Trp Asp Pro Lys Lys Tyr Cys Gly Phe Asp Ser Pro
1  5  10
| Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu | 50  
| 60 |
| aag aga acc gcc aga aga aga tac acc aga cgg cgg cgg atc tgc | 240 |
| Lys Arg Thr Ala Arg Arg Tyr Thr Arg Arg Lys Arg Arg Ile Cys | 65  
| 70  
| 75  
| 80  |
| tat ctg cca gag atc ttc agc aac gag atg gcc aag gtc gac gac agc | 288 |
| Tyr Leu Glu Glu Ile Phe Ser Arg Glu Met Ala lys Val Asp Ser | 95  
| 90  
| 95  |
| ttc ttc cac aga ctt gaa gag tcc ttc ctg gta gaa gag gaz aag aag | 336 |
| Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Asp Lys Lys | 100  
| 105  
| 110  |
| cac gac cgg cac ccc ttc gcc aac atc gta gac gac ggt gcc tcc | 384 |
| His Glu Arg His Pro Ile Phe Gly Arg Ile Val Asp Val Ala Tyr | 110  
| 120  
| 125  |
| cac gac aag tac ccc gcc ttc gcc aac atc tac cac ctg aag aag aag ctg gta gac | 432 |
| His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp | 130  
| 135  
| 140  |
| agc acc gac aag gcc gcc ctg ggg atc tat ctg gcc ctg gcc cac | 480 |
| Ser Thr Asp Ala Asp Leu Arg Leu Ile Tyr Leu Ala Ala Ala His | 145  
| 150  
| 155  
| 160  |
| agt atc aag ttc cgg gcc cac ttc ctg atc gaa ggg gac ctg aac gcc | 528 |
| Met Ile Lys Phe Arg Leu Glu His Phe Leu Ile Glu Gly Asp Leu Asp Pro | 165  
| 170  
| 175  |
| gac aac agc gac gtt gac aag ttc atc atc cag ctg gta gac acc tac | 576 |
| Asp Asn Ser Asp Val Asp Leu Tyr Phe Ile Glu Leu Val Glu Thr Tyr | 180  
| 185  
| 190  |
| aac cag ctg ttc gag gaa gcc aac atc aac gcc agc gcc gtt gac gcc | 624 |
| Asp Glu Leu Glu Pro Glu Glu Arg Pro Ile Arg Ala Ser Gly Val Asp Ala | 195  
| 200  
| 205  |
| aag gcc atc ctg tct gcc aga aag aag aag agc cgg ctt gaa aat | 672 |
| Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Arg | 210  
| 215  
| 220  |
| ctg atc gcc cag ctg gcc gcc gag aag ac aag aag ggg ctg ttc gcc aac | 720 |
| Leu Ile Ala Glu Leu Pro Gly Glu Lys Arg Leu Arg Phe Gly Arg | 225  
| 230  
| 235  
| 240  |
| ctt att gcc ctg agc ctg gcc ctg acc gcc aac tcc aag agc aac tcc | 768 |
| Leu Ile Ala Leu Ser Leu Tyr Pro Gly Phe Ser Arg Ser | 245  
| 250  
| 255  |
| gac ttc gcc gag gat gcc aac ctt cag ctg agc aag gac acc tac gcc | 816 |
| Asp Leu Ala Glu Asp Ala Lys Glu Leu Ser Lys Asp Thr Tyr Asp | 260  
| 265  
| 270  |
| gac gac ctg gcc aac ctg ctg gcc cag atc ggc gac cag tac gcc gcc | 864 |
| Asp Asp Leu Leu Ala Glu llle Gly Arg Gin Thr Tyr Asp | 280  
| 285  
| 290  
| 295  
| 300  |
| ctt ctt ctg gcc gcc aag ac ctt tcc gcc gcc atc ctg ctg agc gac | 912 |
| Leu Phe Leu Ala Ala Lys Arg Arg Asp Ala Ile Leu Leu Ser Asp | 290  
| 295  
| 300  |
| atc ctg aga gtt aac cgg gac atc aac gcc ccc agc gcc ttc | 960 |
| Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser | 305  
| 310  
| 315  
| 320  |
| atc atc aag aag tac gag gac cac cac cag gaz ctg acc ctg gaz | 1008 |
| Met Ile Lys Arg Tyr Asp Glu His His Glu Asp Thr Leu Leu Lys | 325  
| 330  
| 335  
| 340  
| 345  
| 350  
| gac gac agc aag gcc tac gcc gcc aag gac gac gga gcc gcc | 1104 |
| Asp Glu Ser Lys Arg Gly Tyr Ala Glu Tyr Ile Asp Gly Gly Ala Ser | 355  
| 360  
| 365  

---
gcc aac aga aac ttc atg cag ctc atc cac gac gac aag ctt acc ttt  
Ala Asn Arg Asn Phe Met Glu Leu Ile His Asp Asp Ser Leu Thr Phe  
690  695  700

aaa gag gac atc cag aac ggc cag gtc ctc ggc cag ggc gag agc ctt  
Lys Glu Asp Ile Glu Ala Gln Val Ser Gly Glu Gly Asp Ser Leu  
705  710  715  720

cac gag cac att gcc aat ctc ggc aag aag aag ggc  
His Glu His Ile Ala Asn Leu Gly Ser Pro Ala Ile His Lys Gly  
725  730  735

atc ctc cag cca gta ctt cac gac gag ctc gta aag cag ggc  
Ile Leu Gln Thr Val Val Asp Glu Leu Val Lys Val Met Gly  
740  745  750

cgg cac aag ccc gag aac atc gta atc gcc gac aag aag cac gac  
Arg His Lys Pro Glu Asn Ile Val Ala Met Ala Asn Gln  
755  760  765

acc acc cac aag gga cag aac agc cgc gag aga atg aag cgg atc  
Thr Thr Glu Lys Gly Glu Lys Asn Arg Glu Arg Met Lys Arg Ile  
770  775  780

gaa gaa ggc ctc aac gaa gtc cgc agc cag atc ctc aag gaa cac ccc  
Glu Gly Ile Lys Glu Leu Gly Ser Gin Ile Leu Lys Lys His Pro  
795  799  800

gtg gaa aac acc cag cta cag aag aag ctc tac ctc tac tac ctc  
Val Gln Asn Thr Glu Leu Gln Asn Leu Arg Tyr Leu Tyr Tyr Leu  
805  810  815

cag aat ggg cgg gat atg tac gtc gac cag gaa cgg aac ctt gac  
Gln Asn Gly Arg Asp Met Tyr Val Asp Leu Ala Arg  
820  825  830

cgg cac tcc gac tac gat gtt gac gcc atc gtt cct cag acc ttc ctc aag  
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gin Ser Phe Leu Lye  
835  840  845

gac gac tcc gac gcc aag gtc cag ctc acc aca gcc gac aag ggc cgg  
Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg  
850  855  860

ggc aag agc gac aac gtc ccc gaa gag gtc gta aag aag atg aag  
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lye  
865  870  875  880

aag tac tgg cag ctc gtt gac aac gaa ctc att acc cag aca aag  
Asn Tyr Thr Arg Glu Leu Leu Asn Ala Lys Leu Ile Thr Arg Gin Lye  
885  890  895

ttc gac aat ctt acc aag ggc gag aag ggc cgg ctg aag gac gtt gac  
Phe Asp Asn Asn Thr Lys Ala Glu Arg Gly Leu Ser Glu Leu Asp  
900  905  910

aag ggc ggc ttc atc aag aga cag ctc gta gaa acc cgg cag atc aca  
Lys Ala Gly Phe Ile Lys Arg Glu Val Glu Thr Arg Gin Ile Thr  
915  920  925

aag cac gta gca cag ctc gcc tct gcc atg acc acc acc tac gac  
Lys His Val Ala Glu Ile Leu Asp Ser Arg Met Asn Thr Tyr Asp  
930  935  940

gag aat gag cag ctc cgg gaa gtc gaa aag gtc atc acc ctc aag ggc  
Glu Asn Asp Lys Leu Arg Glu Val Lys Val Ile Thr Leu Lys Ser  
945  950  955  960

aag ctc gtt gat ttc cgg aag gat ttc cag ttt ttc aag ggt gtc  
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gin Phe Tyr Lys Val Arg  
965  970  975

gag atc aac tac cac cac gcc cac gcc gcc tac ctc aag gcc gtc  
Glu Ile Asn Asn Tyr His His Ala His Ala Tyr Leu Asn Val Ala  
980  985  990

gtg gga acc gcc ctc aac aag tac cct aag ctt gaa aag gag ttc  
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe  
3024
<table>
<thead>
<tr>
<th></th>
<th>995</th>
<th>1000</th>
<th>1005</th>
</tr>
</thead>
<tbody>
<tr>
<td>gtg tac</td>
<td>ggc gag tac aag gtc</td>
<td>tac gag tgc cgg aag</td>
<td>atg atc gcc</td>
</tr>
<tr>
<td>Val</td>
<td>Tyr</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td>1010</td>
<td>1015</td>
<td>1020</td>
<td></td>
</tr>
<tr>
<td>aag acc</td>
<td>gag cag gaa atc cgc</td>
<td>aag gct acc ggc aag</td>
<td>tac ttc ttc</td>
</tr>
<tr>
<td>Lys</td>
<td>Ser</td>
<td>Glu</td>
<td>Gin</td>
</tr>
<tr>
<td>1025</td>
<td>1030</td>
<td>1035</td>
<td></td>
</tr>
<tr>
<td>tac acc</td>
<td>aac atc atg acc ttt</td>
<td>ttc aag acc gag att</td>
<td>acc ctg gcc</td>
</tr>
<tr>
<td>Tyr</td>
<td>Ser</td>
<td>Arg</td>
<td>Arg</td>
</tr>
<tr>
<td>1040</td>
<td>1045</td>
<td>1050</td>
<td></td>
</tr>
<tr>
<td>aac ggc</td>
<td>gag atc cgg aag cgg</td>
<td>cct ctc atc gag aca</td>
<td>aac ggc gaa</td>
</tr>
<tr>
<td>Amm</td>
<td>Gly</td>
<td>Glu</td>
<td>Ile</td>
</tr>
<tr>
<td>1055</td>
<td>1060</td>
<td>1075</td>
<td></td>
</tr>
<tr>
<td>acc ggg</td>
<td>gag atc gtt gct gat</td>
<td>aag ggc cgc gat ttc</td>
<td>gcc acc ctg</td>
</tr>
<tr>
<td>Thr</td>
<td>Gly</td>
<td>Ile</td>
<td>Val</td>
</tr>
<tr>
<td>1070</td>
<td>1075</td>
<td>1080</td>
<td></td>
</tr>
<tr>
<td>cgg aaa</td>
<td>gtt ctg agc atg ccc</td>
<td>cca gtt aat atc ttt</td>
<td>aaa aag acc</td>
</tr>
<tr>
<td>Arg</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>1085</td>
<td>1090</td>
<td>1095</td>
<td></td>
</tr>
<tr>
<td>gag gtt</td>
<td>cag cca ggc ggc ttc</td>
<td>aac aaa gag tcc aac</td>
<td>ctg ccc aag</td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Gin</td>
<td>Thr</td>
</tr>
<tr>
<td>1100</td>
<td>1105</td>
<td>1110</td>
<td></td>
</tr>
<tr>
<td>aag acc</td>
<td>gag atg aag ctg atc</td>
<td>gcc aga aag aag gag</td>
<td>tgg gac cct</td>
</tr>
<tr>
<td>Arg</td>
<td>Asp</td>
<td>Ser</td>
<td>Arg</td>
</tr>
<tr>
<td>1115</td>
<td>1120</td>
<td>1125</td>
<td></td>
</tr>
<tr>
<td>aag aag</td>
<td>tac ggc ggc ttc gac</td>
<td>agc ccc acc gtg gcc</td>
<td>tat tct gtt</td>
</tr>
<tr>
<td>Lys</td>
<td>Lys</td>
<td>Tyr</td>
<td>Gly</td>
</tr>
<tr>
<td>1130</td>
<td>1135</td>
<td>1140</td>
<td></td>
</tr>
<tr>
<td>ctt gtt</td>
<td>ggt ggc aaaa tgt gaa</td>
<td>aag ggc aag tcc aag</td>
<td>aaa ctg aag</td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Ala</td>
<td>Val</td>
</tr>
<tr>
<td>1145</td>
<td>1150</td>
<td>1155</td>
<td></td>
</tr>
<tr>
<td>aat gtt</td>
<td>aaa gag ctc tgt ggg</td>
<td>atc acc atc atg gaa</td>
<td>aga aga acc</td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td>1160</td>
<td>1165</td>
<td>1170</td>
<td></td>
</tr>
<tr>
<td>ttc gag</td>
<td>aag aat ccc atc gac</td>
<td>ttt ctt gaa gcc aag</td>
<td>ggc tac aaa</td>
</tr>
<tr>
<td>Phe</td>
<td>Glu</td>
<td>Lys</td>
<td>Pro</td>
</tr>
<tr>
<td>1175</td>
<td>1180</td>
<td>1185</td>
<td></td>
</tr>
<tr>
<td>gaa gtt</td>
<td>aag gag gac tgt gtt</td>
<td>atc aag ctc gat aag</td>
<td>tac tcc ctt</td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Lys</td>
<td>Arg</td>
</tr>
<tr>
<td>1190</td>
<td>1195</td>
<td>1200</td>
<td></td>
</tr>
<tr>
<td>ttc gag</td>
<td>ctg gaa aac gcc cgg</td>
<td>aag aga atg ctc gcc</td>
<td>tct gcc ggc</td>
</tr>
<tr>
<td>Phe</td>
<td>Glu</td>
<td>Leu</td>
<td>Lys</td>
</tr>
<tr>
<td>1205</td>
<td>1210</td>
<td>1215</td>
<td></td>
</tr>
<tr>
<td>gaa ctt</td>
<td>cag aag gga aac gaa</td>
<td>ctg gcc ctt cgg ttc</td>
<td>aaa tat gtt</td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Gin</td>
<td>Lys</td>
</tr>
<tr>
<td>1220</td>
<td>1225</td>
<td>1230</td>
<td></td>
</tr>
<tr>
<td>aac tcc</td>
<td>ctt tac ctc gcc aag</td>
<td>cac tat gag aag ctt</td>
<td>aag gcc tcc</td>
</tr>
<tr>
<td>Arg</td>
<td>Phe</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td>1235</td>
<td>1240</td>
<td>1245</td>
<td></td>
</tr>
<tr>
<td>ccc gag</td>
<td>gat aat gag cag aag</td>
<td>cag ctt tgt gaa gaa</td>
<td>ctg cac aag</td>
</tr>
<tr>
<td>Pro</td>
<td>Glu</td>
<td>Asp</td>
<td>Arg</td>
</tr>
<tr>
<td>1250</td>
<td>1255</td>
<td>1260</td>
<td></td>
</tr>
<tr>
<td>cac tac</td>
<td>ctt gag gac cac atc</td>
<td>gac cag atc gcc gag</td>
<td>ttc tcc aag</td>
</tr>
<tr>
<td>His</td>
<td>Tyr</td>
<td>Arg</td>
<td>Asp</td>
</tr>
<tr>
<td>1265</td>
<td>1270</td>
<td>1275</td>
<td></td>
</tr>
<tr>
<td>aga gtt</td>
<td>atc ctg gcc gac gct</td>
<td>aat ctt gac aag gtt</td>
<td>ctg tcc gcc</td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
<td>Ile</td>
<td>Leu</td>
</tr>
<tr>
<td>1280</td>
<td>1285</td>
<td>1290</td>
<td></td>
</tr>
<tr>
<td>tac acc</td>
<td>aag cac cgg gat aag</td>
<td>ccc atc gag cag</td>
<td>ggc gag aat</td>
</tr>
<tr>
<td>3924</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino Acids</td>
<td>Sequence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>acg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino Acids</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>Asp</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Ala</td>
<td>ggc</td>
</tr>
<tr>
<td>Glu</td>
<td>ggc</td>
</tr>
<tr>
<td>Tyr</td>
<td>ggc</td>
</tr>
<tr>
<td>Lys</td>
<td>ggc</td>
</tr>
<tr>
<td>Thr</td>
<td>ggc</td>
</tr>
<tr>
<td>Ser</td>
<td>ggc</td>
</tr>
<tr>
<td>Ile</td>
<td>ggc</td>
</tr>
<tr>
<td>Arg</td>
<td>ggc</td>
</tr>
<tr>
<td>275</td>
<td>280</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
</tr>
<tr>
<td>Thr</td>
<td>290</td>
</tr>
<tr>
<td>295</td>
<td>Asp</td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
</tr>
<tr>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td>Thr</td>
<td>Ala</td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
</tr>
<tr>
<td>315</td>
<td>320</td>
</tr>
<tr>
<td>Met</td>
<td>Ile</td>
</tr>
<tr>
<td>Arg</td>
<td>Tyr</td>
</tr>
<tr>
<td>Glu</td>
<td>His</td>
</tr>
<tr>
<td>Gln</td>
<td>Asp</td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>Arg</td>
<td>Gln</td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td>Tyr</td>
<td>Lys</td>
</tr>
<tr>
<td>Ile</td>
<td>Phe</td>
</tr>
<tr>
<td>Thr</td>
<td>Tyr</td>
</tr>
<tr>
<td>Asp</td>
<td>Gly</td>
</tr>
<tr>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td>Ser</td>
<td>Lys</td>
</tr>
<tr>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>Thr</td>
<td>Asp</td>
</tr>
<tr>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td>Gln</td>
<td>340</td>
</tr>
<tr>
<td>Glu</td>
<td>350</td>
</tr>
<tr>
<td>Phe</td>
<td>355</td>
</tr>
<tr>
<td>Tyr</td>
<td>365</td>
</tr>
<tr>
<td>Pro</td>
<td>370</td>
</tr>
<tr>
<td>Lys</td>
<td>380</td>
</tr>
<tr>
<td>Met</td>
<td>385</td>
</tr>
<tr>
<td>Arg</td>
<td>395</td>
</tr>
<tr>
<td>Leu</td>
<td>405</td>
</tr>
<tr>
<td>Val</td>
<td>415</td>
</tr>
<tr>
<td>Lys</td>
<td>420</td>
</tr>
<tr>
<td>Asp</td>
<td>430</td>
</tr>
<tr>
<td>Arg</td>
<td>435</td>
</tr>
<tr>
<td>Leu</td>
<td>445</td>
</tr>
<tr>
<td>Lys</td>
<td>450</td>
</tr>
<tr>
<td>Arg</td>
<td>460</td>
</tr>
<tr>
<td>Asp</td>
<td>465</td>
</tr>
<tr>
<td>Leu</td>
<td>475</td>
</tr>
<tr>
<td>Val</td>
<td>485</td>
</tr>
<tr>
<td>Arg</td>
<td>495</td>
</tr>
<tr>
<td>Leu</td>
<td>500</td>
</tr>
<tr>
<td>Pro</td>
<td>510</td>
</tr>
<tr>
<td>Asp</td>
<td>515</td>
</tr>
<tr>
<td>Leu</td>
<td>525</td>
</tr>
<tr>
<td>Tyr</td>
<td>530</td>
</tr>
<tr>
<td>Thr</td>
<td>540</td>
</tr>
<tr>
<td>Asp</td>
<td>545</td>
</tr>
<tr>
<td>Leu</td>
<td>555</td>
</tr>
<tr>
<td>Ala</td>
<td>565</td>
</tr>
<tr>
<td>Val</td>
<td>575</td>
</tr>
<tr>
<td>Gly</td>
<td>580</td>
</tr>
<tr>
<td>Ser</td>
<td>590</td>
</tr>
<tr>
<td>Val</td>
<td>595</td>
</tr>
<tr>
<td>Gly</td>
<td>605</td>
</tr>
<tr>
<td>Thr</td>
<td>610</td>
</tr>
<tr>
<td>Tyr</td>
<td>620</td>
</tr>
<tr>
<td>Asp</td>
<td>625</td>
</tr>
<tr>
<td>Leu</td>
<td>635</td>
</tr>
<tr>
<td>Phe</td>
<td>645</td>
</tr>
<tr>
<td>Thr</td>
<td>655</td>
</tr>
<tr>
<td>Val</td>
<td>660</td>
</tr>
<tr>
<td>Thr</td>
<td>670</td>
</tr>
<tr>
<td>Lys</td>
<td>675</td>
</tr>
<tr>
<td>Ser</td>
<td>695</td>
</tr>
<tr>
<td>Ala</td>
<td>690</td>
</tr>
<tr>
<td>Leu</td>
<td>700</td>
</tr>
<tr>
<td>Thr</td>
<td>705</td>
</tr>
</tbody>
</table>

---

US 8,871,445 B2

-continued
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 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 Gln  
755 760 765
Thr Thr Glu Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile  
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln 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 Gln Glu Leu Asp Ile Asn Arg  
820 825 830
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys  
835 840 845
Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg  
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Gln Val Val Lys Lys Met Lys  
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys  
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gln Leu Ser Glu Leu Asp  
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr  
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp  
930 935 940
Glu Asn 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 Asp Gln Phe Tyr Lys Val Arg  
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Ala Ala Tyr Leu Asn Ala Val  
980 985 990
Val Gly Thr Ala Leu Ile Lys 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 Gin 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
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn 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 Asn Ile Val Lys Lys Thr  
1095 1099 1095
Glu Val Gin Thr Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys  
1100 1105 1110
ctcttgccac tctctt
<210> SEQ ID NO 506
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 506
acatcagtgt ccacctccaat gacaagcttt gcagcgggtg gcaccccaac ac 52
<210> SEQ ID NO 507
<211> LENGTH: 1733
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<400> SEQUENCE: 507
cogtttaaag atatctcgag gatactgatt attaataatga ataattaatag cggtaatgat 60
ttcatagcc atatatgag ttctcggtta ctaacttac ggttaatggc cggctggtgc 120
gacacccgcc cgacccccgc ccattgcgct caataatacg gtatgtccc atatastaccgc 180
caatagggac ttctcactgga cggctaatggc cggctgacttt cggctgac 240
cagtacacta agtctactc atgaactgta ggtcccttat tgcagctcat gacctgaat 300
gcgcgccttg gcatctacg cagtcagcag ccttcactag ggcgtcaa 360
ttcacatgatt attcacatgtt tcggagtttag cccactggtcg tcggctatc 420
tcccatctct ccacccctcc gcaaccccaag tttttctttt attttttttt taattttttttt 480
gtgcggggtg gggggcggtgg gggggggggg ggctggcgcc aggccgggagc gggggggggg 540
agggggcgggg cgccgggcaag gggagaggtg cggctggacc caatccagag ggcgcggtc 600
ggasgttccc tttattgagc agggggggcc ggcggccgc ccataaaagc cgaagccgct 660
gccgggggg gatcgggtcg ccgcctgttcc cgcggctgcc cgcggctggtc 720
gcgcgcgcgc gcggccggga cggggggaca atgcggcgcgc cagttcgccc gggggcgtc 780
ccttcctcct ggggctctat gcgtcttttg ttttaataac gcgttttctt ttgtttgttg 840
tgcgtgaaag ccggggccc gtcgggagg cggcttctgt tggggggagc cggctggggg 900
gtgcggtgtg gcggcggggg gtcggggcc ggctggcgcc cggccgggctc gcggggggc 960
gtgccgcgag ccgggggcctct ttctgctgctc cagttcgccc caggggagctc 1020
cggccccggc ggcggccggc ggctggggtc ggctggccgg ggggccacacc gcgtggggtc 1080
ggggggtggtg cgggggggggg ggctgggcgg gggggagtgg cggccagggcc gcgggagcgg 1140
cctcgacacc cccctccgaa cggggtggtgc ccggcgggagt ggcgtggggc ggcgttggct 1200
gcgggggggc cgcggccggc cggcggggcg gcgtggccgg cggccggccgg gcgggcgggcgg 1260
gcgggggagt ggcgtgggac cggccggcgc gcggccaggg tcgggggggg ggcggtgggg 1320
gcgccggccg ccgggtgggg gcgggatgtc gcgggggggg gcgggggggg gcgggggggg 1380
aggggggggca gcggctctct ttcggggggc gagcaatcgc cggggggggc ggccgggggg 1440
ggcggcggcgg cctctaggggc ggcggggggc gcgggggggg cggcgggggg cgccgggggg 1500
ccgggggagg gttcgggggg cggcgggtgg cggggtgggg gcgtggccgg gccggtgggg 1560
gtgctgggg gcgggaggtg ggccggccgg gccggggtgg gcgggggggg gcgggggggg 1620
ggcgggtggcc ggcggggtgg gcgggggggg gcgggggggg gcgggggggg gcgggggggg 1680
<210> SEQ ID NO: 508
<211> LENGTH: 4269
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 508

atggcttata aggccgacgca aggagtcagt atatgattta caaagcgat 60

gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 120
gacgccagc tcaacagcact ccgccacagc ctggaggtgg ctggaggtgcc 180
 accgcaagc aagagaaggc accgagcttg gctttcgacc gcagagctga acgcagcgg 240
 accgccgaac acgcaggaac cagaaggagc tccaccagac gc gagccgaccgagc 300
 atggcttct cggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 360
 tggagccagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 420
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 480
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 540
 atggcttct cggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 600
 itgcagaggt ccagaggtcc gctggagagt gagccgaccgagc 660
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 720
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 780
 cgaacgcagc cccgagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 840
 attcgatgac gctggagagt cccgagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 900
 cgaacgcagc cccgagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 960
 atggcttct cggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1020
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 1080
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1140
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1200
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1260
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1320
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1380
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1440
 cagcagcagc tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1500
 tccgtcggga tggagcaggt ccagaggtcc gctggagagt gagccgaccgagc 1560
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 1620
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 1680
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 1740
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 1800
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 1860
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 1920
 gacgataaga tgtgccccaa gaagaagccg aaggtcggta tccaccgagt cccacgcccc 1980
---continued

acctgaccc tgtttgagg a cagagagatg atcagagaaac ggcggaaaac ctatgcccac 2040
ctgtcgacg acacaggtct aacggcggca gatacagctgc ctcgggcaagg 2100
tctgacgagc gacgtgcca cggccatcgg gacagacagt ggcgaagag ctatcgggat 2160
ctgcagtatg ctgctgtgctc ggcgaacgctc gacatgcgtc acgatgcagc 2220
tctgacctta aagagagct ccagaagacg caggtgctcg ggcaggggaa agatctgtac 2280
gaggacaggt cccatctggc ggcagccccg gccaattccg ggccatctcg ccagagctgt 2340
aaggtggtgg acgagctgcg gaaagtttagtg ggccgacgca aaggccgaaag aatccctgtc 2400
aaatgacgca gacagagacg gacaccacccg aagacagctg cagagaaatg 2460
aagccgacgc aagagagcgt cagaagagct ggccgacgc aacctgagaac 2520
gaaacccccc ctctgacgcc gccgaccggtg aactgtgaatg ggaccacact 2580
agtatacctgg accagactac gcggttggctcg aactgtgctg agataccagtc 2640
gtgcgccaga ggttcctggca ggaacgacatat cgcgcagtcc acgatggctc acgagaagcg 2700
aagacgcggc gcaagagcgag cctactgctg ctcgaagagct gctgacaagac gacagaaag 2760
tttcgtgctgg acggcagcag ccgcaagcaggt atttcgtgctg ctcgaagagct 2820
aagacgcggc gacggcgccttg gacggcgcatt ggtcctagtg gttctccgca 2880
agtatacgcg agaagacgact tgcgctggcag taacgtaaacg gtcgatctgg 2940
aagacgcggc gcggctgctt gcggcttgcttg gcggcttgcttg gcggctgctt gcggcttgcttg 3000
cgcgccgcgc gcgcgctgctt gcgcgctgctt gcgcgctgctt gcgcgctgctt gcgcgctgctt 3060
ccggccgcgc gcgcgctgctt gcgcgctgctt gcgcgctgctt gcgcgctgctt gcgcgctgctt 3120
cgcgccgcgc gcgcgctgctt gcgcgctgctt gcgcgctgctt gcgcgctgctt gcgcgctgctt 3180
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3240
acatcaggact ttcagcgaacct gtcggcacagc gggagatcgcg aagacgcggc 3300
tctgctgcttg ctcggcgggc aacggggaacg atctgtggag gtctgctgcttg 3360
acggccgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3420
acggccgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3480
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3540
ctgtaagggct ggagggctgc cccaggggat aagagagcag cagagcaaccg ttggagggct 3600
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3660
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3720
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3780
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3840
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3900
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 3960
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 4020
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 4080
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 4140
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 4200
acgacgcgctg gcggtgccga aacgggcaagc gcaagtgcgca aagcttcttt ctagacgcaac 4260
aagaaaaag 4269
<210> SEQ ID NO: 509
<211> LENGTH: 780
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 509

ggaaacgccg ccactaacct cttccctggtg aacaagcag gggatgtaga agagaatctcc 60
gggccagta gcagaagcga ggaacgttc ccgccccggcct tggccatctt gggagacgctg 120
gacggcgcag taacgccccaa gttccgagc gttcgcgccc aggcccaggg cgatggccac 180
tacggaactc ttgccctagt gttcagcgc ccaccaggggt gagtgccgccgag tgggac 240
acccctgtca ccacccgagc ttagggctgt cagttgtta ccggccgagc ccacccagag 300
aacagccgac ccttctttca aacgccccggg cgggacgagc gacgccccg 360
ttcctcagcg aacgccgac ctacacgaccc cgcgcgcagc ttaaagttaa ccgggagcagc 420
cgggtgagc aaagggcg acagggcc cgggagcagc gacgccccggg cgggacgagc gacgccccg 480
caccagcg ccaccgc ccccctta caggagc cgggagcagc gacgccccggg cgggacgagc gacgccccg 540
gcagccac ccaccgc ccccctta caggagc cgggagcagc gacgccccggg cgggacgagc gacgccccg 600
cagctctata ccagaggctt ttggttggtgcc ggttcgagccgc cccgccccggg cgggacgagc gacgccccg 660
gctctctgttag agtttctgcc ccagccccggg cgggagcagc gacgccccggg cgggacgagc gacgccccg 720
gtgccagcctg gctccttta ccagccccggg cgggagcagc gacgccccggg cgggacgagc gacgccccg 780

<210> SEQ ID NO: 510
<211> LENGTH: 597
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 510

cgataaatca cctctggatt aacaatattg tgaagaatag acattggatt ttaactatgt 60
tgctctctttt aagctatattg gtagcgcg ttaaagtccg ttgtctacgt ctacattctt 120
cggtctatcgt tttccgattt tttctctgt aataactttg tttggattt cttgctttttt 180
gtggggtgcc ggtgctagcc aacgcttgct ggtctttggt acagcgccccg 240
cacgctttggc gcaggccgcc cccgcttgcc cgggacttctt cttcccccttc 300
cctccctcgc ggcgcgcac ccctctttcc cgacgccggc cccgcttgcc 360
gccgtctgagg acctgccctc ctggccgttc cgacgccgc cccgcttgcc 420
gtcctctgagg acctgccctc ctggccgttc cgacgccgc cccgcttgcc 480
cctccgctt ccagcctgcc ccctctttcc cgggtctgagg acctgccctc ctggccgttc cgacgccgc cccgcttgcc 540
tcttgtctgagg acctgccctc ccctctttcc cgggtctgagg acctgccctc ctggccgttc cgacgccgc cccgcttgcc 597

<210> SEQ ID NO: 511
<211> LENGTH: 210
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 511

cgacgctcgct gttgtccttt cgtgtccctc ctggccttct cctcgctgct gcctgcctcc 60
ccttgacctt ggaaggtgcc acctcactgt tcctttctca ataaatgag gaaatgcat 120
cgatattgct ggtgagttgt cattcattcc tgggggttg ggtgggccag gacagcagag 180
ggagaggattt ggaagaccat gcggagcatg 210

<210> SEQ ID NO: 512
<211> LENGTH: 906
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<222> LOCATION: (109)..(109)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<222> LOCATION: (135)..(135)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 512
ataacctgtg ataattcttg ctatacgaag ttattggaga tgaataaag taagattgca 60
gatctgcgac ctctcggatt ctggagctct ggagatctt gctatgaccnt accaacatg 120
gttagagttt ctcttggtttt aaaaaaccct cccacactcc ccttgacactt gaaacattaa 180
atgaatcgaa ctgttctgtt atttctgtgtt ataatcaattt gtaatgttta aataaaagc 240
atctatcata ccattcttcc aatataagca ttttttctcc tggattctgg tttggtttg 300
tccactaatc ttaattcgtc tgcgatctgg ctcgctaggg atctagcatc aatactacta 360
gcataccac atcttctaggg gttttactgg ttttaaaata cttctaccaac cttctcccatc 420
acgtcaaaca taaaatgtat gcaatgtttg ttggtaactt tttattgcga gctttatattg 480
gttacaaatta aagcatatgc atcaccasatt tccaaaatatt aacagatttttt tcaagctatt 540
cagcttggtt tttgctacaa cttcataactc t tacattctac ttcgaacagct 600
gaggtctaatacagcata ccacatgtt a gactcctggta ctgcgacctc 660
acacatcctt cagacctgga acataaaat gataacaa tttggtgttta actgtgattt 720
tgcgacattt aatgatcata aataaagcag taaggctcaac aataacattttta aataagcatc 780
nttttcacg atctcagttg atttgttctgc caaactccct cactatcatt actgtaacctg 840
gctccacatt cagctgacgg ggaacccatt atataaccct gcattagttca tgaactacga 900
agttat 906

<210> SEQ ID NO: 513
<211> LENGTH: 1079
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<400> SEQUENCE: 513
cgggcccctcg gacgctggtgg gacgcgtttct gtggagacagc cgggtacagc tggtgactct 60
ggagggcggc agggctgggt cgggtcagcct ctgcagacac gcggaggggca 120
ggagagggc aggggcagga tctcaccggc acggagcagt ggtgccccgg ggggggggca 180
gcgagggagc gctcgcggac gccttgactg gctgtatgctg ctcttttttcc cttgagccttg 240
tgttaaaaca cacaaagcct gttttgtggc gcgtaagcgc cttgtcagtt aacggcagcc 300
413  -continued

ggagtgccga gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 360
gaggtcgacag ctggactcgt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga 420
gggtaacgca aagtagtctgg cgcgcaagct gcggcccaac ctcccccttct ccctggggag 480
tgttttttacct gcggccgcca gcctgcgttc gcctgcgttc gcctccaggg gggtggcggga 540
tgctgctctgc catgggtgttg tgtggtgcca agttggtctcc atccgcggca caccggggcc 600
gggtgacggt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 660
gggtgacggt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 720
tgagagcgggt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 780
gggcagccag acctccacg gcggccgcca gcctgcgttc gcctccaggg gggtggcggga 840
gggtgacggt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 900
gggcagccag acctccacg gcggccgcca gcctgcgttc gcctccaggg gggtggcggga 960
gagtgtgttc ataccccttt gcgggtgcttc tgtggtgcca gcacgctggt gcggccgcca 1020
gggcagccag acctccacg gcggccgcca gcctgcgttc gcctccaggg gggtggcggga 1079

<210> SEQ ID NO 514
<211> LENGTH: 4356
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 514
agatgggggg gagttctctgg gcggcgttca aagggctaac ttgtgtgtgg gcgtgtgtct 60
gcgaggggaa tcgaaagcgg taaattggga gggcaagacc tccccacaga ttctggattt 120
tgctggaggg tttttcattaa gggcctaaa agaaatatgg gcggcatggt gcgtcgctctgg 180
ggttttttcgc gcggagcattgc cggtttttgtgc cgggctgggt ggtttttttc 240
tcgaggtgct attaagacgc tggctccaccc apgttttttac tccctgtttt ggcagtttca 300
cgggctgggt gcggcgttca aagggctaac ttgtgtgtgg gcgtgtgtct 360
taacagccgg agaatcttcgt atccgctcgtc ctggctgctt gcgtcagcct 420
ttttttttcgc gcggagcattgc cggtttttgtgc cgggctgggt ggtttttttc 480
gacagagatt cggccttttt cttttctcca tccctgtttt ggcagtttca 540
acagagatt cggccttttt cttttctcca tccctgtttt ggcagtttca 600
ttccttactt cttttctcca tccctgtttt ggcagtttca 660
ttccttactt cttttctcca tccctgtttt ggcagtttca 720
ttccttactt cttttctcca tccctgtttt ggcagtttca 780
gggcctgctc tttttccttt cttttccttt cttttccttt cttttccttt 840
ggcgctgcgt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 900
ggcgctgcgt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 960
ggcgctgcgt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 1020
ggcgctgcgt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 1080
ggcgctgcgt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 1140
ggcgctgcgt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 1200
ggcgctgcgt gcggccgcca gcctgcgttc gcctccaggg gggtggcggga gcgtaggggg 1260
agaagatatt caggactatt ttggaactgg gtaaggatatt gattctgtatt agcaacctgta 3720
gtgagcattg agtcgcccct ggctgtagat tggagttacat tgtaatatct aataataagt 3780
ggctttttaa aagagcctt gttttcttct aactcattt tttgctttaa 3840
gaatacttg ttgctgtacc ctctgacagc aaataaaggct tttaaaagct atatatctat 3900
tcctttttttttttt gtttctttaa gcaaggctttg ctgtgctttt cctgctgtct 3960
ccttgaaactc actttcttga ccagctcttc ctcgaaactca gaaatccgcc tgcctctgcct 4020
tcctgagctgc eggatataaa ggcgcctgcgcc accagccttg gtaagtttgg atatatgtgt 4080
ataataactt aacaactaat aacctcaagt ggtagttttt taatacctgc agtagcttta 4140
agtagctttt attggtcctctt cactaaactt cactctctgc cttacaaagc gctgttggctt 4200
cctggtcgc ctaagcactt tcctacgctt atactagcag attaggggtc agggcttagaa 4260
acttagtcat cgttttggtg aclattcagc atctactac cttgttcgta tagaatattt 4320
toaggacact tagett 4336

<210> SEQ ID NO: 515
<211> LENGTH: 1646
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: Description of Artificial Sequence: Synthetic polymer nucleotide
<400> SEQUENCE: 515

aatctccag ggtagggagg gctctttttc cagggcgcgt tggagcgtgc gctttggcag 60
cooogggctg cacttggcgcc tacaactagt gctctgggccc tggcagatc accatcataa 120
ccgtaggag ccacggccgct ccgttctttt gttgccccttg cgccgcacct tctactcctc 180
cctagtgcg gaaaggctgg cccggccggct agtcggcctg gctgcagagc tgcaaaatgg 240
aaggtagcgc ttcacaagtgt ctctgagctgc tggagcgcgc ccagctgcgaat tggagcgggg 300
tagctcttt gggccaggggc caatactcagc tttgcttctc gctcttgggg gtcctagggg 360
tggaggggg ggtgtccggc ggccgctcca cggggggtcg ccggggggcg ggggggggcgg 420
gaagtcctgt cgagggccct catttgctga ccgtcttttc ggcgtgtggtg gactggt tgt 480
tctctctct ctcactctct gcctttcgg cttgctcttg ccgctcaaat gttgaacagc atggattgca 540
tctgagatg tataaactat tgcacccat ccggagccgc attggaaacc atggattgca 600
cgcaggttct cggccgcttt ggctggagatg gacttggcgc tattgaccca cacaaccagc 660
aatgcttgcg tctgcttcgg ccggctccgg gccttgcatt gccgctttgtt gacttcggcct 720
tgtcaagcgc gcctcggcttg gtggctctga tggagctccg cggagggcag ccgggtcgtc 780
gtgctggcc agcagggccc tctctggcgg acgtcctgcg gcagctgctga ctggagcgggg 840
aagggaatcg gtctctatgg cggcaagctgc ggccggagatc tctctgtgct tgtctgttgcc 900
tcttgcggag aagatcttcca tctggcgtga tgaagctggc cggctctgata cggctcgctt 960
ggccttttt ggctggctgc accataagaa atctagcagc ggcgagccac gcatctggat 1020
gggagcgtc cttgctgcag ggagtatctg ggcaagcgac tctggggcgg ggtgggccgc 1080
ggatcggccc gaggctgcct cggccggcgc gctctctttcc ctctacatcg 1140
tggccagtgc cctgttgcag atacatatgt ggaaataagg gctcttttctg gattctgctg 1200
tctggtggcc ctgggtgggtt cggacgcgct gagctcgata cttggcagctc cggctgatat 1260
tggcagggc ctttgccgag attggggcgtc cggctcttct gttgttatag gtatggccg 1320
tccogatcgc cagcgcatcg ccttctatcg ccttttgac gagtcttttt gaggggatcc 1380
gctgtaagtc tgcgaasatg gatgatctat taacaactaaa agatgtccac taaaatggas 1440
gtttttcttg taataacttta ttaaaagcgg taacacacta attttgagtcg 1500
agagttggag ctacgaggggt ggagggtggg gttggattgaa taataatcttg ctcttcctcttg 1560
aagctttttc attaatttcc tatttattag ttttttattg gttatcatag attttaaacc 1620
goaaaacca atttaagggc ccgcttattt ctctctctct ctctctata gctctctgatc 1680
tctctgggat tccatttttt tctttttttt ccacactttg gtttctattg acgtgtctttg 1740
ccaaatgtgtt cgttttcaaa gcgtaagaga cgaagttcagc gcgtctggtt ccacataac 1800
tctcttcct gttatgttttt ggcaatgtag aatttcatcg gaaagccgtaagtgtagtctatatggatccg
<210> SEQ ID NO: 516
<211> LENGTH: 1519
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic nucleotide
<400> SEQUENCE: 516

taccgggttag ggagggcctgt tttcccaaggg cagttgctacg cattcgtctttt agcaagccccc 60
cggggaattt tggcctaca tgcatttcaca ccgctctacg ttcctacttt gggggtccgg 120
agggcgcacg gggcctattt cttcttctttctc tctctctctct ttcctctctct ttcctctctct 180
gcctaggggt gcccctccgg gcccctccgg gcccctccgg gcccctccgg gcccctccgg 240
agcgcaacct cgtttttttttt ttccttcttttt ttcctttttttttt ttcctttttttttt ttcctttttttttt 300
tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
gaagacaata gcaggcatg

<210> SEQ ID NO 517
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 517

gaggccat ttocatgat tcc

<210> SEQ ID NO 518
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 518

cctgtggaaa gagcgaacaa cc

<210> SEQ ID NO 519
<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: (4)...(23)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 519

aacnnnnnn nnnnnnnnn nnnnggttgtt cgtcttttcc acaag

<210> SEQ ID NO 520
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 520

cctgagtttt gagcccccag tggctgct

<210> SEQ ID NO 521
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 521

acggagccag agtgtggttt gctgtgggcc aggcccc

<210> SEQ ID NO 522
<211> LENGTH: 68
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 522

catcagcttc tcagctcagc ctgagcttgtt gaggccctgtc ctgcgtggcc 60
cgaagctg 68

<210> SEQ ID NO 523
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 523

gaggccctat tcccatgat ttcttca 27

<210> SEQ ID NO 524
<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: (84) ...(103)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 524

aagaaaaagca ccgacgtcggt ggcacttttt tcaagttgata acggacagtgc cttatatatttt 60
c tgtgtctatt ctgtctctaa aacmmmmnm nnnnnnnnn nnnngttggtt cgtcttctc 120
acaag 125

<210> SEQ ID NO 525
<211> LENGTH: 111
<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: (9) ...(28)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 525

gaaacacmm nnnnnnnnn nnnnnnnngtt ttttagacga gsaatagoaa gttaaataaa 60
ggtagtctgg ttatcaacctt gaaaaagttgg caccgagtgc gttcttttttt t 111

<210> SEQ ID NO 526
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 526

agctgatatta cttgatggtct 20

<210> SEQ ID NO 527
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
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,

2. The method of claim 1, wherein the expression of two or more gene products is altered.

3. The method of claim 1, wherein the CRISPR-Cas system comprises a trans-activating cr (tracr) sequence.

4. The method of claim 1, wherein at least one NLS is at or within 50 amino acids of the amino-terminus of the Cas9 protein and/or at least one NLS is at or within 50 amino acids of the carboxy-terminus of the Cas9 protein.

5. The method of claim 1, wherein the one or more vectors are viral vectors.

6. The method of claim 5, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, aden-associated and herpes simplex viral vectors.

7. The method of claim 1, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

8. The method of claim 1, wherein the eukaryotic cell is a mammalian or human cell.

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. A CRISPR-Cas system-mediated genome editing method comprising introducing into a eukaryotic cell containing a polynucleotide having a sequence wherein the polynucleotide sequence includes a target sequence, 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,

12. The method of claim 11, wherein the sequence-specific genome editing comprises creation of a double strand break (DSB) which is repaired by a non-homologous end joining (NHEJ) cell repair mechanism generating indels thereby modifying the polynucleotide sequence.
13. The method of claim 11, wherein the sequence-specific genome editing comprises creation of a DSB which is repaired by a homologous recombination (HR) cell repair mechanism incorporating a HR template into the polynucleotide thereby modifying the polynucleotide sequence.

14. The method of claim 11, wherein at least one NLS is at or within 50 amino acids of the amino-terminus of the Cas9 protein or/and at least one NLS is at or within 50 amino acids of the carboxy-terminus of the Cas9 protein.

15. The method of claim 11, wherein the CRISPR-Cas system comprises a tracr sequence.

16. The method of claim 11, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

17. The method of claim 11, wherein the eukaryotic cell is a mammalian or human cell.

18. 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 a target sequence of a DNA molecule in a eukaryotic cell that contains the DNA molecule, wherein the DNA molecule encodes and the eukaryotic cell expresses at least one gene product,
   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 tile system,
   wherein the CRISPR-Cas system comprises a tracr sequence,
   wherein the CRISPR-Cas system comprises two or more nuclear localization signals (NLSs),
   whereby the guide RNA targets and hybridizes with the target sequence and the Cas9 protein cleaves 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.

19. The system of claim 18, wherein at least one NLS is at or within 50 amino acids of the amino-terminus of the Cas9 protein and/or at least one NLS is at or within 50 amino acids of the carboxy-terminus of the Cas9 protein.

20. The system of claim 18, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

21. The system of claim 18, wherein the eukaryotic cell is a mammalian or human cell.

22. The system of claim 18, wherein the expression of one or more gene products is increased.

23. The system of claim 18, wherein the expression of one or more gene products is decreased.

24. The system of claim 18, wherein the one or more vectors are viral vectors.

25. The system of claim 24, wherein the one or more viral vectors are selected from the group consisting of retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors.

26. 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 CRISPR-Cas system comprises two or more nuclear localization signals (NLSs), wherein the DNA molecule encodes and the eukaryotic cell expresses at least one gene product and the Cas9 protein cleaves 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.

27. The CRISPR-Cas system of claim 26, wherein at least one NLS is at or within 50 amino acids of the amino-terminus of the Cas9 protein and/or at least one NLS is at or within 50 amino acids of the carboxy-terminus of the Cas9 protein.

28. The CRISPR-Cas system of claim 26, wherein the CRISPR-Cas system comprises a tracr sequence.

29. The CRISPR-Cas system of claim 26, wherein the Cas9 protein is codon optimized for expression in the eukaryotic cell.

30. The CRISPR-Cas system of claim 26, wherein the eukaryotic cell is a mammalian or human cell.