Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2841710 C 2021/03/16

(11)(21) 2 841 710

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2012/07/12
(87) Date publication PCT/PCT Publication Date: 2013/01/24
(45) Date de délivrance/Issue Date: 2021/03/16
(85) Entrée phase nationale/National Entry: 2014/01/14
(86) N° demande PCT/PCT Application No.: US 2012/046451
(87) N° publication PCT/PCT Publication No.: 2013/012674
(30) Priorités/Priorities: 2011/07/15 (US61/508,366); 2012/02/21 (US61/601,409); 2012/03/13 (US61/610,212)

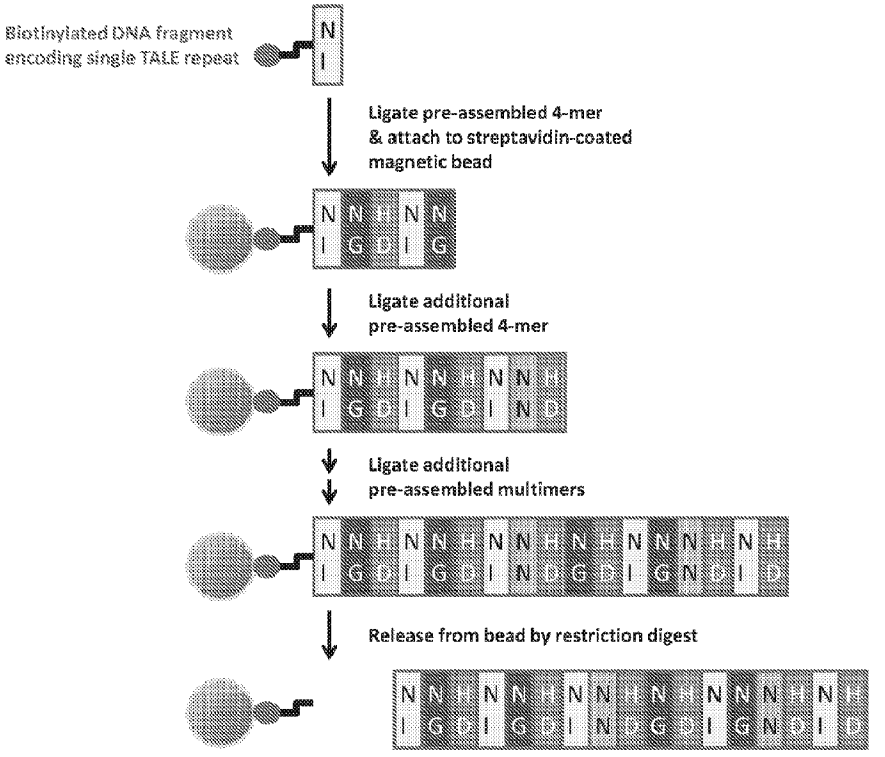
(51) Cl.Int./Int.Cl. *C12N 15/66* (2006.01), *C07K 14/24* (2006.01), *C07K 19/00* (2006.01), *C12N 1/00* (2006.01), *C12N 15/31* (2006.01), *C12N 15/62* (2006.01), *C12N 5/00* (2006.01), *C40B 40/06* (2006.01)

(72) Inventeurs/Inventors:
JOUNG, J. KEITH, US;
SANDER, JEFFRY D., US

(73) Propriétaire/Owner:
THE GENERAL HOSPITAL CORPORATION, US

(74) Agent: SMART & BIGGAR LLP

(54) Titre : PROCEDES D'ASSEMBLAGE D'EFFECTEURS DE TYPE ACTIVATEUR DE LA TRANSCRIPTION
(54) Title: METHODS OF TRANSCRIPTION ACTIVATOR LIKE EFFECTOR ASSEMBLY




(57) Abrégé/Abstract:
The disclosure describes methods that include providing a first nucleic acid having a sequence encoding a first set comprising one or more transcription activator-like effector (TALE) repeat domains and/or one or more portions of one or more TALE repeat

(11)(21) 2 841 710
(13) C

(57) Abrégé(suite)/Abstract(continued):
domains; contacting the first nucleic acid with a first enzyme, wherein the first enzyme creates a first ligatable end; providing a second nucleic acid having a sequence encoding a second set comprising one or more TALE repeat domains and/or one or more portions of one or more TALE repeat domains; contacting the second nucleic acid with a second enzyme, wherein the second enzyme creates a second ligatable end, and wherein the first and second ligatable ends are compatible; and ligating the first and second nucleic acids through the first and second ligatable ends to produce a first ligated nucleic acid, wherein the first ligated nucleic acid is linked to a solid support, and wherein the first ligated nucleic acid encodes a polypeptide comprising said first and second sets.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 January 2013 (24.01.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/012674 A1

(51) International Patent Classification:
*C12N 15/66* (2006.01) *C07K 14/24* (2006.01)
*C12N 15/62* (2006.01) *C12N 1/00* (2006.01)
*C12N 15/31* (2006.01) *C12N 5/00* (2006.01)
*C12Q 1/68* (2006.01) *C40B 40/06* (2006.01)
*C07K 19/00* (2006.01)

(21) International Application Number:
PCT/US2012/046451

(22) International Filing Date:
12 July 2012 (12.07.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/508,366 15 July 2011 (15.07.2011) US
61/601,409 21 February 2012 (21.02.2012) US
61/610,212 13 March 2012 (13.03.2012) US

(71) Applicant *(for all designated States except US)*: **THE GENERAL HOSPITAL CORPORATION** [US/US]; 55 Fruit Street, Boston, MA 02114 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **JOUNG, J., Keith** [US/US]; 1 Magnolia Way, Winchester, MA 01890 (US). **SANDER, Jeffry, D.** [US/US]; 111 Locust St Bldg #3 Apt A1, Woburn, MA 01801 (US).

(74) Agents: **LODOVICE, Ian, J.** et al.; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, MN 55440-1022 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

*[Continued on next page]*

(54) Title: METHODS OF TRANSCRIPTION ACTIVATOR LIKE EFFECTOR ASSEMBLY

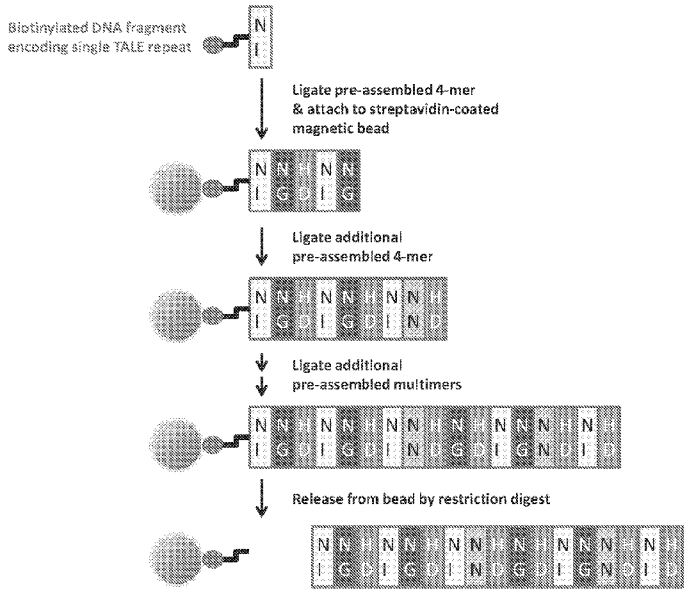



FIG. 1

(57) Abstract: The disclosure describes methods that include providing a first nucleic acid having a sequence encoding a first set comprising one or more transcription activator-like effector (TALE) repeat domains and/or one or more portions of one or more TALE repeat domains; contacting the first nucleic acid with a first enzyme, wherein the first enzyme creates a first ligatable end; providing a second nucleic acid having a sequence encoding a second set comprising one or more TALE repeat domains and/or one or more portions of one or more TALE repeat domains; contacting the second nucleic acid with a second enzyme, wherein the second enzyme creates a second ligatable end, and wherein the first and second ligatable ends are compatible; and ligating the first and second nucleic acids through the first and second ligatable ends to produce a first ligated nucleic acid, wherein the first ligated nucleic acid is linked to a solid support, and wherein the first ligated nucleic acid encodes a polypeptide comprising said first and second sets.

WO 2013/012674 A1

WO 2013/012674 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published**:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

# METHODS OF TRANSCRIPTION ACTIVATOR LIKE EFFECTOR ASSEMBLY

## CLAIM OF PRIORITY

## STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

This invention was made with government support under grant number DP1 OD006862 awarded by the National Institutes of Health. The government has certain rights in the invention.

## TECHNICAL FIELD

This invention relates to methods of producing nucleic acids encoding peptides and polypeptides encoding multiple transcription-like activator effector (TALE) repeat domains and the proteins themselves.

## BACKGROUND

TALE proteins of plant pathogenic bacteria in the genus Xanthomonas play important roles in disease, or trigger defense, by binding host DNA and activating effector-specific host genes (see, e.g., Gu et al., 2005, Nature 435:1122; Yang et al., 2006 Proc. Natl. Acad. Sci. USA 103:10503; Kay et al., 2007, Science 318:648; Sugio et al., 2007, Proc. Natl. Acad. Sci. USA 104:10720; and Romer et al., 2007, Science 318:645).

Specificity for nucleic acid sequences depends on an effector-variable number of imperfect, typically ~33-35 amino acid repeats (Schornack et al., 2006, J. Plant Physiol. 163:256). Each repeat binds to one nucleotide in the target sequence, and the specificity of each repeat for its nucleotide is largely context-independent, allowing for the development of custom sequence-specific TALE proteins (Moscou et al., 2009, Science 326:1501; Boch et al., 2009, Science 326:1509-1512).

## SUMMARY

This application is based, at least in part, on the development of rapid, simple, and easily automatable methods for assembling nucleic acids encoding custom TALE repeat array proteins.

Accordingly, this disclosure features a process that includes: (a) providing a first nucleic acid having a sequence encoding a first set comprising one or more (e.g., two or more, three or more, four or more, five or more, six or more, one to six, two to six, three to six, four to six, five or six, one two to five, three to five, four or five, one to four, two to four, three or four, one to three, two or three, one or two, one, two, three, four, five, or six) transcription activator-like effector (TALE) repeat domains and/or one or more portions of one or more TALE repeat domains; (b) contacting the first nucleic acid with a first enzyme, wherein the first enzyme creates a first ligatable end; (c) providing a second nucleic acid having a sequence encoding a second set comprising one or more (e.g., two or more, three or more, four or more, five or more, six or more, one to six, two to six, three to six, four to six, five or six, one two to five, three to five, four or five, one to four, two to four, three or four, one to three, two or three, one or two, one, two, three, four, five, or six) TALE repeat domains and/or one or more portions of one or more TALE repeat domains; (d) contacting the second nucleic acid with a second enzyme, wherein the second enzyme creates a second ligatable end, and wherein the first and second ligatable ends are compatible; and (e) ligating the first and second nucleic acids through the first and second ligatable ends to produce a first ligated nucleic acid, wherein the first ligated nucleic acid is linked to a solid support, and wherein the first ligated nucleic acid encodes a polypeptide comprising said first and second sets.

In some embodiments, the disclosure relates to a process for assembling nucleic acids encoding custom TALE repeat array proteins comprising: (a) providing a first nucleic acid comprising a sequence encoding a first set comprising one or more transcription activator-like effector (TALE) repeat domains; (b) contacting the first nucleic acid with a first enzyme, wherein the first enzyme creates a first ligatable end; (c) providing a second nucleic acid comprising a sequence encoding a second set comprising one or more TALE repeat domains; (d) contacting the second nucleic acid with a second enzyme, wherein the second enzyme creates a second ligatable end, and wherein the first and second ligatable ends are compatible; and (e) ligating the first and second nucleic acids through the first and second ligatable ends to produce a first ligated nucleic acid, wherein the first ligated nucleic acid is linked to a solid support, and wherein the first ligated nucleic acid encodes a polypeptide comprising said first and second sets comprising one or more TALE repeat domains.

In some embodiments, the methods include linking the first nucleic acid to a solid support prior to (b) contacting the first nucleic acid with the first enzyme or prior to (e) ligating the first and second nucleic acids. In some embodiments, the methods include linking the first ligated nucleic acid to a solid support.

In some embodiments, the first set is N-terminal to the second set in the polypeptide. In some embodiments, the second set is N-terminal to the first set in the polypeptide.

In some embodiments, the first and second enzymes are a first and second restriction endonuclease, wherein the first restriction endonuclease cleaves at a site within the first nucleic acid and creates a first cut end, and the second restriction endonuclease cleaves at a site within the second nucleic acid and creates a second cut end, and wherein the first and second ligatable ends are the first and second cut ends. When restriction endonucleases are used, the first ligated nucleic acid cannot include a restriction site recognized by the first restriction endonuclease.

The process can further include: (f) contacting the first ligated nucleic acid with a third enzyme, wherein the third enzyme creates a third ligatable end; (g) providing a third nucleic acid comprising a sequence encoding a third set comprising one or more (e.g., two or more,

three or more, four or more, five or more, six or more, one to six, two to six, three to six, four to six, five or six, one two to five, three to five, four or five, one to four, two to four, three or four, one to three, two or three, one or two, one, two, three, four, five, or six) TALE repeat domains and/or one or more portions of one or more TALE repeat domains; (h) contacting the third nucleic acid with a fourth enzyme, wherein the fourth enzyme creates a fourth ligatable end, and wherein the third and fourth ligatable ends are compatible; and (i) ligating the first ligated and third nucleic acids through the third and fourth ligatable ends to produce a second ligated nucleic acid linked to the solid support, wherein the second ligated nucleic acid encodes a polypeptide comprising said first, second, and third sets.

In some embodiments, the third and fourth enzymes are a third and fourth restriction endonuclease, wherein the third restriction endonuclease cleaves at a site within the first ligated nucleic acid and creates a third cut end, and the fourth restriction

endonuclease cleaves at a site within the third nucleic acid and creates a fourth cut end, and wherein the third and fourth ligatable ends are the third and fourth cut ends.

In some embodiments, the ligated nucleic acid does not include a restriction site recognized by the first endonuclease, and the first and third restriction endonucleases are the same. In some embodiments, the second and fourth restriction endonucleases are the same.

The process can further include: (j) contacting the second ligated nucleic acid with a fifth enzyme, wherein the fifth enzyme creates a fifth ligatable end; (k) providing a fourth nucleic acid having a sequence encoding a fourth set comprising one or more (e.g., two or more, three or more, four or more, five or more, six or more, one to six, two to six, three to six, four to six, five or six, one two to five, three to five, four or five, one to four, two to four, three or four, one to three, two or three, one or two, one, two, three, four, five, or six) TALE repeat domains and/or one or more portions of one or more TALE repeat domains; (l) contacting the fourth nucleic acid with a sixth enzyme, wherein the sixth enzyme creates a sixth ligatable end, and wherein the fifth and sixth ligatable ends are compatible; and (m) ligating the second ligated and fourth nucleic acids through the fifth and sixth ligatable ends to produce a third ligated nucleic acid linked to the solid support, wherein the third ligated nucleic acid encodes a polypeptide comprising said first, second, third, and fourth sets. One of ordinary skill would recognize that the process can be repeated with similar additional steps. Such methods are included within this disclosure.

In some embodiments, the fifth and sixth enzymes are a fifth and sixth restriction endonuclease, wherein the fifth restriction endonuclease cleaves at a site within the second ligated nucleic acid and creates a fifth cut end, and the sixth restriction endonuclease cleaves at a site within the fourth nucleic acid and creates a sixth cut end, and wherein the fifth and sixth ligatable ends are the fifth and sixth cut ends.

In some embodiments, the second ligated nucleic acid does not include a restriction site recognized by the first endonuclease, and the first, third, and fifth restriction endonucleases are the same.

In some embodiments, the second, fourth, and sixth restriction endonucleases are the same.

In some embodiments, the solid support and linked nucleic acid are isolated, e.g., following any of the above steps (a) – (m).

In some embodiments, the second, third, or fourth set comprises one to four TALE repeat domains.

In some embodiments, the ligatable ends include an overhang of 1-10 nucleotides. In some embodiments, the ligatable ends are blunt ends. In some embodiments, an overhang can be generated using an exonuclease and polymerase in the presence of one or more nucleotides.

In some embodiments, an enzyme or restriction endonuclease used in the above processes is a type IIS restriction endonuclease.

The processes can further comprise unlinking a ligated nucleic acid from the solid support and inserting the ligated nucleic acid (or a processed derivative thereof comprising the TALE repeat array coding sequences) into a vector, e.g., an expression vector. The expression vector can include a sequence encoding an effector domain (e.g., a nuclease domain) configured to create a sequence encoding a fusion protein of the polypeptide and the effector domain. The expression vector can be inserted into a cell to affect the cell directly or for expression of the polypeptide or fusion protein. When the polypeptide or fusion protein is to be expressed, the processes can further include expressing and purifying the polypeptide or fusion protein.

In another aspect, this disclosure features TALE proteins that bind to a target nucleotide sequence (e.g., a "half site") disclosed herein (e.g., in Table 6 or 7), TALE nucleases that include the TALE proteins, pairs of TALE proteins (e.g., TALENs) that bind to the target sites disclosed herein (e.g., in Table 6 or 7), and nucleic acids that encode any of the above. In some embodiments, the TALE proteins, TALE nucleases, and pairs of TALE proteins (e.g., TALENs) are those disclosed in Example 7. The nucleic acids encoding the TALE proteins, TALE nucleases, and pairs of TALE proteins (e.g., TALENs) can be those disclosed in Example 7 or other sequences that encode the proteins disclosed in Example 7. The disclosure also includes vectors and cells that

include the nucleic acids encoding the TALE proteins, TALE nucleases, or pairs of TALE proteins (e.g., TALENs) disclosed herein and methods of expressing the TALE proteins, TALE nucleases, or pairs of TALE proteins (e.g., TALENs) that include culturing the cells. The methods of expressing the TALE proteins, TALE nucleases, or pairs of TALE proteins (e.g., TALENs) can also include isolating the TALE proteins, TALE nucleases, or pairs of TALE proteins (e.g., TALENs) from the cell culture.

In another aspect, the invention features a set, archive, or library of nucleic acids (e.g., plasmids) that include sequences encoding one or more TALE domains. In some embodiments, the set, archive, or library includes sequences encoding one, two, three, and/or four (or more than four (e.g., five, six, or more)) TALE repeat domains. In some embodiments, the set, library, or archive of nucleic acids includes sequences encoding TALE repeat domains that bind to nucleotide sequences having one, two, three, four (or more than four (e.g., five, six, or more)) nucleotides. In some embodiments, the set, library, or archive includes restriction sites (e.g., sites for type IIS restriction endonucleases) surrounding the sequences encoding the TALE repeat domains.

The methods described herein provide several advantages, including avoiding extensive PCR amplification of the TALE repeats, thereby avoiding the introduction of mutations from PCR errors. Further, TALE repeat arrays of any desired length can be constructed, and the methods can be easily multiplexed and/or automated.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and

advantages of the invention will be apparent from the description and drawings, and from the claims.

## DESCRIPTION OF DRAWINGS

FIG. 1 is a schematic depiction of an exemplary method of assembling a nucleic acid encoding a TALE protein.

FIG. 2 is a schematic depiction of exemplary archives of nucleic acids encoding single (one-mer), two-mer, three-mer, and four-mer TALE repeat domains.

FIG. 3 depicts the sequence of the pUC57-ΔBsaI plasmid. This plasmid is identical to plasmid pUC57 except for mutation of a single base (in bold, underlined and lowercase) that destroys a BsaI restriction site.

FIG. 4A depicts the polypeptide sequences of exemplary TALE repeats of type α/ε, β, γ, and δ. Polymorphic residues characteristic of each type are indicated in bold and italic. The hypervariable triplet SNI for binding to A is indicated in underscore.

FIG. 4B depicts the polynucleotide sequences of the exemplary TALE repeats of FIG. 4A.

FIGs. 5A-5B depict the common sequence of expression plasmids pJDS70, pJDS71, pJDS74, pJDS76, and pJDS78. The region of the variable sequences is depicted as XXXXXXXXX (underlined and bold).

FIG. 6 is a schematic diagram of the enhanced green fluorescent protein (eGFP) gene and the location of the binding sites for synthetic TALE proteins described herein.

FIG. 7 is a bar graph depicting the % of TALE nuclease-modified, eGFP-negative cells at 2 and 5 days following transfection with plasmids encoding TALE nucleases designed to bind and cleave the eGFP reporter gene.

FIG. 8 is a depiction of the sequences of insertion-deletion mutants of eGFP induced by TALE nucleases. Deleted bases are indicated by dashes and inserted bases indicated by double underlining; the TALEN target half-sites are single underlined. The net number of bases inserted or deleted is shown to the right.

FIG. 9 is a depiction of an electrophoresis gel of assembled DNA fragments encoding 17-mer TALE array preparations.

FIG. 10 is a depiction of an electrophoresis gel of 16-mer TALE array preparations.

FIGs. 11A-11B depict the nucleotide (11A) and polypeptide (11B) sequence of engineered DR-TALE-0003.

FIGs. 12A-12B depict the nucleotide (12A) and polypeptide (12B) sequence of engineered DR-TALE-0006.

FIGs. 13A-13B depict the nucleotide (13A) and polypeptide (13B) sequence of engineered DR-TALE-0005.

FIGs. 14A-14B depict the nucleotide (14A) and polypeptide (14B) sequence of engineered DR-TALE-0010.

FIGs. 15A-15B depict the nucleotide (15A) and polypeptide (15B) sequence of engineered DR-TALE-0023.

FIGs. 16A-16B depict the nucleotide (16A) and polypeptide (16B) sequence of engineered DR-TALE-0025.

FIGs. 17A-17B depict the nucleotide (17A) and polypeptide (17B) sequence of engineered DR-TALE-0020.

FIGs. 18A-18B depict the nucleotide (18A) and polypeptide (18B) sequence of engineered DR-TALE-0022.

FIG. 19A is a bar graph depicting activities of 48 TALEN pairs and four ZFN pairs in the EGFP gene-disruption assay. Percentages of EGFP-negative cells as measured 2 and 5 days following transfection of U2OS cells bearing a chromosomally integrated EGFP reporter gene with nuclease-encoding plasmids are shown. Mean percent disruption of EGFP and standard error of the mean from three independent transfections are shown.

FIG. 19B is a bar graph depicting mean EGFP-disruption activities from FIG. 19A, grouped by length of the TALENs.

FIG. 20A is a graph depicting the ratio of mean percent EGFP disruption values from day 2 to day 5. Ratios were calculated for groups of each length TALEN using the data from FIG. 19B. Values greater than 1 indicate a decrease in the average of EGFP-disrupted cells at day 5 relative to day 2.

FIG. 20B is a graph depicting the ratio of mean tdTomato-positive cells from day 2 to day 5 grouped by various lengths of TALENs. tdTomato-encoding control plasmids were transfected together with nuclease-encoding plasmids on day 0.

FIGs. 21A-E depict DNA sequences and frequencies of assembled TALEN-induced mutations at endogenous human genes. For each endogenous gene target, the wild-type (WT) sequence is shown at the top with the TALEN target half-sites underlined and the translation start codon of the gene (ATG) indicated by a box. Deletions are indicated by dashes and insertions by lowercase letters and double underlining. The sizes of the insertions (+) or deletions (Δ) are indicated to the right of each mutated site. The number of times that each mutant was isolated is shown in parentheses. Mutation frequencies are calculated as the number of mutants identified divided by the total number of sequences analyzed. Note that for several of the genes, we also identified larger deletions that extend beyond the sequences of the TALEN target sites.

FIG. 22 is a schematic depiction of an exemplary method of assembling a nucleic acid encoding a TALE protein containing TALE repeat domains or portions of TALE repeat domains.

## DETAILED DESCRIPTION

The methods described herein can be used to assemble engineered proteins containing TALE repeat domains for binding to specific sequences of interest. Assembling long arrays (e.g., 12 or more) of TALE repeat domain repeats can be challenging because the repeats differ only at a small number of amino acids within their highly conserved ~33-35 amino acid consensus sequence. PCR assembly can lead to the introduction of unwanted mutations. Hierarchical assembly methods that involve one or more passages of intermediate plasmid constructs in *E. coli* can also be problematic because the highly repetitive nature of these constructs can make them unstable and prone to recombination and because the need to passage these intermediate constructs makes these approaches difficult to automate.

TAL Effectors

TAL effectors of plant pathogenic bacteria in the genus Xanthomonas play important roles in disease, or trigger defense, by binding host DNA and activating effector-specific host genes. Specificity depends on an effector-variable number of imperfect, typically ~33-35 amino acid repeats. Polymorphisms are present primarily at repeat positions 12 and 13, which are referred to herein as the "repeat variable-diresidue" (RVD). The RVDs of TAL effectors correspond to the nucleotides in their target sites in a direct, linear fashion, one RVD to one nucleotide, with some degeneracy and no apparent context dependence. In some embodiments, the polymorphic region that grants nucleotide specificity may be expressed as a triresidue or triplet e.g., encompassing residues 11, 12, and 13.

Each DNA binding repeat can include an RVD that determines recognition of a base pair in the target DNA sequence, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA sequence, and wherein the RVD comprises, but is not limited to, one or more of the following: HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; SN for recognizing G or A; YG for recognizing T; and NK for recognizing G, and one or more of: HD for recognizing C; NG for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T, wherein * represents a gap in the second position of the RVD; HG for recognizing T; H* for recognizing T, wherein * represents a gap in the second position of the RVD; and IG for recognizing T.

TALE proteins are useful in research and biotechnology as targeted chimeric nucleases that can facilitate homologous recombination in genome engineering (e.g., to add or enhance traits useful for biofuels or biorenewables in plants). These proteins also are useful as, for example, transcription factors, and especially for therapeutic applications requiring a very high level of specificity such as therapeutics against pathogens (e.g., viruses) as non-limiting examples.

Assembly Methods

An example of the methods described herein of assembling a TALE repeat domain array is shown in FIG. 1 and includes the following steps: (1) provision a single biotinylated PCR product encoding one single N-terminal TALE repeat domain (a one-mer) with a linker suitable for attachment to a solid support (in the example shown here, a magnetic streptavidin coated bead is used but other solid supports can also be utilized as well as other ways of tethering the initial DNA fragment to the solid support); (2) creation of an overhang at the 3' end of the one-mer DNA (e.g., using a Type IIS restriction enzyme); (3) ligation of a second fragment containing four TALE repeat domain (i.e., a pre-assembled four-mer), creating a five-mer; (4) attachment of the five-mer to the solid support; (5) ligation of additional pre-assembled TALE repeat domains to create a long array, e.g., a piece or pieces of DNA encoding one, two, three, or four TALE repeat domains depending upon the length of the desired final array, and (6) release of the extended DNA encoding the TALE repeats from the solid support (e.g., by using a Type IIS restriction enzyme whose site is built in at the 5' end of the initial biotinylated DNA product). The final fragment can then be prepared for ligation to an appropriate expression plasmid.

Alternatively, the method can proceed as follows: (1) attachment of a single biotinylated PCR product encoding one single N-terminal TALE repeat domains to a solid support (in the example shown here, a magnetic streptavidin coated bead is used but other solid supports such as the streptavidin-coated wells of a multi-well plate can also be utilized as well as other ways of tethering the initial DNA fragment to the solid support), (2) creation of an overhang at the 3' end of the anchored DNA (e.g., using a Type IIS restriction enzyme), (3) ligation of a second fragment containing four TALE repeat domain, (4) additional cycles of steps (2) and (3) to create a long array, (5) in the final cycle performing ligation of a piece of DNA encoding one, two, three, or four TALE repeat domains depending upon the length of the desired final array, and (6) release of the extended DNA encoding the TALE repeats from the solid support (e.g., by using a Type IIS restriction enzyme whose site is built in at the 5' end of the initial biotinylated DNA product).

Another example of a method of assembling a TALE repeat domain array based on the methods described herein is shown in FIG. 22 and includes the following steps: (1) provision a single biotinylated PCR product encoding a portion of one single N-terminal TALE repeat domain (a partial one-mer) with a linker suitable for attachment to a solid support (in the example shown here, a magnetic streptavidin coated bead is used but other solid supports can also be utilized as well as other ways of tethering the initial DNA fragment to the solid support); (2) creation of an overhang at the 3' end of the partial one-mer DNA (e.g., using a Type IIS restriction enzyme); (3) ligation of a second fragment containing consisting of two partial and three full TALE repeats; (4) attachment of the second fragment to the solid support; (5) ligation of additional pre-assembled TALE repeat domains or portions of TALE repeat domains to create a long array, e.g., a piece or pieces of DNA encoding one, two, three, or four TALE repeat domains (or portions of TALE repeat domains) depending upon the length of the desired final array, and (6) release of the extended DNA encoding the TALE repeats from the solid support (e.g., by using a Type IIS restriction enzyme whose site is built in at the 5' end of the initial biotinylated DNA product). The final fragment can then be prepared for ligation to an appropriate expression plasmid.

The initial nucleic acid encoding one or more TALE repeat domains (or portions) is linked to a solid support. The initial nucleic acid can be prepared by any means (e.g., chemical synthesis, PCR, or cleavage from a plasmid). Additionally, the nucleic acid can be linked to the solid support by any means, e.g., covalently or noncovalently.

In some embodiments, the nucleic acid is linked noncovalently by using a nucleic acid modified with one member of a binding pair and incorporating the other member of the binding pair on the solid support. A member of a binding pair is meant to be one of a first and a second moiety, wherein said first and said second moiety have a specific binding affinity for each other. Suitable binding pairs for use in the invention include, but are not limited to, antigens/antibodies (for example, digoxigenin/anti-digoxigenin, dinitrophenyl (DNP)/anti-DNP, dansyl-X/anti-dansyl, Fluorescein/anti-fluorescein, lucifer yellow/anti-lucifer yellow, peptide/anti-peptide, ligand/receptor and rhodamine/anti-rhodamine), biotin/avidin (or biotin/streptavidin) and calmodulin binding

protein (CBP)/calmodulin. Other suitable binding pairs include polypeptides such as the FLAG-peptide (Hopp et al., 1988, BioTechnology, 6:1204 10); the KT3 epitope peptide (Martin et al., Science 255:192 194 (1992)); tubulin epitope peptide (Skinner et al., J. Biol. Chem. 266:15163-66 (1991)); and the T7 gene 10 protein peptide tag (Lutz-Freyerinuth et al., Proc. Natl. Acad. Sci. USA, 87:6393 97 (1990)) and the antibodies each thereto.

In some embodiments, the individual nucleic acids encoding one or more TALE repeat domains are present in an archive or library of plasmids (see FIG. 2). Although nucleic acids encoding one to four TALE repeat domains are shown, the library of plasmids can contain nucleic acids encoding more than four (e.g., five, six, or more) TALE repeat domains. Alternatively, as shown Figure 22, the nucleic acids encoding parts or portions of one or more TALE repeat domains can also be joined together to create final DNA fragments encoding the desired full-length arrays of TALE repeat domains. Numerous TALE repeat domain sequences with binding specificity for specific nucleotides or sets of nucleotides are known in the art, and one of ordinary skill can design and prepare a library of plasmids based on these known sequences and the disclosures herein.

As used herein, a solid support refers to any solid or semisolid or insoluble support to which the nucleic acid can be linked. Such materials include any materials that are used as supports for chemical and biological molecule syntheses and analyses, such as, but not limited to: polystyrene, polycarbonate, polypropylene, nylon, glass, dextran, chitin, sand, pumice, agarose, polysaccharides, dendrimers, buckyballs, polyacryl-amide, silicon, rubber, and other materials used as supports for solid phase syntheses, affinity separations and purifications, hybridization reactions, immunoassays and other such applications. The solid support can be particulate or can be in the form of a continuous surface, such as a microtiter dish or well, a glass slide, a silicon chip, a nitrocellulose sheet, nylon mesh, or other such materials. When particulate, typically the particles have at least one dimension in the 5-10 mm range or smaller. Such particles, referred collectively herein as "beads," are often, but not necessarily, spherical. Such reference, however, does not constrain the geometry of the matrix, which can be any

shape, including random shapes, needles, fibers, and elongated. Roughly spherical "beads," particularly microspheres that can be used in the liquid phase, also are contemplated. The "beads" can include additional components, such as magnetic or paramagnetic particles (see, e.g., Dynabeads (Dynal, Oslo, Norway)) for separation using magnets, as long as the additional components do not interfere with the methods described herein.

The ligatable ends can be produced by cutting with a restriction endonuclease (e.g., a type II or type IIS restriction endonuclease) or by "chewing back" the end using an enzyme (or enzymes) with exonuclease and polymerase activities in the presence of one or more nucleotides (see, Aslanidis et al., 1990, Nucl. Acids Res., 18:6069-74). Suitable enzymes are known to those of ordinary skill in the art. When restriction endonucleases are used, the nucleic acids can be designed to include restriction sites for the enzymes at suitable locations.

Following a ligation reaction, any unligated ends with 5' or 3' overhangs can be "blunted" by use of a polymerase, e.g., a DNA polymerase with both 3'→ 5' exonuclease activity and 5'→ 3' polymerase activity. This blunting step can reduce the appearance of undesired or partial assembly products. Alternatively, these ends can be capped using either a "hairpin" oligo bearing a compatible overhang (Briggs et al., 2012, Nucleic Acids Res, PMID: 22740649) or by short double-stranded DNAs bearing a compatible overhang on one end and a blunt end on the other.

To prepare the ligated nucleic acid for further downstream processing, it can be useful to select nucleic acids of the expected size, to reduce the presence of minor products created by incomplete ligations. Methods of selecting nucleic acids by size are known in the art, and include gel electrophoresis (e.g., slab gel electrophoresis or capillary gel electrophoresis (see, e.g., Caruso et al., 2003, Electrophoresis, 24:1-2:78-85)), liquid chromatography (e.g., size exclusion chromatography or reverse phase chromatography (see, e.g., Huber et al., 1995, Anal. Chem., 67:578-585)), and lab-on-a-chip systems (e.g., LabChip® XT system, Caliper Life Sciences, Hopkinton, MA). In some embodiments, a size exclusion step can be performed using an automated system,

e.g., an automated gel electrophoresis system (e.g., a Pippin Prep™ automated DNA size selection system, Sage Science, Beverly, MA).

Automation

The methods disclosed herein can be performed manually or implemented in laboratory automation hardware (e.g., SciClone G3 Liquid Handling Workstation, Caliper Life Sciences, Hopkinton, MA) controlled by a compatible software package (e.g., Maestro™ liquid handling software) programmed according to the new methods described herein or a new software package designed and implemented to carry out the specific method steps described herein. When performed by laboratory automation hardware, the methods can be implemented by computer programs using standard programming techniques following the method steps described herein.

Examples of automated laboratory system robots include the Sciclone™ G3 liquid handling workstation (Caliper Life Sciences, Hopkinton, MA), Biomek® FX liquid handling system (Beckman-Coulter, Fullerton, California), TekBench™ automated liquid handling platform (TekCel, Hopkinton, Massachusetts), and Freedom EVO® automation platform (Tecan Trading AG, Switzerland).

The programs can be designed to execute on a programmable computer including at least one processor, at least one data storage system (including volatile and non-volatile memory and/or storage elements, e.g., RAM and ROM), at least one communications port that provides access for devices such as a computer keyboard, telephone, or a wireless, hand-held device, such as a PDA, and optionally at least one output device, such as a monitor, printer, or website. The central computer also includes a clock and a communications port that provides control of the lab automation hardware. These are all implemented using known techniques, software, and devices. The system also includes a database that includes data, e.g., data describing the procedure of one or more method steps described herein.

Program code is applied to data input by a user (e.g., location of samples to be processed, timing and frequency of manipulations, amounts of liquid dispensed or aspirated, transfer of samples from one location in the system to another) and data in the

database, to perform the functions described herein. The system can also generate inquiries and provide messages to the user. The output information is applied to instruments, e.g., robots, that manipulate, heat, agitate, etc. the vessels that contain the reactants as described herein. In addition, the system can include one or more output devices such as a telephone, printer, or a monitor, or a web page on a computer monitor with access to a website to provide to the user information regarding the synthesis and/or its progress.

Each program embodying the new methods is preferably implemented in a high level procedural or object-oriented programming language to communicate with a computer system. However, the programs can also be implemented in assembly or machine language if desired. In any case, the language can be a compiled or interpreted language.

Each such computer program is preferably stored on a storage medium or device (e.g., RAM, ROM, optical, magnetic) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The system can also be considered to be implemented as a computer- or machine-readable storage medium (electronic apparatus readable medium), configured with a program, whereby the storage medium so configured causes a computer or machine to operate in a specific and predefined manner to perform the functions described herein.

The new methods can be implemented using various means of data storage. The files can be transferred physically on recordable media or electronically, e.g., by email on a dedicated intranet, or on the Internet. The files can be encrypted using standard encryption software from such companies as RSA Security (Bedford, Massachusetts) and Baltimore®. The files can be stored in various formats, e.g., spreadsheets or databases.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; communications networks, including local area networks (LAN), wide area networks (WAN), Internet, Intranet, and Extranet;

electronic appliances such as a personal digital assistants (PDAs), cellular telephones, "smartphones," pagers and the like; and local and distributed processing systems.

As used herein, "stored" refers to a process for encoding information on an electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the sequence information.

A variety of software programs and formats can be used to store method data on an electronic apparatus readable medium. For example, the data and machine instructions can be incorporated in the system of the software provided with the automated system, represented in a word processing text file, formatted in commercially-available software such as WordPerfect® and Microsoft® Word®, or represented in the form of an ASCII file, stored in a database application, such as Microsoft Access®, Microsoft SQL Server®, Sybase®, Oracle®, or the like, as well as in other forms. Any number of data processor structuring formats (e.g., text file or database) can be employed to obtain or create a medium having recorded thereon the relevant data and machine instructions to implement the methods described herein.

By providing information in electronic apparatus readable form, the programmable computer can communicate with and control the lab automation hardware to perform the methods described herein. One skilled in the art can input data in electronic apparatus readable form (or a form that is converted to electronic apparatus readable form) to describe the completion of various method steps by the lab automation hardware.

Polypeptide Expression Systems

In order to use the engineered proteins of the present invention, it is typically necessary to express the engineered proteins from a nucleic acid that encodes them. This can be performed in a variety of ways. For example, the nucleic acid encoding the engineered TALE repeat protein is typically cloned into an intermediate vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or

insect vectors, for storage or manipulation of the nucleic acid encoding the engineered TALE protein or production of protein. The nucleic acid encoding the engineered TALE repeat protein is also typically cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell.

To obtain expression of a cloned gene or nucleic acid, the engineered TALE repeat protein is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook et al., Molecular Cloning, A Laboratory Manual (3d ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 2010). Bacterial expression systems for expressing the engineered TALE repeat protein are available in, e.g., *E. coli*, *Bacillus* sp., and *Salmonella* (Palva et al., 1983, Gene 22:229-235). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

The promoter used to direct expression of the engineered TALE repeat protein nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of the engineered TALE repeat protein. In contrast, when the engineered TALE repeat protein is to be administered in vivo for gene regulation, either a constitutive or an inducible promoter can be used, depending on the particular use of the engineered TALE repeat protein. In addition, a preferred promoter for administration of the engineered TALE repeat protein can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter typically can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tet-regulated systems and the RU-486 system (see, e.g., Gossen & Bujard, 1992, Proc. Natl. Acad. Sci. USA, 89:5547; Oligino et al., 1998, Gene Ther., 5:491-496; Wang et al., 1997, Gene Ther., 4:432-441; Neering et al., 1996, Blood, 88:1147-55; and Rendahl et al., 1998, Nat. Biotechnol., 16:757-761).

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the TALE repeat protein signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette can include, e.g., enhancers, and heterologous spliced intronic signals.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the engineered TALE repeat protein, e.g., expression in plants, animals, bacteria, fungus, protozoa, etc. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available fusion expression systems such as GST and LacZ. A preferred fusion protein is the maltose binding protein, "MBP." Such fusion proteins can be used for purification of the engineered TALE repeat protein. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, e.g., c-myc or FLAG.

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with the engineered TALE repeat protein encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (see, e.g., Colley et al., 1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (see, e.g., Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983).

Any of the well-known procedures for introducing foreign nucleotide sequences into host cells can be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

Characterization of TALE Proteins

Engineered TALE repeat array proteins designed using methods of the present invention can be further characterized to ensure that they have the desired characteristics for their chosen use. For example, TALE repeat array protein can be assayed using a bacterial two-hybrid, bacterial promoter repression, phage-display, or ribosome display system or using an electrophoretic mobility shift assay or "EMSA" (Buratowski & Chodosh, in Current Protocols in Molecular Biology pp. 12.2.1-12.2.7). Equally, any other DNA binding assay known in the art could be used to verify the DNA binding properties of the selected protein.

In one embodiment, a bacterial "two-hybrid" system is used to express and test a TALE repeat protein of the present invention. The bacterial two-hybrid system has an additional advantage, in that the protein expression and the DNA binding "assay" occur within the same cells, thus there is no separate DNA binding assay to set up.

Methods for the use of the bacterial two-hybrid system to express and assay DNA binding proteins are described in Joung et al., 2000, Proc. Natl. Acad. Sci. USA, 97:7382, Wright et al., 2006, Nat. Protoc, 1:1637-52; Maeder et al., 2008, Mol. Cell, 31:294-301; Maeder et al., 2009, Nat. Protoc., 4:1471-1501; and US Patent Application No. 2002/0119498. Briefly, in a bacterial two-hybrid system, the DNA binding protein is expressed in a bacterial strain bearing the sequence of interest upstream of a weak promoter controlling expression of a reporter gene (e.g., histidine 3 (HIS3), the beta-lactamase antibiotic resistance gene, or the beta-galactosidase (lacZ) gene). Expression of the reporter gene occurs in cells in which the DNA binding protein expressed by the cell binds to the target site sequence. Thus, bacterial cells expressing DNA binding proteins that bind to their target site are identified by detection of an activity related to the reporter gene (e.g., growth on selective media, expression of beta-galactosidase).

In some embodiments, calculations of binding affinity and specificity are also made. This can be done by a variety of methods. The affinity with which the selected TALE repeat array protein binds to the sequence of interest can be measured and quantified in terms of its $K_D$. Any assay system can be used, as long as it gives an accurate measurement of the actual $K_D$ of the TALE repeat array protein. In one embodiment, the $K_D$ for the binding of a TALE repeat array protein to its target is measured using an EMSA

In one embodiment, EMSA is used to determine the $K_D$ for binding of the selected TALE repeat array protein both to the sequence of interest (i.e., the specific $K_D$) and to non-specific DNA (i.e., the non-specific $K_D$). Any suitable non-specific or "competitor" double stranded DNA known in the art can be used. In some embodiments, calf thymus DNA or human placental DNA is used. The ratio of the non-specific $K_D$ to the specific KD is the specificity ratio. TALE repeat array proteins that bind with high specificity

have a high specificity ratio. This measurement is very useful in deciding which of a group of selected TALE should be used for a given purpose. For example, use of TALE repeat array protein in vivo requires not only high affinity binding but also high-specificity binding.

Construction of Chimeric TALE Proteins

Often, the aim of producing a custom-designed TALE repeat array DNA binding domain is to obtain a TALE repeat array protein that can be used to perform a function. The TALE repeat array DNA binding domain can be used alone, for example to bind to a specific site on a gene and thus block binding of other DNA-binding domains. However, in some embodiments, the TALE repeat array protein will be used in the construction of a chimeric TALE protein containing a TALE repeat array DNA binding domain and an additional domain having some desired specific function (e.g., gene activation) or enzymatic activity i.e., a "functional domain."

Chimeric TALE repeat array proteins designed and produced using the methods described herein can be used to perform any function where it is desired to target, for example, some specific enzymatic activity to a specific DNA sequence, as well as any of the functions already described for other types of synthetic or engineered DNA binding molecules. Engineered TALE repeat array DNA binding domains, can be used in the construction of chimeric proteins useful for the treatment of disease (see, for example, U.S. patent application 2002/0160940, and U.S. Pat. Nos. 6,511,808, 6,013,453 and 6,007,988, and International patent application WO 02/057308), or for otherwise altering the structure or function of a given gene in vivo. The engineered TALE repeat array proteins of the present invention are also useful as research tools, for example, in performing either in vivo or in vitro functional genomics studies (see, for example, U.S. Pat. No. 6,503,717 and U.S. patent application 2002/0164575).

To generate a functional recombinant protein, the engineered TALE repeat array DNA binding domain will typically be fused to at least one "functional" domain. Fusing functional domains to synthetic TALE repeat array proteins to form functional transcription factors involves only routine molecular biology techniques which are

commonly practiced by those of skill in the art, see for example, U.S. Pat. Nos. 6,511,808, 6,013,453, 6,007,988, 6,503,717 and U.S. patent application 2002/0160940).

Functional domains can be associated with the engineered TALE repeat array domain at any suitable position, including the C- or N-terminus of the TALE protein. Suitable "functional" domains for addition to the engineered protein made using the methods of the invention are described in U.S. Pat. Nos. 6,511,808, 6,013,453, 6,007,988, and 6,503,717 and U.S. patent application 2002/0160940.

In one embodiment, the functional domain is a nuclear localization domain which provides for the protein to be translocated to the nucleus. Several nuclear localization sequences (NLS) are known, and any suitable NLS can be used. For example, many NLSs have a plurality of basic amino acids, referred to as a bipartite basic repeats (reviewed in Garcia-Bustos et al, 1991, Biochim. Biophys. Acta, 1071:83-101). An NLS containing bipartite basic repeats can be placed in any portion of chimeric protein and results in the chimeric protein being localized inside the nucleus. It is preferred that a nuclear localization domain is routinely incorporated into the final chimeric protein, as the ultimate functions of the chimeric proteins of the present invention will typically require the proteins to be localized in the nucleus. However, it may not be necessary to add a separate nuclear localization domain in cases where the engineered TALE repeat array domain itself, or another functional domain within the final chimeric protein, has intrinsic nuclear translocation function.

In another embodiment, the functional domain is a transcriptional activation domain such that the chimeric protein can be used to activate transcription of the gene of interest. Any transcriptional activation domain known in the art can be used, such as for example, the VP16 domain form herpes simplex virus (Sadowski et al., 1988, Nature, 335:563-564) or the p65 domain from the cellular transcription factor NF-kappaB (Ruben et al., 1991, Science, 251:1490-93).

In yet another embodiment, the functional domain is a transcriptional repression domain such that the chimeric protein can be used to repress transcription of the gene of interest. Any transcriptional repression domain known in the art can be used, such as for

example, the KRAB (Kruppel-associated box) domain found in many naturally occurring KRAB proteins (Thiesen et al., 1991, Nucleic Acids Res., 19:3996).

In a further embodiment, the functional domain is a DNA modification domain such as a methyltransferase (or methylase) domain, a de-methylation domain, a deaminase domain, a hydroxylase domain, an acetylation domain, or a deacetylation domain. Many such domains are known in the art and any such domain can be used, depending on the desired function of the resultant chimeric protein. For example, it has been shown that a DNA methylation domain can be fused to a TALE repeat array DNA binding protein and used for targeted methylation of a specific DNA sequence (Xu et al., 1997, Nat. Genet., 17:376-378). The state of methylation of a gene affects its expression and regulation, and furthermore, there are several diseases associated with defects in DNA methylation.

In a still further embodiment the functional domain is a chromatin modification domain such as a histone acetylase or histone de-acetylase (or HDAC) domain. Many such domains are known in the art and any such domain can be used, depending on the desired function of the resultant chimeric protein. Histone deacetylases (such as HDAC1 and HDAC2) are involved in gene repression. Therefore, by targeting HDAC activity to a specific gene of interest using an engineered TALE protein, the expression of the gene of interest can be repressed.

In an alternative embodiment, the functional domain is a nuclease domain, such as a restriction endonuclease (or restriction enzyme) domain. The DNA cleavage activity of a nuclease enzyme can be targeted to a specific target sequence by fusing it to an appropriate engineered TALE repeat array DNA binding domain. In this way, sequence specific chimeric restriction enzyme can be produced. Several nuclease domains are known in the art and any suitable nuclease domain can be used. For example, an endonuclease domain of a type IIS restriction endonuclease (e.g., FokI) can be used, as taught by Kim et al., 1996, Proc. Natl. Acad. Sci. USA, 6:1156-60). In some embodiments, the endonuclease is an engineered FokI variant as described in US 2008/0131962. Such chimeric endonucleases can be used in any situation where cleavage of a specific DNA sequence is desired, such as in laboratory procedures for the

construction of recombinant DNA molecules, or in producing double-stranded DNA breaks in genomic DNA in order to promote homologous recombination (Kim et al., 1996, Proc. Natl. Acad. Sci. USA, 6:1156-60; Bibikova et al., 2001, Mol. Cell. Biol., 21:289-297; Porteus & Baltimore, 2003, Science, 300:763; Miller et al., 2011, Nat. Biotechnol., 29:143-148; Cermak et al., 2011, Nucl. Acids Res., 39:e82). Repair of TALE nuclease-induced double-strand breaks (DSB) by error-prone non-homologous end-joining leads to efficient introduction of insertion or deletion mutations at the site of the DSB (Miller et al., 2011, Nat. Biotechnol., 29:143-148; Cermak et al., 2011, Nucl. Acids Res., 39:e82). Alternatively, repair of a DSB by homology-directed repair with an exogenously introduced "donor template" can lead to highly efficient introduction of precise base alterations or insertions at the break site (Bibikova et al., 2003, Science, 300:764; Urnov et al., 2005, Nature, 435:646-651; Porteus et al., 2003, Science, 300:763; Miller et al., 2011, Nat. Biotechnol., 29:143-148).

In some embodiments, the functional domain is an integrase domain, such that the chimeric protein can be used to insert exogenous DNA at a specific location in, for example, the human genome.

Other suitable functional domains include silencer domains, nuclear hormone receptors, resolvase domains oncogene transcription factors (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos family members etc.), kinases, phosphatases, and any other proteins that modify the structure of DNA and/or the expression of genes. Suitable kinase domains, from kinases involved in transcription regulation are reviewed in Davis, 1995, Mol. Reprod. Dev., 42:459-67. Suitable phosphatase domains are reviewed in, for example, Schonthal & Semin, 1995, Cancer Biol. 6:239-48.

Fusions of TALE repeat arrays to functional domains can be performed by standard recombinant DNA techniques well known to those skilled in the art, and as are described in, for example, basic laboratory texts such as Sambrook et al., Molecular Cloning; A Laboratory Manual 3d ed. (2001), and in U.S. Pat. Nos. 6,511,808, 6,013,453, 6,007,988, and 6,503,717 and U.S. patent application 2002/0160940.

In some embodiments, two or more engineered TALE repeat array proteins are linked together to produce the final DNA binding domain. The linkage of two or more

engineered proteins can be performed by covalent or non-covalent means. In the case of covalent linkage, engineered proteins can be covalently linked together using an amino acid linker (see, for example, U.S. patent application 2002/0160940, and International applications WO 02/099084 and WO 01/53480). This linker can be any string of amino acids desired. In one embodiment the linker is a canonical TGEKP linker. Whatever linkers are used, standard recombinant DNA techniques (such as described in, for example, Sambrook et al., Molecular Cloning; A Laboratory Manual 3d ed. (2001)) can be used to produce such linked proteins.

In embodiments where the engineered proteins are used in the generation of chimeric endonuclease, the chimeric protein can possess a dimerization domain as such endonucleases are believed to function as dimers. Any suitable dimerization domain can be used. In one embodiment the endonuclease domain itself possesses dimerization activity. For example, the nuclease domain of FokI which has intrinsic dimerization activity can be used (Kim et al., 1996, Proc. Natl. Acad. Sci., 93:1156-60).

Assays for Determining Regulation of Gene Expression by Engineered Proteins

A variety of assays can be used to determine the level of gene expression regulation by the engineered TALE repeat proteins, see for example U.S. Pat. No. 6,453,242. The activity of a particular engineered TALE repeat protein can be assessed using a variety of in vitro and in vivo assays, by measuring, e.g., protein or mRNA levels, product levels, enzyme activity, tumor growth; transcriptional activation or repression of a reporter gene; second messenger levels (e.g., cGMP, cAMP, IP3, DAG, $Ca^{2+}$); cytokine and hormone production levels; and neovascularization, using, e.g., immunoassays (e.g., ELISA and immunohistochemical assays with antibodies), hybridization assays (e.g., RNase protection, northerns, in situ hybridization, oligonucleotide array studies), colorimetric assays, amplification assays, enzyme activity assays, tumor growth assays, phenotypic assays, and the like.

TALE proteins can be first tested for activity in vitro using cultured cells; e.g., 293 cells, CHO cells, VERO cells, BHK cells, HeLa cells, COS cells, and the like. In some embodiments, human cells are used. The engineered TALE repeat array protein is

often first tested using a transient expression system with a reporter gene, and then regulation of the target endogenous gene is tested in cells and in animals, both in vivo and ex vivo. The engineered TALE repeat array protein can be recombinantly expressed in a cell, recombinantly expressed in cells transplanted into an animal, or recombinantly expressed in a transgenic animal, as well as administered as a protein to an animal or cell using delivery vehicles described below. The cells can be immobilized, be in solution, be injected into an animal, or be naturally occurring in a transgenic or non-transgenic animal.

Modulation of gene expression is tested using one of the in vitro or in vivo assays described herein. Samples or assays are treated with the engineered TALE repeat array protein and compared to un-treated control samples, to examine the extent of modulation. For regulation of endogenous gene expression, the TALE repeat array protein ideally has a $K_D$ of 200 nM or less, more preferably 100 nM or less, more preferably 50 nM, most preferably 25 nM or less. The effects of the engineered TALE repeat array protein can be measured by examining any of the parameters described above. Any suitable gene expression, phenotypic, or physiological change can be used to assess the influence of the engineered TALE repeat array protein. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as tumor growth, neovascularization, hormone release, transcriptional changes to both known and uncharacterized genetic markers (e.g., northern blots or oligonucleotide array studies), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as cGMP.

Preferred assays for regulation of endogenous gene expression can be performed in vitro. In one in vitro assay format, the engineered TALE repeat array protein regulation of endogenous gene expression in cultured cells is measured by examining protein production using an ELISA assay. The test sample is compared to control cells treated with an empty vector or an unrelated TALE repeat array protein that is targeted to another gene.

In another embodiment, regulation of endogenous gene expression is determined in vitro by measuring the level of target gene mRNA expression. The level of gene

expression is measured using amplification, e.g., using RT-PCR, LCR, or hybridization assays, e.g., northern hybridization, RNase protection, dot blotting. RNase protection is used in one embodiment. The level of protein or mRNA is detected using directly or indirectly labeled detection agents, e.g., fluorescently or radioactively labeled nucleic acids, radioactively or enzymatically labeled antibodies, and the like, as described herein.

Alternatively, a reporter gene system can be devised using the target gene promoter operably linked to a reporter gene such as luciferase, green fluorescent protein, CAT, or beta-galactosidase. The reporter construct is typically co-transfected into a cultured cell. After treatment with the TALE repeat array protein, the amount of reporter gene transcription, translation, or activity is measured according to standard techniques known to those of skill in the art.

Another example of an assay format useful for monitoring regulation of endogenous gene expression is performed in vivo. This assay is particularly useful for examining TALE repeat array proteins that inhibit expression of tumor promoting genes, genes involved in tumor support, such as neovascularization (e.g., VEGF), or that activate tumor suppressor genes such as p53. In this assay, cultured tumor cells expressing the engineered TALE protein are injected subcutaneously into an immune compromised mouse such as an athymic mouse, an irradiated mouse, or a SCID mouse. After a suitable length of time, preferably 4-8 weeks, tumor growth is measured, e.g., by volume or by its two largest dimensions, and compared to the control. Tumors that have statistically significant reduction (using, e.g., Student's T test) are said to have inhibited growth. Alternatively, the extent of tumor neovascularization can also be measured. Immunoassays using endothelial cell specific antibodies are used to stain for vascularization of the tumor and the number of vessels in the tumor. Tumors that have a statistically significant reduction in the number of vessels (using, e.g., Student's T test) are said to have inhibited neovascularization.

Transgenic and non-transgenic animals can also be used for examining regulation of endogenous gene expression in vivo. Transgenic animals can express the engineered TALE repeat array protein. Alternatively, animals that transiently express the engineered TALE repeat array protein, or to which the engineered TALE repeat array protein has

been administered in a delivery vehicle, can be used. Regulation of endogenous gene expression is tested using any one of the assays described herein.

Use of Engineered TALE Repeat-Containing Proteins in Gene Therapy

The engineered proteins of the present invention can be used to regulate gene expression or alter gene sequence in gene therapy applications in the same. Similar methods have been described for synthetic zinc finger proteins, see for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,013,453, U.S. Pat. No. 6,007,988, U.S. Pat. No. 6,503,717, U.S. patent application 2002/0164575, and U.S. patent application 2002/0160940.

Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding the engineered TALE repeat array protein into mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding engineered TALE repeat array proteins to cells in vitro. Preferably, the nucleic acids encoding the engineered TALE repeat array proteins are administered for in vivo or ex vivo gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, 1992, Science, 256:808-813; Nabel & Felgner, 1993, TIBTECH, 11:211-217; Mitani & Caskey, 1993, TIBTECH, 11:162-166; Dillon, 1993, TIBTECH, 11:167-175; Miller, 1992, Nature, 357:455-460; Van Brunt, 1988, Biotechnology, 6:1149-54; Vigne, 1995, Restorat. Neurol. Neurosci., 8:35-36; Kremer & Perricaudet, 1995, Br. Med. Bull., 51:31-44; Haddada et al., in Current Topics in Microbiology and Immunology Doerfler and Bohm (eds) (1995); and Yu et al., 1994, Gene Ther., 1:13-26.

Methods of non-viral delivery of nucleic acids encoding the engineered TALE repeat array proteins include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA or RNA, artificial virions, and agent-enhanced uptake of DNA or RNA. Lipofection is described

in e.g., U.S. Pat. No. 5,049,386, No. 4,946,787; and No.4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ 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 (ex vivo administration) or target tissues (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, 1995, Science, 270:404-410; Blaese et al., 1995, Cancer Gene Ther., 2:291-297; Behr et al., 1994, Bioconjugate Chem. 5:382-389; Remy et al., 1994, Bioconjugate Chem., 5:647-654; Gao et al., Gene Ther., 2:710-722; Ahmad et al., 1992, Cancer Res., 52:4817-20; U.S. Pat. Nos. 4,186,183, 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 encoding the engineered TALE repeat array proteins 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 are administered to patients (ex vivo). Conventional viral based systems for the delivery of TALE repeat array proteins could include retroviral, lentivirus, adenoviral, adeno-associated, Sendai, and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene 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 (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., 1992, J. Virol., 66:2731-39; Johann et al., 1992, J. Virol., 66:1635-40; Sommerfelt et al., 1990, Virololgy, 176:58-59; Wilson et al., 1989, J. Virol., 63:2374-78; Miller et al., 1991, J. Virol., 65:2220-24; WO 94/26877).

In applications where transient expression of the engineered TALE repeat array protein is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., 1987, Virology 160:38-47; U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, 1994, Hum. Gene Ther., 5:793-801; Muzyczka, 1994, J. Clin. Invest., 94:1351). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., 1985, Mol. Cell. Biol. 5:3251-60; Tratschin et al.,1984, Mol. Cell. Biol., 4:2072-81; Hermonat & Muzyczka, 1984, Proc. Natl. Acad. Sci. USA, 81:6466-70; and Samulski et al., 1989, J. Virol., 63:3822-28.

In particular, at least six viral vector approaches are currently available for gene transfer in clinical trials, with retroviral vectors by far the most frequently used system. All of these viral vectors utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar et al., 1995, Blood, 85:3048; Kohn et al.,1995, Nat. Med., 1:1017; Malech et al., 1997, Proc. Natl. Acad. Sci. USA, 94:12133-38). PA317/pLASN was the

first therapeutic vector used in a gene therapy trial. (Blaese et al., 1995, Science, 270:475-480). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors (Ellem et al., 1997, Immunol Immunother., 44:10-20; Dranoffet al., 1997, Hum. Gene Ther., 1:111-112).

Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. Typically, the vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system (Wagner et al., 1998, Lancet, 351:1702-1703; Kearns et al., 1996, Gene Ther., 9:748-55).

Replication-deficient recombinant adenoviral vectors (Ad) are predominantly used for colon cancer gene therapy, because they can be produced at high titer and they readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiple types of tissues in vivo, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman et al., 1998, Hum. Gene Ther. 7:1083-89). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker et al., 1996, Infection, 24:15-10; Sterman et al., 1998, Hum. Gene Ther., 9:7 1083-89; Welsh et al., 1995, Hum. Gene Ther., 2:205-218; Alvarez et al., 1997, Hum. Gene Ther. 5:597-613; Topf et al., 1998, Gene Ther., 5:507-513; Sterman et al., 1998, Hum. Gene Ther., 7:1083-89.

Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and Ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors

typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han et al., 1995, Proc. Natl. Acad. Sci. USA, 92:9747-51, reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., Fab or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

Gene therapy vectors can be delivered in vivo by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described

below. Alternatively, vectors can be delivered to cells ex vivo, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or stem cells (e.g., universal donor hematopoietic stem cells, embryonic stem cells (ES), partially differentiated stem cells, non-pluripotent stem cells, pluripotent stem cells, induced pluripotent stem cells (iPS cells) (see e.g., Sipione et al., Diabetologia, 47:499-508, 2004)), followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

Ex vivo cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with nucleic acid (gene or cDNA), encoding the engineered TALE repeat array protein, and re-infused back into the subject organism (e.g., patient). Various cell types suitable for ex vivo transfection are well known to those of skill in the art (see, e.g., Freshney et al., Culture of Animal Cells, A Manual of Basic Technique (5th ed. 2005)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

In one embodiment, stem cells (e.g., universal donor hematopoietic stem cells, embryonic stem cells (ES), partially differentiated stem cells, non-pluripotent stem cells, pluripotent stem cells, induced pluripotent stem cells (iPS cells) (see e.g., Sipione et al., Diabetologia, 47:499-508, 2004)) are used in ex vivo procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types in vitro, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells in vitro into clinically important immune cell types using cytokines such a GM-CSF, IFN-gamma and TNF-alpha are known (see Inaba et al., 1992, J. Exp. Med., 176:1693-1702).

Stem cells can be isolated for transduction and differentiation using known methods. For example, stem cells can be isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+

(T cells), CD45+ (panB cells), GR-1 (granulocytes), and 1ad (differentiated antigen presenting cells) (see Inaba et al., 1992, J. Exp. Med., 176:1693-1702).

Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing nucleic acids encoding the engineered TALE repeat array protein can be also administered directly to the organism for transduction of cells in vivo. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route. Alternatively, stable formulations of the engineered TALE repeat array protein can also be administered.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions available, as described below (see, e.g., Remington: The Science and Practice of Pharmacy, 21st ed., 2005).

Delivery Vehicles

An important factor in the administration of polypeptide compounds, such as the engineered TALE repeat array proteins of the present invention, is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins and other compounds such as liposomes have been described, which have the ability to translocate polypeptides such as engineered TALE repeat array protein across a cell membrane.

For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-

translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, Antennapedia, was found to be the third helix of the protein, from amino acid position 43 to 58 (see, e.g., Prochiantz, 1996, Curr. Opin. Neurobiol., 6:629-634). Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics (see, e.g., Lin et al., 1995, J. Biol. Chem., 270:14255-58).

Examples of peptide sequences that can be linked to a protein, for facilitating uptake of the protein into cells, include, but are not limited to: peptide fragments of the tat protein of HIV (Endoh et al., 2010, Methods Mol. Biol., 623:271-281; Schmidt et al., 2010, FEBS Lett., 584:1806-13; Futaki, 2006, Biopolymers, 84:241-249); a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (see Fahraeus et al., 1996, Curr. Biol., 6:84); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi et al., 1994, J. Biol. Chem., 269:10444); the h region of a signal peptide, such as the Kaposi fibroblast growth factor (K-FGF) h region (Lin et al., supra); or the VP22 translocation domain from HSV (Elliot & O'Hare, 1997, Cell, 88:223-233). See also, e.g., Caron et al., 2001, Mol Ther., 3:310-318; Langel, Cell-Penetrating Peptides: Processes and Applications (CRC Press, Boca Raton FL 2002); El-Andaloussi et al., 2005, Curr. Pharm. Des., 11:3597-3611; and Deshayes et al., 2005, Cell. Mol. Life Sci., 62:1839-49. Other suitable chemical moieties that provide enhanced cellular uptake can also be chemically linked to TALE repeat array proteins described herein.

Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules are composed of at least two parts (called "binary toxins"): a translocation or binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular receptor, and then the toxin is transported into the cell. Several bacterial toxins, including Clostridium perfringens iota toxin, diphtheria toxin (DT), Pseudomonas exotoxin A (PE), pertussis toxin (PT), Bacillus anthracis toxin, and pertussis adenylate cyclase (CYA), have been used in attempts to deliver peptides to the cell cytosol as internal or amino-

terminal fusions (Arora et al., 1993, J. Biol. Chem., 268:3334-41; Perelle et al., 1993, Infect. Immun., 61:5147-56; Stenmark et al., 1991, J. Cell Biol., 113:1025-32; Donnelly et al., 1993, Proc. Natl. Acad. Sci. USA, 90:3530-34; Carbonetti et al., 1995, Abstr. Annu. Meet. Am. Soc. Microbiol. 95:295; Sebo et al., 1995, Infect. Immun., 63:3851-57; Klimpel et al., 1992, Proc. Natl. Acad. Sci. USA, 89:10277-81; and Novak et al., 1992, J. Biol. Chem., 267:17186-93).

Such subsequences can be used to translocate engineered TALE repeat array proteins across a cell membrane. The engineered TALE repeat array proteins can be conveniently fused to or derivatized with such sequences. Typically, the translocation sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the engineered TALE repeat array protein and the translocation sequence. Any suitable linker can be used, e.g., a peptide linker.

The engineered TALE repeat array protein can also be introduced into an animal cell, preferably a mammalian cell, via liposomes and liposome derivatives such as immunoliposomes. The term "liposome" refers to vesicles comprised of one or more concentrically ordered lipid bilayers, which encapsulate an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell, i.e., the engineered TALE repeat array protein.

The liposome fuses with the plasma membrane, thereby releasing the compound into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the liposome either degrades or fuses with the membrane of the transport vesicle and releases its contents.

In current methods of drug delivery via liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound (e.g., the engineered TALE repeat array protein or a nucleic acid encoding the same) at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer degrades over time through the action of various agents in the body. Alternatively, active compound release involves using an agent to induce a permeability change in the liposome vesicle. Liposome membranes can be constructed so that they become destabilized when the environment becomes acidic near

the liposome membrane (see, e.g., Proc. Natl. Acad. Sci. USA, 84:7851 (1987); Biochemistry, 28:908 (1989)). When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents. This destabilization is termed fusogenesis. Dioleoylphosphatidylethanolamine (DOPE) is the basis of many "fusogenic" systems.

Such liposomes typically comprise the engineered TALE repeat array protein and a lipid component, e.g., a neutral and/or cationic lipid, optionally including a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (e.g., an antigen). A variety of methods are available for preparing liposomes as described in, e.g., Szoka et al., 1980, Annu. Rev. Biophys. Bioeng., 9:467, U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,946,787, PCT Publication. No. WO 91/17424, Deamer & Bangham, 1976, Biochim. Biophys. Acta, 443:629-634; Fraley, et al., 1979, Proc. Natl. Acad. Sci. USA, 76:3348-52; Hope et al., 1985, Biochim. Biophys. Acta, 812:55-65; Mayer et al., 1986, Biochim. Biophys. Acta, 858:161-168; Williams et al., 1988, Proc. Natl. Acad. Sci. USA, 85:242-246; Liposomes (Ostro (ed.), 1983, Chapter 1); Hope et al., 1986, Chem. Phys. Lip., 40:89; Gregoriadis, Liposome Technology (1984) and Lasic, Liposomes: from Physics to Applications (1993)). Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are well known in the art.

In certain embodiments, it is desirable to target liposomes using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (e.g., ligands, receptors, and monoclonal antibodies) has been previously described (see, e.g., U.S. Pat. Nos. 4,957,773 and 4,603,044).

Examples of targeting moieties include monoclonal antibodies specific to antigens associated with neoplasms, such as prostate cancer specific antigen and MAGE. Tumors can also be diagnosed by detecting gene products resulting from the activation or over-expression of oncogenes, such as ras or c-erbB2. In addition, many tumors express antigens normally expressed by fetal tissue, such as the alphafetoprotein (AFP) and

carcinoembryonic antigen (CEA). Sites of viral infection can be diagnosed using various viral antigens such as hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV1) and papilloma virus antigens. Inflammation can be detected using molecules specifically recognized by surface molecules which are expressed at sites of inflammation such as integrins (e.g., VCAM-1), selectin receptors (e.g., ELAM-1) and the like.

Standard methods for coupling targeting agents to liposomes can be used. These methods generally involve incorporation into liposomes lipid components, e.g., phosphatidylethanolamine, which can be activated for attachment of targeting agents, or derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A (see Renneisen et al., 1990, J. Biol. Chem., 265:16337-42 and Leonetti et al., 1990, Proc. Natl. Acad. Sci. USA, 87:2448-51).

Dosages

For therapeutic applications, the dose of the engineered TALE repeat array protein to be administered to a patient is calculated in a similar way as has been described for zinc finger proteins, see for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,492,117, U.S. Pat. No. 6,453,242, U.S. patent application 2002/0164575, and U.S. patent application 2002/0160940. In the context of the present disclosure, the dose should be sufficient to effect a beneficial therapeutic response in the patient over time. In addition, particular dosage regimens can be useful for determining phenotypic changes in an experimental setting, e.g., in functional genomics studies, and in cell or animal models. The dose will be determined by the efficacy, specificity, and $K_D$ of the particular engineered TALE repeat array protein employed, the nuclear volume of the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular patient.

Pharmaceutical Compositions and Administration

Appropriate pharmaceutical compositions for administration of the engineered TALE repeat array proteins of the present invention can be determined as described for zinc finger proteins, see for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,492,117, U.S. Pat. No. 6,453,242, U.S. patent application 2002/0164575, and U.S. patent application 2002/0160940. Engineered TALE repeat array proteins, and expression vectors encoding engineered TALE repeat array proteins, can be administered directly to the patient for modulation of gene expression and for therapeutic or prophylactic applications, for example, cancer, ischemia, diabetic retinopathy, macular degeneration, rheumatoid arthritis, psoriasis, HIV infection, sickle cell anemia, Alzheimer's disease, muscular dystrophy, neurodegenerative diseases, vascular disease, cystic fibrosis, stroke, and the like. Examples of microorganisms that can be inhibited by TALE repeat array protein-mediated gene therapy include pathogenic bacteria, e.g., chlamydia, rickettsial bacteria, mycobacteria, staphylococci, streptococci, pneumococci, meningococci and conococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospirosis, and Lyme disease bacteria; infectious fungus, e.g., Aspergillus, Candida species; protozoa such as sporozoa (e.g., Plasmodia), rhizopods (e.g., Entamoeba) and flagellates (Trypanosoma, Leishmania, Trichomonas, Giardia, etc.); viral diseases, e.g., hepatitis (A, B, or C), herpes virus (e.g., VZV, HSV-1, HSV-6, HSV-II, CMV, and EBV), HIV, Ebola, adenovirus, influenza virus, flaviviruses, echovirus, rhinovirus, coxsackie virus, comovirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV virus, dengue virus, papillomavirus, poliovirus, rabies virus, and arboviral encephalitis virus, etc.

Administration of therapeutically effective amounts is by any of the routes normally used for introducing TALE repeat array proteins into ultimate contact with the tissue to be treated. The TALE repeat array proteins are administered in any suitable manner, preferably with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a

particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions that are available (see, e.g., Remington: The Science and Practice of Pharmacy, 21st ed., 2005).

The engineered TALE repeat array proteins, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. The disclosed compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

Use of TALE Nucleases

TALE nucleases engineered using the methods described herein can be used to induce mutations in a genomic sequence, e.g., by cleaving at two sites and deleting sequences in between, by cleavage at a single site followed by non-homologous end joining, and/or by cleaving at a site so as to remove or replace one or two or a few nucleotides. In some embodiments, the TALE nuclease is used to induce mutation in an

animal, plant, fungal, or bacterial genome. Targeted cleavage can also be used to create gene knock-outs (e.g., for functional genomics or target validation) and to facilitate targeted insertion of a sequence into a genome (i.e., gene knock-in); e.g., for purposes of cell engineering or protein overexpression. Insertion can be by means of replacements of chromosomal sequences through homologous recombination or by targeted integration, in which a new sequence (i.e., a sequence not present in the region of interest), flanked by sequences homologous to the region of interest in the chromosome, is used to insert the new sequence at a predetermined target site via homologous recombination. Exogenous DNA can also be inserted into TALE nuclease-induced double stranded breaks without the need for flanking homology sequences (see, Orlando et al., 2010, Nucl. Acids Res., 1-15, doi:10.1093/nar/gkq512).

As demonstrated in Example 3 below, the TALE nucleases produced by the methods described herein were capable of inducing site-specific mutagenesis in mammalian cells. A skilled practitioner will readily appreciate that TALE nucleases produced by the methods described herein woluld also function to induce efficient site-specific mutagenesis in other cell types and organisms (see, for example, Cade et al., 2012, Nucleic Acids Res., PMID: 22684503 and Moore et al., 2012, PLoS One, PMID: 22655075).

The same methods can also be used to replace a wild-type sequence with a mutant sequence, or to convert one allele to a different allele.

Targeted cleavage of infecting or integrated viral genomes can be used to treat viral infections in a host. Additionally, targeted cleavage of genes encoding receptors for viruses can be used to block expression of such receptors, thereby preventing viral infection and/or viral spread in a host organism. Targeted mutagenesis of genes encoding viral receptors (e.g., the CCR5 and CXCR4 receptors for HIV) can be used to render the receptors unable to bind to virus, thereby preventing new infection and blocking the spread of existing infections. Non-limiting examples of viruses or viral receptors that can be targeted include herpes simplex virus (HSV), such as HSV-1 and HSV-2, varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV), HHV6 and HHV7. The hepatitis family of viruses includes hepatitis A virus (HAV), hepatitis B virus

(HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV) and hepatitis G virus (HGV). Other viruses or their receptors can be targeted, including, but not limited to, Picornaviridae (e.g., polioviruses, etc.); Caliciviridae; Togaviridae (e.g., rubella virus, dengue virus, etc.); Flaviviridae; Coronaviridae; Reoviridae; Bimaviridae; Rhabodoviridae (e.g., rabies virus, etc.); Filoviridae; Paramyxoviridae (e.g., mumps virus, measles virus, respiratory syncytial virus, etc.); Orthomyxoviridae (e.g., influenza virus types A, B and C, etc.); Bunyaviridae; Arenaviridae; Retroviradae; lentiviruses (e.g., HTLV-I; HTLV-II; HIV-1 (also known as HTLV-III, LAV, ARV, hTLR, etc.) HIV-II); simian immunodeficiency virus (SIV), human papillomavirus (HPV), influenza virus and the tick-borne encephalitis viruses. See, e.g., Virology, 3rd Edition (W. K. Joklik, ed. 1988); Fundamental Virology, 4th Edition (Knipe and Howley, eds. 2001), for a description of these and other viruses. Receptors for HIV, for example, include CCR-5 and CXCR-4.

In similar fashion, the genome of an infecting bacterium can be mutagenized by targeted DNA cleavage followed by non-homologous end joining, to block or ameliorate bacterial infections.

The disclosed methods for targeted recombination can be used to replace any genomic sequence with a homologous, non-identical sequence. For example, a mutant genomic sequence can be replaced by its wild-type counterpart, thereby providing methods for treatment of e.g., genetic disease, inherited disorders, cancer, and autoimmune disease. In like fashion, one allele of a gene can be replaced by a different allele using the methods of targeted recombination disclosed herein.

Exemplary genetic diseases include, but are not limited to, achondroplasia, achromatopsia, acid maltase deficiency, adenosine deaminase deficiency (OMIM No.102700), adrenoleukodystrophy, aicardi syndrome, alpha-1 antitrypsin deficiency, alpha-thalassemia, androgen insensitivity syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia telangictasia, barth syndrome, beta-thalassemia, blue rubber bleb nevus syndrome, canavan disease, chronic granulomatous diseases (CGD), cri du chat syndrome, cystic fibrosis, dercum's disease, ectodermal dysplasia, Fanconi anemia, fibrodysplasia ossificans progressive, fragile X syndrome, galactosemis,

Gaucher's disease, generalized gangliosidoses (e.g., GM1), hemochromatosis, the hemoglobin C mutation in the 6th codon of beta-globin (HbC), hemophilia, Huntington's disease, Hurler Syndrome, hypophosphatasia, Klinefelter's syndrome, Krabbes Disease, Langer-Giedion Syndrome, leukocyte adhesion deficiency (LAD, OMIM No. 116920), leukodystrophy, long QT syndrome, Marfan syndrome, Moebius syndrome, mucopolysaccharidosis (MPS), nail patella syndrome, nephrogenic diabetes insipdius, neurofibromatosis, Neimann-Pick disease, osteogenesis imperfecta, porphyria, Prader-Willi syndrome, progeria, Proteus syndrome, retinoblastoma, Rett syndrome, Rubinstein-Taybi syndrome, Sanfilippo syndrome, severe combined immunodeficiency (SCID), Shwachman syndrome, sickle cell disease (sickle cell anemia), Smith-Magenis syndrome, Stickler syndrome, Tay-Sachs disease, Thrombocytopenia Absent Radius (TAR) syndrome, Treacher Collins syndrome, trisomy, tuberous sclerosis, Turner's syndrome, urea cycle disorder, von Hippel-Landau disease, Waardenburg syndrome, Williams syndrome, Wilson's disease, Wiskott-Aldrich syndrome, X-linked lymphoproliferative syndrome (XLP, OMIM No. 308240).

Additional exemplary diseases that can be treated by targeted DNA cleavage and/or homologous recombination include acquired immunodeficiencies, lysosomal storage diseases (e.g., Gaucher's disease, GM1, Fabry disease and Tay-Sachs disease), mucopolysaccahidosis (e.g., Hunter's disease, Hurler's disease), hemoglobinopathies (e.g., sickle cell diseases, HbC, alpha-thalassemia, beta-thalassemia) and hemophilias.

In certain cases, alteration of a genomic sequence in a pluripotent cell (e.g., a hematopoietic stem cell) is desired. Methods for mobilization, enrichment and culture of hematopoietic stem cells are known in the art. See for example, U.S. Pat. Nos. 5,061,620; 5,681,559; 6,335,195; 6,645,489 and 6,667,064. Treated stem cells can be returned to a patient for treatment of various diseases including, but not limited to, SCID and sickle-cell anemia.

In many of these cases, a region of interest comprises a mutation, and the donor polynucleotide comprises the corresponding wild-type sequence. Similarly, a wild-type genomic sequence can be replaced by a mutant sequence, if such is desirable. For example, overexpression of an oncogene can be reversed either by mutating the gene or

by replacing its control sequences with sequences that support a lower, non-pathologic level of expression. As another example, the wild-type allele of the ApoAI gene can be replaced by the ApoAI Milano allele, to treat atherosclerosis. Indeed, any pathology dependent upon a particular genomic sequence, in any fashion, can be corrected or alleviated using the methods and compositions disclosed herein.

Targeted cleavage and targeted recombination can also be used to alter non-coding sequences (e.g., sequences encoding microRNAs and long non-coding RNAs, and regulatory sequences such as promoters, enhancers, initiators, terminators, splice sites) to alter the levels of expression of a gene product. Such methods can be used, for example, for therapeutic purposes, functional genomics and/or target validation studies.

The compositions and methods described herein also allow for novel approaches and systems to address immune reactions of a host to allogeneic grafts. In particular, a major problem faced when allogeneic stem cells (or any type of allogeneic cell) are grafted into a host recipient is the high risk of rejection by the host's immune system, primarily mediated through recognition of the Major Histocompatibility Complex (MHC) on the surface of the engrafted cells. The MHC comprises the HLA class I protein(s) that function as heterodimers that are comprised of a common beta subunit and variable alpha subunits. It has been demonstrated that tissue grafts derived from stem cells that are devoid of HLA escape the host's immune response. See, e.g., Coffman et al., 1993, J. Immunol., 151:425-35; Markmann et al., 1992, Transplantation, 54:1085-89; Koller et al., 1990, Science, 248:1227-30. Using the compositions and methods described herein, genes encoding HLA proteins involved in graft rejection can be cleaved, mutagenized or altered by recombination, in either their coding or regulatory sequences, so that their expression is blocked or they express a non-functional product. For example, by inactivating the gene encoding the common beta subunit gene (beta2 microglobulin) using TALE nuclease fusion proteins as described herein, HLA class I can be removed from the cells to rapidly and reliably generate HLA class I null stem cells from any donor, thereby reducing the need for closely matched donor/recipient MHC haplotypes during stem cell grafting.

Inactivation of any gene (e.g., the beta2 microglobulin gene) can be achieved, for example, by a single cleavage event, by cleavage followed by non-homologous end joining, by cleavage at two sites followed by joining so as to delete the sequence between the two cleavage sites, by targeted recombination of a missense or nonsense codon into the coding region, or by targeted recombination of an irrelevant sequence (i.e., a "stuffer" sequence) into the gene or its regulatory region, so as to disrupt the gene or regulatory region.

Targeted modification of chromatin structure, as disclosed in WO 01/83793, can be used to facilitate the binding of fusion proteins to cellular chromatin.

In additional embodiments, one or more fusions between a TALE binding domain and a recombinase (or functional fragment thereof) can be used, in addition to or instead of the TALE-cleavage domain fusions disclosed herein, to facilitate targeted recombination. See, for example, co-owned U.S. Pat. No. 6,534,261 and Akopian et al. (2003) Proc. Natl. Acad. Sci. USA 100:8688-8691.

In additional embodiments, the disclosed methods and compositions are used to provide fusions of TALE repeat DNA-binding domains with transcriptional activation or repression domains that require dimerization (either homodimerization or heterodimerization) for their activity. In these cases, a fusion polypeptide comprises a TALE repeat DNA-binding domain and a functional domain monomer (e.g., a monomer from a dimeric transcriptional activation or repression domain). Binding of two such fusion polypeptides to properly situated target sites allows dimerization so as to reconstitute a functional transcription activation or repression domain.

Regulation of Gene Expression in Plants

Engineered TALE repeat array proteins can be used to engineer plants for traits such as increased disease resistance, modification of structural and storage polysaccharides, flavors, proteins, and fatty acids, fruit ripening, yield, color, nutritional characteristics, improved storage capability, and the like. In particular, the engineering of crop species for enhanced oil production, e.g., the modification of the fatty acids produced in oilseeds, is of interest.

Seed oils are composed primarily of triacylglycerols (TAGs), which are glycerol esters of fatty acids. Commercial production of these vegetable oils is accounted for primarily by six major oil crops (soybean, oil palm, rapeseed, sunflower, cotton seed, and peanut). Vegetable oils are used predominantly (90%) for human consumption as margarine, shortening, salad oils, and frying oil. The remaining 10% is used for non-food applications such as lubricants, oleochemicals, biofuels, detergents, and other industrial applications.

The desired characteristics of the oil used in each of these applications varies widely, particularly in terms of the chain length and number of double bonds present in the fatty acids making up the TAGs. These properties are manipulated by the plant in order to control membrane fluidity and temperature sensitivity. The same properties can be controlled using TALE repeat array proteins to produce oils with improved characteristics for food and industrial uses.

The primary fatty acids in the TAGs of oilseed crops are 16 to 18 carbons in length and contain 0 to 3 double bonds. Palmitic acid (16:0 [16 carbons: 0 double bonds]), oleic acid (18:1), linoleic acid (18:2), and linolenic acid (18:3) predominate. The number of double bonds, or degree of saturation, determines the melting temperature, reactivity, cooking performance, and health attributes of the resulting oil.

The enzyme responsible for the conversion of oleic acid (18:1) into linoleic acid (18:2) (which is then the precursor for 18:3 formation) is delta-12-oleate desaturase, also referred to as omega-6 desaturase. A block at this step in the fatty acid desaturation pathway should result in the accumulation of oleic acid at the expense of polyunsaturates.

In one embodiment engineered TALE repeat array proteins are used to regulate expression of the FAD2-1 gene in soybeans. Two genes encoding microsomal delta-6 desaturases have been cloned recently from soybean, and are referred to as FAD2-1 and FAD2-2 (Heppard et al., 1996, Plant Physiol. 110:311-319). FAD2-1 (delta-12 desaturase) appears to control the bulk of oleic acid desaturation in the soybean seed. Engineered TALE repeat array proteins can thus be used to modulate gene expression of FAD2-1 in plants. Specifically, engineered TALE repeat array proteins can be used to inhibit expression of the FAD2-1 gene in soybean in order to increase the accumulation

of oleic acid (18:1) in the oil seed. Moreover, engineered TALE proteins can be used to modulate expression of any other plant gene, such as delta-9 desaturase, delta-12 desaturases from other plants, delta-15 desaturase, acetyl-CoA carboxylase, acyl-ACP-thioesterase, ADP-glucose pyrophosphorylase, starch synthase, cellulose synthase, sucrose synthase, senescence-associated genes, heavy metal chelators, fatty acid hydroperoxide lyase, polygalacturonase, EPSP synthase, plant viral genes, plant fungal pathogen genes, and plant bacterial pathogen genes.

Recombinant DNA vectors suitable for transformation of plant cells are also used to deliver protein (e.g., engineered TALE repeat array protein)-encoding nucleic acids to plant cells. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature (see, e.g., Weising et al., 1988, Ann. Rev. Genet., 22:421-477). A DNA sequence coding for the desired TALE repeat array protein is combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the TALE protein in the intended tissues of the transformed plant.

For example, a plant promoter fragment can be employed which will direct expression of the engineered TALE repeat array protein in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35 S transcription initiation region, the 1'- or 2'-promoter derived from T-DNA of Agrobacterium tumefaciens, and other transcription initiation regions from various plant genes known to those of skill.

Alternatively, the plant promoter can direct expression of the engineered TALE repeat array protein in a specific tissue or can be otherwise under more precise environmental or developmental control. Such promoters are referred to here as "inducible" promoters. Examples of environmental conditions that can affect transcription by inducible promoters include anaerobic conditions or the presence of light.

Examples of promoters under developmental control include promoters that initiate transcription only in certain tissues, such as fruit, seeds, or flowers. For example, the use of a polygalacturonase promoter can direct expression of the TALE repeat array protein in the fruit, a CHS-A (chalcone synthase A from petunia) promoter can direct expression of the TALE repeat array protein in the flower of a plant.

The vector comprising the TALE repeat array protein sequences will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker can encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

Such DNA constructs can be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct can be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs can be combined with suitable T-DNA flanking regions and introduced into a conventional Agrobacterium tumefaciens host vector. The virulence functions of the Agrobacterium tumefaciens host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al., 1984, EMBO J., 3:2717-22. Electroporation techniques are described in Fromm et al. 1985, Proc. Natl. Acad. Sci. USA, 82:5824. Biolistic transformation techniques are described in Klein et al., 1987, Nature, 327:70-73.

Agrobacterium tumefaciens-meditated transformation techniques are well described in the scientific literature (see, e.g., Horsch et al., 1984, Science, 233:496-498; and Fraley et al., 1983, Proc. Natl. Acad. Sci. USA, 80:4803).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired TALE repeat array protein-controlled phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the TALE repeat array protein nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp. 124-176 (1983); and Binding, Regeneration of Plants, Plant Protoplasts, pp. 21-73 (1985). Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al., 1987, Ann. Rev. Plant Phys., 38:467-486.

Functional Genomics Assays

Engineered TALE repeat array proteins also have use for assays to determine the phenotypic consequences and function of gene expression. Recent advances in analytical techniques, coupled with focused mass sequencing efforts have created the opportunity to identify and characterize many more molecular targets than were previously available. This new information about genes and their functions will improve basic biological understanding and present many new targets for therapeutic intervention. In some cases analytical tools have not kept pace with the generation of new data. An example is provided by recent advances in the measurement of global differential gene expression. These methods, typified by gene expression microarrays, differential cDNA cloning frequencies, subtractive hybridization and differential display methods, can very rapidly identify genes that are up or down-regulated in different tissues or in response to specific stimuli. Increasingly, such methods are being used to explore biological processes such as, transformation, tumor progression, the inflammatory response, neurological disorders etc. Many differentially expressed genes correlate with a given physiological phenomenon, but demonstrating a causative relationship between an individual differentially expressed gene and the phenomenon is labor intensive. Until now, simple

methods for assigning function to differentially expressed genes have not kept pace with the ability to monitor differential gene expression.

The engineered TALE repeat array proteins described herein can be used to rapidly analyze the function of a differentially expressed gene. Engineered TALE proteins can be readily used to up or down-regulate or knockout any endogenous target gene, or to knock in an endogenous or endogenous gene. Very little sequence information is required to create a gene-specific DNA binding domain. This makes the engineered TALE repeat array technology ideal for analysis of long lists of poorly characterized differentially expressed genes. One can simply build a TALE repeat array protein-based DNA binding domain for each candidate gene, create chimeric up and down-regulating artificial transcription factors and test the consequence of up or down-regulation on the phenotype under study (e.g., transformation or response to a cytokine) by switching the candidate genes on or off one at a time in a model system.

Additionally, greater experimental control can be imparted by engineered TALE repeat array proteins than can be achieved by more conventional methods. This is because the production and/or function of engineered TALE repeat array proteins can be placed under small molecule control. Examples of this approach are provided by the Tet-On system, the ecdysone-regulated system and a system incorporating a chimeric factor including a mutant progesterone receptor. These systems are all capable of indirectly imparting small molecule control on any endogenous gene of interest or any transgene by placing the function and/or expression of a engineered TALE repeat array protein under small molecule control.

Transgenic Animals

A further application of engineered TALE repeat array proteins is manipulating gene expression in animal models. As with cell lines, the introduction of a heterologous gene into or knockout of an endogenous in a transgenic animal, such as a transgenic mouse or zebrafish, is a fairly straightforward process. Thus, transgenic or transient expression of an engineered TALE repeat array protein in an animal can be readily performed.

By transgenically or transiently expressing a suitable engineered TALE repeat array protein fused to an activation domain, a target gene of interest can be over-expressed. Similarly, by transgenically or transiently expressing a suitable engineered TALE repeat array protein fused to a repressor or silencer domain, the expression of a target gene of interest can be down-regulated, or even switched off to create "functional knockout". Knock-in or knockout mutations by insertion or deletion of a target gene of interest can be prepared using TALE nucleases.

Two common issues often prevent the successful application of the standard transgenic and knockout technology; embryonic lethality and developmental compensation. Embryonic lethality results when the gene plays an essential role in development. Developmental compensation is the substitution of a related gene product for the gene product being knocked out, and often results in a lack of a phenotype in a knockout mouse when the ablation of that gene's function would otherwise cause a physiological change.

Expression of transgenic engineered TALE repeat array proteins can be temporally controlled, for example using small molecule regulated systems as described in the previous section. Thus, by switching on expression of an engineered TALE repeat array protein at a desired stage in development, a gene can be over-expressed or "functionally knocked-out" in the adult (or at a late stage in development), thus avoiding the problems of embryonic lethality and developmental compensation.

## EXAMPLES

Example 1. Assembly of TALE Repeat Arrays using streptavidin coated magnetic beads

An archive of DNA plasmids (~850 different plasmids) encoding one, two, three, or four TALE repeat domains was created for assembly of nucleic acids encoding multiple TALE arrays of any desired length. The plasmids were created by cloning synthetic arrays of one, two, three or four TALE repeat domains into the pUC57-ΔBsal backbone (FIG. 3). The TALE repeats were of the arrangement α, βγδε, βγδ, βγ', βγ, δε', and β, and included hypervariable triplet residues at each position to bind to the nucleotides as shown in Table 1. Polypeptide and nucleotide sequences of the TALE

repeat types are shown in FIGs. 4A and 4B, respectively. The polypeptide and polynucleotide sequences were varied slightly among the four types to reduce the possibility of recombination-mediated mutations due to long sequences of exact repeats.

**Table 1. Nucleotide binding code of TALE triplets**

| Triplet | Bound Nucleotide |
|---|---|
| SNI | A |
| SHD | C |
| NNN | G |
| SNK | G |
| SNG | T |

A 16-mer TALE repeat array targeted to the eGFP gene was created by in vitro assembly of 16 TALE repeats designed to bind the target sequence GCAGTGCTTCAGCCGC (SEQ ID NO: 41). In the first step, a plasmid carrying an α-type TALE repeat with an NNN triplet (G) was amplified by PCR using a biotinylated forward primer Biotin-TCTAGAGAAGACAAGAACCTGACC (SEQ ID NO: 42) and a reverse primer GGATCCGGTCTCTTAAGGCCGTGG (SEQ ID NO: 43). The amplified fragment (50 µl) was purified using a QIA Quick PCR purification kit (QIAGEN), eluted in 40 µl 0.1X elution buffer (as provided in the QIA Quick PCR purification kit), and digested with BsaI HF (New England Biolabs (NEB)) in NEB Buffer 4 for 15 minutes at 50 °C (40 µl elution, 5 µl NEBuffer 4, 5 µl BsaI HF). The digested fragment was purified using a QIA Quick PCR purification kit and eluted in 0.1X elution buffer (50 µl).

A plasmid containing a four TALE repeat domain sub-array unit (βγδε) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SNG (designed to bind the sequence 5'-CAGT-3') was digested with BbsI (NEB) in NEBuffer 2 for 2 hours at 37 °C in 100 µl (50 µl plasmid [~200 ng/µl], 10 µl NEBuffer 2, 10 µl BbsI, 30 µl water). To the 100 µl digest was added 25 µl NEBuffer 4, 2.5 µl 100X BSA (NEB), 107.5 µl water, and 5 µl XbaI (NEB), and the digest was incubated for 5

minutes at 37 °C. To the mixture, 5 µl of BamHI HF was then added for a 5 minute digest at 37 °C, and then 5 µl SalI HF (NEB) was added for an additional 5 minute digest at 37 °C. The resulting fragment was purified using a QIA Quick PCR purification kit (QIAGEN) and eluted in 180 µl 0.1X elution buffer.

For the initial ligation, 2 µl of the alpha unit digest was mixed with 2.5 µl of T4 DNA ligase (400 U/µl; NEB) and 27 µl Quick Ligase Buffer (QLB) (NEB). To this 31.5 µl mixture was added 22.5 µl of the first digested subarray, and the mixture was ligated for 15 minutes at room temperature. Magnetic beads were prepared by washing 5 µl of Dynabeads MyOne Streptavidin C1 (Invitrogen) three times with 50 µl 1x B&W Buffer (5.0 mM Tris-HCl [pH 7.5], 0.5 mM EDTA, 1.0 M NaCl, 0.005% Tween™ 20) and resuspending in 54 µl B&W Buffer. The ligated mixture was added to the washed beads and incubated for 15 minutes at room temperature (with mixing every five minutes). The mixture was then placed on a SPRIplate 96-well Ring magnet for 3 minutes. The supernatant was then aspirated, and 100 µl 1x B&W Buffer was added to wash, with mixing by moving the beads 31 times from side to side within the tube using a DynaMag-96 Side magnet (Invitrogen). The B&W Buffer was then aspirated, and 100 µl 1x BSA was added, with mixing, then aspirated. The ligated, bead-bound nucleic acids (αβγδε) were resuspended in 50 µl BsaI HF mix (5 µl NEBuffer 4, 2 µl BsaI HF, 43 µl water).

The digest was incubated at 50 °C for 10 minutes, and 50 µl 1X B&W buffer was added. The digest was placed on a magnet for 3 minutes, and the supernatant was aspirated. The beads were washed with 100 µl 1X B&W Buffer and 100 µl 1X BSA as above. To the washed beads were added a digested plasmid containing a four TALE repeat domain sub-array unit (βγδε) coding for repeats that each harbor one of the following variable amino acids NNN, SHD, SNG, and SNG (designed to bind the DNA sequence 5'-GCTT-3') (22.5 µl) and 27.5 µl ligase mix (25 µl Quick Ligase Buffer, 2 µl DNA ligase). The beads were resuspended by pipetting up and down, and the mixture was incubated for 15 minutes at room temperature with mixing every five minutes. To the ligation was added 50 µl 1X B&W Buffer, and the mixture was placed on the magnet for 3 minutes. The supernatant was aspirated, and the beads were washed with 100 µl 1X B&W Buffer and 100 µl 1X BSA as above. The ligated, bead-bound nucleic acids

(αβγδεβγδε) were resuspended in 50 µl BsaI HF mix (5 µl NEBuffer 4, 2 µl BsaI HF, 43 µl water). Two more TALE repeat sub-array units were ligated sequentially as above, the first a four TALE repeat sub-array unit (βγδε) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SHD (designed to bind the DNA sequence 5'-CAGC-3') and the second a three TALE repeat sub-array unit (βγδ) coding for repeats that each harbor one of the following variable amino acids SHD, NNN, and SHD (designed to bind the DNA sequence 5'-CGC-3'). The final TALE repeat array contained subunits of the format αβγδεβγδεβγδεβγδ with individual TALE repeats designed to bind the target DNA sequence 5'-GCAGTGCTTCAGCCGC-3' (SEQ ID NO: 44).

Following the final ligation step, the construct was digested with BsaI HF for eventual cloning into an expression vector and the beads were washed with 1X B&W Buffer and 1X BSA. The washed beads were resuspended in 50 µl BbsI mix (5 µl NEBuffer 2, 5 µl BbsI, 40 µl water) and incubated at 37 °C for 2 hours with agitation at 1500 rpm to cleave the biotinylated 5' end and release the assembled TALE repeat array from the magnetic beads. The digested mixture was purified by MinElute column purified (QIAGEN) and ligated into a BsmBI-digested TALE expression vector. The ligated mixture was transformed into chemically competent XL1 Blue cells and plated on LB/Carb$^{100}$ plates overnight.

The expression vectors each harbor the following elements: a T7 promoter, a nuclear localization signal, a FLAG tag, amino acids 153 to 288 from the TALE13 protein (numbering as defined by Miller et al., 2011, Nat. Biotechnol., 29:143-148), two adjacent BsmBI restriction sites into which a DNA fragment encoding a TALE repeat array can be cloned, a 0.5 TALE repeat, amino acids 715 to 777 from the C-terminal end of the TALE13 protein (numbering as defined by Miller et al., 2011, Nat. Biotechnol., 29:143-148), and the wild-type FokI cleavage domain.

The plasmids differ in the identity of the C-terminal 0.5 TALE repeat. Plasmid pJDS70 encodes a 0.5 TALE repeat with a SNI RVD (for recognition of an A nucleotide), plasmid pJDS71 encodes a 0.5 TALE repeat with a SHD RVD (for recognition of a C nucleotide), plasmid pJDS74 encodes a 0.5 TALE repeat with a NNN RVD (for

recognition of a G nucleotide), plasmid pJDS76 encodes a 0.5 TALE repeat with a SNK RVD (for recognition of a G nucleotide), and plasmid pJDS78 encodes a 0.5 TALE repeat with a NG RVD (for recognition of a T nucleotide). All plasmids share the common sequence shown in FIGs. 5A-5B and differ at just nine nucleotide positions marked as XXXXXXXXX (underlined and bold). The sequence of these 9 bps and plasmid names are also shown below in Table 2.

**Table 2. DNA sequences of expression vectors**

| Plasmid name | Sequence of variable 9 bps | SEQ ID NO: | RVD of C-terminal 0.5 TALE repeat |
|---|---|---|---|
| pJDS70 | TCTAACATC | 45 | SNI (for binding to an A nucleotide) |
| pJDS71 | TCCCACGAC | 46 | SHD (for binding to a C nucleotide) |
| pJDS74 | AATAATAAC | 47 | NNN (for binding to a G nucleotide) |
| pJDS76 | TCCAATAAA | 48 | SNK (for binding to a G nucleotide) |
| pJDS78 | TCTAATGGG | 49 | SNG (for binding to a T nucleotide) |

This example demonstrates the construction of TALE repeat arrays on an immobilized substrate using preassembled TALE repeat sub-array units. The above method, up to the cloning step, can be performed in one day.

Example 2. Assembly of TALE Repeat Arrays using a streptavidin coated plate

TALE repeats are assembled using the archive of DNA plasmids (~850 different plasmids) as described in Example 1. A 16-mer TALE repeat array was created by in vitro assembly of 16 TALE repeats designed to bind a target sequence. In the first step, a plasmid carrying an α-type TALE repeat with an NNN triplet (G) was amplified by PCR using a biotinylated forward primer Biotin-TCTAGAGAAGACAAGAACCTGACC (SEQ ID NO: 42) and a reverse primer GGATCCGGTCTCTTAAGGCCGTGG (SEQ ID NO: 43). The amplified fragment (50 µl) was purified using a QIA Quick PCR purification kit (QIAGEN), eluted in 40 µl 0.1X elution buffer (as provided in the QIA

Quick PCR purification kit), and digested with BsaI HF (New England Biolabs (NEB)) in NEB Buffer 4 for 15 minutes at 50 °C (40 µl elution; 5 µl NEBuffer 4, 5 µl BsaI HF). The digested fragment was purified using a QIA Quick PCR purification kit and eluted in 0.1X elution buffer (50 µl).

A plasmid containing a four TALE repeat domain sub-array unit (βγδε) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SNG (designed to bind the sequence 5'-CAGT-3') was digested with BbsI (NEB) in NEBuffer 2 for 2 hours at 37 °C in 100 µl (50 µl plasmid [~200 ng/µl], 10 µl NEBuffer 2, 10 µl BbsI, 30 µl water). To the 100 µl digest was added 25 µl NEBuffer 4, 2.5 µl 100X BSA (NEB), 107.5 µl water, and 5 µl XbaI (NEB), and the digest was incubated for 5 minutes at 37 °C. To the mixture, 5 µl of BamHI HF was then added for a 5 minute digest at 37 °C, and then 5 µl SalI HF (NEB) was added for an additional 5 minute digest at 37 °C. The resulting fragment was purified using a QIA Quick PCR purification kit (QIAGEN) and eluted in 180 µl 0.1X elution buffer.

For the initial ligation, 2 µl of the alpha unit digest was mixed with 2.5 µl of T4 DNA ligase (400 U/µl; NEB) and 27 µl Quick Ligase Buffer (QLB) (NEB). To this 31.5 µl mixture was added 22.5 µl of the first digested subarray, and the mixture was ligated for 15 minutes at room temperature. The ligation mixture was then mixed with 2X B&@ buffer (Invitrogen) and added to a well in a 96-well plate coated with streptavidin (Thermo Scientific) and incubated at room temperature for 15min. The supernatant was aspirated. Each well in the 96 well plate was washed with 200ul of 1x Bovine Serum Albumin (BSA) by pipetting up and down 10 times before discarding the 1x BSA. This was repeated for a total of two washes with 1x BSA. Then 50 µl BsaI HF mix (5 µl NEBuffer 4, 2 µl BsaI HF, 43 µl water) was added to the ligated,nucleic acids (αβγδε) bound to the streptavidin-coated well.

The digest was incubated at 50 °C for 10 minutes and then the supernatant was aspirated. The wells were then washed with 200 µl 1X B&W Buffer and 200 µl 1X BSA twice by pipetting up and down ten times before removal of each supernatant. 22.5 µl of digested plasmid encoding a four TALE repeat domain sub-array unit (βγδε) coding for repeats that each harbor one of the following variable amino acids NNN, SHD, SNG, and

SNI and 27.5 µl ligase mix (25 µl Quick Ligase Buffer, 2 µl DNA ligase) were added to the well. The supernatant was mixed by pipetting up and down, and the mixture was incubated for 15 minutes at room temperature. The supernatant was removed and the well was washed with 1X B&W and 1x BSA as above. Then 50 µl BsaI HF mix (5 µl NEBuffer 4, 2 µl BsaI HF, 43 µl water) was added to the ligated nucleic acids (αβγδεβγδε) bound to the well. Two more TALE repeat sub-array units were ligated sequentially as above, the first a four TALE repeat sub-array unit (βγδε) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SNG and the second a three TALE repeat sub-array unit (βγδ) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SHD. The final TALE repeat array contained subunits of the format αβγδεβγδεβγδεβγδ with individual TALE repeats designed to bind a target DNA sequence.

Following the final ligation step, the fragments in the well were digested with BsaI HF for eventual cloning into an expression vector. The well was then washed with 1X B&W Buffer and twice with 1X BSA. Then 50 µl BbsI mix (5 µl NEBuffer 2, 5 µl BbsI, 40 µl water) was added to the well and incubated at 37 °C for 2 hours to cleave the biotinylated 5' end and release the assembled TALE repeat array from the well. The digested mixture was purified, ligated, and transformed as described in Example 1.

Example 3. Site-Specific Mutagenesis Using TALE Nucleases

To demonstrate the effectiveness of TALE repeat domains created by the methods described herein, TALE repeat arrays were constructed and cloned into TALE nuclease expression vectors (as described in Example 1) to produce plasmids encoding TALE nuclease monomers targeted to the eGFP coding sequences shown in FIG. 6 and Table 3. Nucleic acid and polypeptide sequences of the TALE nuclease monomers are shown in FIGs. 11A-18B.

**Table 3. TALE nuclease monomer target sequences**

| TALE Fragment | Target Sequence | Length of target sequence | SEQ ID NO: | Site | Position (half-site) | Plasmid name |
|---|---|---|---|---|---|---|
| DR-TALE-0003 | TGCAGTGCTTCAGCCGC | 17 | 50 | eGFP223 | left | SQT70 |
| DR-TALE-0006 | TGCAGTGCTTCAGCCGCT | 18 | 51 | eGFP223 | left | SQT114 |
| DR-TALE-0005 | TTGAAGAAGTCGTGCTGC | 18 | 52 | eGFP223 | right | SQT72 |
| DR-TALE-0010 | TGAAGAAGTCGTGCTGCT | 18 | 53 | eGFP223 | right | SQT56 |
| DR-TALE-0023 | TCGAGCTGAAGGGCATC | 17 | 54 | eGFP382 | left | SQT84 |
| DR-TALE-0025 | TCGAGCTGAAGGGCATCG | 18 | 55 | eGFP382 | left | SQT120 |
| DR-TALE-0020 | TTGTGCCCCAGGATGTTG | 18 | 56 | eGFP382 | right | SQT135 |
| DR-TALE-0022 | TGTGCCCCAGGATGTTGC | 18 | 57 | eGFP382 | right | SQT118 |

4E5 U2OS-eGFP cells were nucleofected with 400 ng plasmid DNA in solution SE with program DN-100 using Nucleofector™ non-viral transfection (Lonza, Walkersville, MD). The cells were analyzed by flow cytometry at days 2 and 5 (FIG. 7). Non-homologous end joining (NHEJ)-mediated mutagenic repair of TALE nuclease-induced double-stranded breaks led to disruption of eGFP expression (eGFP-negative cells). All eight TALE nuclease pairs tested induced a high percentage of eGFP-negative (eGFP-) cells (y-axis). The percentage of eGFP- cells declined only modestly between day 2 and 5 suggesting that the alterations were stably induced.

A subset of mutated eGFP genes were amplified from cells and sequenced. The resulting mutations are shown in FIG. 8. Sequences targeted by the TALE nucleases encoded by expression plasmids SQT70/SQT56 in human USOS-eGFP cells are

underlined in the wild-type (WT) sequence shown at the top of FIG. 8. Insertion and deletion mutations induced by the TALE nuclease pair are shown below with deleted bases indicated by dashes and inserted bases indicated by double underlining. The net number of bases inserted or deleted is shown to the right. All mutations were isolated once unless otherwise indicated in brackets. The overall frequency of mutagenesis (46%) is also indicated.

Example 4. Automated Assembly of TALE Repeat Arrays

The assembly method described in Example 1 has been automated so as to be performed using a Sciclone™ G3 liquid handling workstation (Caliper Life Sciences, Hopkinton, MA) in 96-well plates. All of the steps were automated except digestion of the nucleic acids prior to ligation and linking to the beads and the steps following release of the assembled TALE repeat array from the magnetic beads. The automated steps were performed essentially as when done manually with minor variations in the number of resuspension and mixing motions. The results of assembly of two 17-mers are shown in FIG. 9. A major product of the expected size can be seen, corresponding to the 17-mer. Additional minor 13-mer, 9-mer, and 5-mer products can also be seen, likely produced by carry forward of incompletely ligated products. A similar result can be seen in FIG. 10, which shows the results of assembly of 16-mers from an N-terminal 1-mer sub-array (1), three 4-mer subarrays ($4_A$, $4_B$, $4_C$), and a C-terminal 3-mer subarray ($3_D$).

This example demonstrates that the methods described herein can be automated for rapid and reproducible synthesis of nucleic acids encoding TALE repeat arrays.

Example 5. Assembly Methods

TALE repeat arrays were created using an architecture in which four distinct TALE repeat backbones that differ slightly in their amino acid and DNA sequences occur in a repeated pattern. The first, amino-terminal TALE repeat in an array was designated as the α unit. This was followed by β, γ, and δ units and then an ε unit that is essentially identical to the α unit except for the different positioning of a Type IIS restriction site on the 5' end (required to enable creation of a unique overhang on the α unit needed for

cloning). The ε unit was then followed again by repeats of β, γ, δ, and ε units. Due to constraints related to creation of a 3' end required for cloning, slightly modified DNA sequences were required for TALE repeat arrays that end with a carboxy-terminal γ or ε unit. We designated these variant units as γ* and ε*.

For each type of TALE repeat unit (i.e.—α, β, γ, δ, ε, γ*, and ε*), we commercially synthesized (Genscript) a series of four plasmids, each harboring one of the five repeat variable di-residues (RVDs) that specifies one of the four DNA bases (NI = A; HD = C; NN = G; NG = T, NK = G). Full DNA sequences of these plasmids are provided in Table 4 and FIG. 3. For all 35 of these plasmids, the sequence encoding the TALE repeat domain is flanked on the 5' end by unique XbaI and BbsI restriction sites and on the 3' end by unique BsaI and BamHI restriction sites. Additionally, the overhangs generated by digestion of any plasmids encoding units designed to be adjacent to one another (e.g.—β and γ, or δ and ε) with BsaI and BbsI are complementary. Using these 35 different plasmids and serial ligation via the BsaI and BbsI restriction sites, we assembled an archive of all possible combinations of βγ, βγδε, βγδ, βγ*, and δε* repeats. In total, this archive consisted of 825 different plasmids encoding 5 α's, 5 β's, 25 βγ combinations, 625 βγδε combinations, 125 βγδ combinations, 25 βγ* combinations, and 25 δε* combinations (Table 5). These 825 plasmids plus ten of the original 35 plasmids encoding single TALE repeats (five α and five β plasmids) are required to practice the methods. With this archive of 835 plasmids listed in Table 5, the methods can be used to construct TALE repeat arrays of any desired length and composition.

**Table 4. DNA sequences encoding individual TALE repeats**

| TAL ID# | Unit Architecture | RVD | Target Base | DNA Sequence (Cloned between XbaI/BamHI in pUC57-ΔBsaI) | SEQ ID NO: |
|---|---|---|---|---|---|
| 6 | α | NI | A | TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTACTCGCAATCGCG TCGAACATTGGGGGAAAGCAAGCC CTGGAAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC | 58 |

| | | | | | |
|---|---|---|---|---|---|
| 7 | α | HD | C | TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTAGTCGCAATCGCG TCACATGACGGGGGAAAGCAAGCC CTGGAAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC | 59 |
| 8 | α | NK | G | TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTAGTCGCAATCGCG TCGAACAAAGGGGGAAAGCAAGCC CTGGAAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC | 60 |
| 9 | α | NN | G | TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTAGTCGCAATCGCG AACAATAATGGGGGAAAGCAAGCC CTGGAAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC | 61 |
| 10 | α | NG | T | TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTAGTCGCAATCGCG TCAAACGGAGGGGGAAAGCAAGCC CTGGAAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC | 62 |
| 11 | β | NI | A | TCTAGAGAAGACAACTTACACCGG AGCAAGTCGTGGCCATTGCAAGCA ACATCGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC | 63 |
| 12 | β | HD | C | TCTAGAGAAGACAACTTACACCGG AGCAAGTCGTGGCCATTGCATCCC ACGACGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC | 64 |
| 13 | β | NK | G | TCTAGAGAAGACAACTTACACCGG AGCAAGTCGTGGCCATTGCATCAA ATAAAGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC | 65 |
| 14 | β | NN | G | TCTAGAGAAGACAACTTACACCGG AGCAAGTCGTGGCCATTGCAAATA ATAACGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC | 66 |

| | | | | | |
|---|---|---|---|---|---|
| 15 | β | NG | T | TCTAGAGAAGACAACTTACACCGG AGCAAGTCGTGGCCATTGCAAGCA ATGGGGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC | 67 |
| 16 | γ | NI | A | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCGA ACATTGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC | 68 |
| 17 | γ | HD | C | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCGC ATGACGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC | 69 |
| 18 | γ | NK | G | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCCA ACAAGGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC | 70 |
| 19 | γ | NN | G | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGAATA ACAATGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC | 71 |
| 20 | γ | NG | T | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCCA ACGGTGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC | 72 |
| 21 | δ | NI | A | TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCTCCA ATATTGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC | 73 |
| 22 | δ | HD | C | TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCAGCC ATGATGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC | 74 |

| | | | | | |
|---|---|---|---|---|---|
| 23 | δ | NK | G | TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCAGCA ATAAGGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC | 75 |
| 24 | δ | NN | G | TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCAACA ACAACGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC | 76 |
| 25 | δ | NG | T | TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCTCGA ATGGCGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC | 77 |
| 26 | ε | NI | A | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGTCGA ACATTGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC | 78 |
| 27 | ε | HD | C | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGTCAC ATGACGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC | 79 |
| 28 | ε | NK | G | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGTCGA ACAAAGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC | 80 |
| 29 | ε | NN | G | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGAACA ATAATGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC | 81 |
| 30 | ε | NG | T | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGTCAA ACGGAGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC | 82 |

| | | | | | |
|---|---|---|---|---|---|
| 31 | γ' | NI | A | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCGA ACATTGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC | 83 |
| 32 | γ' | HD | C | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCGC ATGACGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC | 84 |
| 33 | γ' | NK | G | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCCA ACAAGGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC | 85 |
| 34 | γ' | NN | G | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGAATA ACAATGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC | 86 |
| 35 | γ' | NG | T | TCTAGAGAAGACAACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCCA ACGGTGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC | 87 |
| 36 | ε' | NI | A | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGTCGA ACATTGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC | 88 |
| 37 | ε' | HD | C | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGTCAC ATGACGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC | 89 |
| 38 | ε' | NK | G | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGTCGA ACAAAGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC | 90 |

| | | | | | |
|---|---|---|---|---|---|
| 39 | ε' | NN | G | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGAACA ATAATGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC | 91 |
| 40 | ε' | NG | T | TCTAGAGAAGACAACTGACCCCAG ACCAGGTAGTCGCAATCGCGTCAA ACGGAGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC | 92 |

**Table 5. Archive of 835 plasmids encoding pre-assembled TALE repeat units**

| Plasmid ID | DNA Target | RVDs | Unit Architecture |
|---|---|---|---|
| TAL006 | A | NI | α |
| TAL007 | C | HD | α |
| TAL008 | G | NK | α |
| TAL009 | G | NN | α |
| TAL010 | T | NG | α |
| TAL011/016/021/026 | AAAA | NI/NI/NI/NI | βγδε |
| TAL011/016/021/027 | AAAC | NI/NI/NI/HD | βγδε |
| TAL011/016/021/028 | AAAG | NI/NI/NI/NK | βγδε |
| TAL011/016/021/029 | AAAG | NI/NI/NI/NN | βγδε |
| TAL011/016/021/030 | AAAT | NI/NI/NI/NG | βγδε |
| TAL011/016/022/026 | AACA | NI/NI/HD/NI | βγδε |
| TAL011/016/022/027 | AACC | NI/NI/HD/HD | βγδε |
| TAL011/016/022/028 | AACG | NI/NI/HD/NK | βγδε |
| TAL011/016/022/029 | AACG | NI/NI/HD/NN | βγδε |
| TAL011/016/022/030 | AACT | NI/NI/HD/NG | βγδε |
| TAL011/016/023/026 | AAGA | NI/NI/NK/NI | βγδε |
| TAL011/016/023/027 | AAGC | NI/NI/NK/HD | βγδε |
| TAL011/016/023/028 | AAGG | NI/NI/NK/NK | βγδε |
| TAL011/016/023/029 | AAGG | NI/NI/NK/NN | βγδε |
| TAL011/016/023/030 | AAGT | NI/NI/NK/NG | βγδε |
| TAL011/016/024/026 | AAGA | NI/NI/NN/NI | βγδε |
| TAL011/016/024/027 | AAGC | NI/NI/NN/HD | βγδε |
| TAL011/016/024/028 | AAGG | NI/NI/NN/NK | βγδε |
| TAL011/016/024/029 | AAGG | NI/NI/NN/NN | βγδε |
| TAL011/016/024/030 | AAGT | NI/NI/NN/NG | βγδε |
| TAL011/016/025/026 | AATA | NI/NI/NG/NI | βγδε |

| | | | |
|---|---|---|---|
| TAL011/016/025/027 | AATC | NI/NI/NG/HD | βγδε |
| TAL011/016/025/028 | AATG | NI/NI/NG/NK | βγδε |
| TAL011/016/025/029 | AATG | NI/NI/NG/NN | βγδε |
| TAL011/016/025/030 | AATT | NI/NI/NG/NG | βγδε |
| TAL011/017/021/026 | ACAA | NI/HD/NI/NI | βγδε |
| TAL011/017/021/027 | ACAC | NI/HD/NI/HD | βγδε |
| TAL011/017/021/028 | ACAG | NI/HD/NI/NK | βγδε |
| TAL011/017/021/029 | ACAG | NI/HD/NI/NN | βγδε |
| TAL011/017/021/030 | ACAT | NI/HD/NI/NG | βγδε |
| TAL011/017/022/026 | ACCA | NI/HD/HD/NI | βγδε |
| TAL011/017/022/027 | ACCC | NI/HD/HD/HD | βγδε |
| TAL011/017/022/028 | ACCG | NI/HD/HD/NK | βγδε |
| TAL011/017/022/029 | ACCG | NI/HD/HD/NN | βγδε |
| TAL011/017/022/030 | ACCT | NI/HD/HD/NG | βγδε |
| TAL011/017/023/026 | ACGA | NI/HD/NK/NI | βγδε |
| TAL011/017/023/027 | ACGC | NI/HD/NK/HD | βγδε |
| TAL011/017/023/028 | ACGG | NI/HD/NK/NK | βγδε |
| TAL011/017/023/029 | ACGG | NI/HD/NK/NN | βγδε |
| TAL011/017/023/030 | ACGT | NI/HD/NK/NG | βγδε |
| TAL011/017/024/026 | ACGA | NI/HD/NN/NI | βγδε |
| TAL011/017/024/027 | ACGC | NI/HD/NN/HD | βγδε |
| TAL011/017/024/028 | ACGG | NI/HD/NN/NK | βγδε |
| TAL011/017/024/029 | ACGG | NI/HD/NN/NN | βγδε |
| TAL011/017/024/030 | ACGT | NI/HD/NN/NG | βγδε |
| TAL011/017/025/026 | ACTA | NI/HD/NG/NI | βγδε |
| TAL011/017/025/027 | ACTC | NI/HD/NG/HD | βγδε |
| TAL011/017/025/028 | ACTG | NI/HD/NG/NK | βγδε |
| TAL011/017/025/029 | ACTG | NI/HD/NG/NN | βγδε |
| TAL011/017/025/030 | ACTT | NI/HD/NG/NG | βγδε |
| TAL011/018/021/026 | AGAA | NI/NK/NI/NI | βγδε |
| TAL011/018/021/027 | AGAC | NI/NK/NI/HD | βγδε |
| TAL011/018/021/028 | AGAG | NI/NK/NI/NK | βγδε |
| TAL011/018/021/029 | AGAG | NI/NK/NI/NN | βγδε |
| TAL011/018/021/030 | AGAT | NI/NK/NI/NG | βγδε |
| TAL011/018/022/026 | AGCA | NI/NK/HD/NI | βγδε |
| TAL011/018/022/027 | AGCC | NI/NK/HD/HD | βγδε |
| TAL011/018/022/028 | AGCG | NI/NK/HD/NK | βγδε |
| TAL011/018/022/029 | AGCG | NI/NK/HD/NN | βγδε |
| TAL011/018/022/030 | AGCT | NI/NK/HD/NG | βγδε |
| TAL011/018/023/026 | AGGA | NI/NK/NK/NI | βγδε |
| TAL011/018/023/027 | AGGC | NI/NK/NK/HD | βγδε |

| | | | |
|---|---|---|---|
| TAL011/018/023/028 | AGGG | NI/NK/NK/NK | βγδε |
| TAL011/018/023/029 | AGGG | NI/NK/NK/NN | βγδε |
| TAL011/018/023/030 | AGGT | NI/NK/NK/NG | βγδε |
| TAL011/018/024/026 | AGGA | NI/NK/NN/NI | βγδε |
| TAL011/018/024/027 | AGGC | NI/NK/NN/HD | βγδε |
| TAL011/018/024/028 | AGGG | NI/NK/NN/NK | βγδε |
| TAL011/018/024/029 | AGGG | NI/NK/NN/NN | βγδε |
| TAL011/018/024/030 | AGGT | NI/NK/NN/NG | βγδε |
| TAL011/018/025/026 | AGTA | NI/NK/NG/NI | βγδε |
| TAL011/018/025/027 | AGTC | NI/NK/NG/HD | βγδε |
| TAL011/018/025/028 | AGTG | NI/NK/NG/NK | βγδε |
| TAL011/018/025/029 | AGTG | NI/NK/NG/NN | βγδε |
| TAL011/018/025/030 | AGTT | NI/NK/NG/NG | βγδε |
| TAL011/019/021/026 | AGAA | NI/NN/NI/NI | βγδε |
| TAL011/019/021/027 | AGAC | NI/NN/NI/HD | βγδε |
| TAL011/019/021/028 | AGAG | NI/NN/NI/NK | βγδε |
| TAL011/019/021/029 | AGAG | NI/NN/NI/NN | βγδε |
| TAL011/019/021/030 | AGAT | NI/NN/NI/NG | βγδε |
| TAL011/019/022/026 | AGCA | NI/NN/HD/NI | βγδε |
| TAL011/019/022/027 | AGCC | NI/NN/HD/HD | βγδε |
| TAL011/019/022/028 | AGCG | NI/NN/HD/NK | βγδε |
| TAL011/019/022/029 | AGCG | NI/NN/HD/NN | βγδε |
| TAL011/019/022/030 | AGCT | NI/NN/HD/NG | βγδε |
| TAL011/019/023/026 | AGGA | NI/NN/NK/NI | βγδε |
| TAL011/019/023/027 | AGGC | NI/NN/NK/HD | βγδε |
| TAL011/019/023/028 | AGGG | NI/NN/NK/NK | βγδε |
| TAL011/019/023/029 | AGGG | NI/NN/NK/NN | βγδε |
| TAL011/019/023/030 | AGGT | NI/NN/NK/NG | βγδε |
| TAL011/019/024/026 | AGGA | NI/NN/NN/NI | βγδε |
| TAL011/019/024/027 | AGGC | NI/NN/NN/HD | βγδε |
| TAL011/019/024/028 | AGGG | NI/NN/NN/NK | βγδε |
| TAL011/019/024/029 | AGGG | NI/NN/NN/NN | βγδε |
| TAL011/019/024/030 | AGGT | NI/NN/NN/NG | βγδε |
| TAL011/019/025/026 | AGTA | NI/NN/NG/NI | βγδε |
| TAL011/019/025/027 | AGTC | NI/NN/NG/HD | βγδε |
| TAL011/019/025/028 | AGTG | NI/NN/NG/NK | βγδε |
| TAL011/019/025/029 | AGTG | NI/NN/NG/NN | βγδε |
| TAL011/019/025/030 | AGTT | NI/NN/NG/NG | βγδε |
| TAL011/020/021/026 | ATAA | NI/NG/NI/NI | βγδε |
| TAL011/020/021/027 | ATAC | NI/NG/NI/HD | βγδε |
| TAL011/020/021/028 | ATAG | NI/NG/NI/NK | βγδε |

| | | | |
|---|---|---|---|
| TAL011/020/021/029 | ATAG | NI/NG/NI/NN | βγδε |
| TAL011/020/021/030 | ATAT | NI/NG/NI/NG | βγδε |
| TAL011/020/022/026 | ATCA | NI/NG/HD/NI | βγδε |
| TAL011/020/022/027 | ATCC | NI/NG/HD/HD | βγδε |
| TAL011/020/022/028 | ATCG | NI/NG/HD/NK | βγδε |
| TAL011/020/022/029 | ATCG | NI/NG/HD/NN | βγδε |
| TAL011/020/022/030 | ATCT | NI/NG/HD/NG | βγδε |
| TAL011/020/023/026 | ATGA | NI/NG/NK/NI | βγδε |
| TAL011/020/023/027 | ATGC | NI/NG/NK/HD | βγδε |
| TAL011/020/023/028 | ATGG | NI/NG/NK/NK | βγδε |
| TAL011/020/023/029 | ATGG | NI/NG/NK/NN | βγδε |
| TAL011/020/023/030 | ATGT | NI/NG/NK/NG | βγδε |
| TAL011/020/024/026 | ATGA | NI/NG/NN/NI | βγδε |
| TAL011/020/024/027 | ATGC | NI/NG/NN/HD | βγδε |
| TAL011/020/024/028 | ATGG | NI/NG/NN/NK | βγδε |
| TAL011/020/024/029 | ATGG | NI/NG/NN/NN | βγδε |
| TAL011/020/024/030 | ATGT | NI/NG/NN/NG | βγδε |
| TAL011/020/025/026 | ATTA | NI/NG/NG/NI | βγδε |
| TAL011/020/025/027 | ATTC | NI/NG/NG/HD | βγδε |
| TAL011/020/025/028 | ATTG | NI/NG/NG/NK | βγδε |
| TAL011/020/025/029 | ATTG | NI/NG/NG/NN | βγδε |
| TAL011/020/025/030 | ATTT | NI/NG/NG/NG | βγδε |
| TAL012/016/021/026 | CAAA | HD/NI/NI/NI | βγδε |
| TAL012/016/021/027 | CAAC | HD/NI/NI/HD | βγδε |
| TAL012/016/021/028 | CAAG | HD/NI/NI/NK | βγδε |
| TAL012/016/021/029 | CAAG | HD/NI/NI/NN | βγδε |
| TAL012/016/021/030 | CAAT | HD/NI/NI/NG | βγδε |
| TAL012/016/022/026 | CACA | HD/NI/HD/NI | βγδε |
| TAL012/016/022/027 | CACC | HD/NI/HD/HD | βγδε |
| TAL012/016/022/028 | CACG | HD/NI/HD/NK | βγδε |
| TAL012/016/022/029 | CACG | HD/NI/HD/NN | βγδε |
| TAL012/016/022/030 | CACT | HD/NI/HD/NG | βγδε |
| TAL012/016/023/026 | CAGA | HD/NI/NK/NI | βγδε |
| TAL012/016/023/027 | CAGC | HD/NI/NK/HD | βγδε |
| TAL012/016/023/028 | CAGG | HD/NI/NK/NK | βγδε |
| TAL012/016/023/029 | CAGG | HD/NI/NK/NN | βγδε |
| TAL012/016/023/030 | CAGT | HD/NI/NK/NG | βγδε |
| TAL012/016/024/026 | CAGA | HD/NI/NN/NI | βγδε |
| TAL012/016/024/027 | CAGC | HD/NI/NN/HD | βγδε |
| TAL012/016/024/028 | CAGG | HD/NI/NN/NK | βγδε |
| TAL012/016/024/029 | CAGG | HD/NI/NN/NN | βγδε |

| | | | |
|---|---|---|---|
| TAL012/016/024/030 | CAGT | HD/NI/NN/NG | βγδε |
| TAL012/016/025/026 | CATA | HD/NI/NG/NI | βγδε |
| TAL012/016/025/027 | CATC | HD/NI/NG/HD | βγδε |
| TAL012/016/025/028 | CATG | HD/NI/NG/NK | βγδε |
| TAL012/016/025/029 | CATG | HD/NI/NG/NN | βγδε |
| TAL012/016/025/030 | CATT | HD/NI/NG/NG | βγδε |
| TAL012/017/021/026 | CCAA | HD/HD/NI/NI | βγδε |
| TAL012/017/021/027 | CCAC | HD/HD/NI/HD | βγδε |
| TAL012/017/021/028 | CCAG | HD/HD/NI/NK | βγδε |
| TAL012/017/021/029 | CCAG | HD/HD/NI/NN | βγδε |
| TAL012/017/021/030 | CCAT | HD/HD/NI/NG | βγδε |
| TAL012/017/022/026 | CCCA | HD/HD/HD/NI | βγδε |
| TAL012/017/022/027 | CCCC | HD/HD/HD/HD | βγδε |
| TAL012/017/022/028 | CCCG | HD/HD/HD/NK | βγδε |
| TAL012/017/022/029 | CCCG | HD/HD/HD/NN | βγδε |
| TAL012/017/022/030 | CCCT | HD/HD/HD/NG | βγδε |
| TAL012/017/023/026 | CCGA | HD/HD/NK/NI | βγδε |
| TAL012/017/023/027 | CCGC | HD/HD/NK/HD | βγδε |
| TAL012/017/023/028 | CCGG | HD/HD/NK/NK | βγδε |
| TAL012/017/023/029 | CCGG | HD/HD/NK/NN | βγδε |
| TAL012/017/023/030 | CCGT | HD/HD/NK/NG | βγδε |
| TAL012/017/024/026 | CCGA | HD/HD/NN/NI | βγδε |
| TAL012/017/024/027 | CCGC | HD/HD/NN/HD | βγδε |
| TAL012/017/024/028 | CCGG | HD/HD/NN/NK | βγδε |
| TAL012/017/024/029 | CCGG | HD/HD/NN/NN | βγδε |
| TAL012/017/024/030 | CCGT | HD/HD/NN/NG | βγδε |
| TAL012/017/025/026 | CCTA | HD/HD/NG/NI | βγδε |
| TAL012/017/025/027 | CCTC | HD/HD/NG/HD | βγδε |
| TAL012/017/025/028 | CCTG | HD/HD/NG/NK | βγδε |
| TAL012/017/025/029 | CCTG | HD/HD/NG/NN | βγδε |
| TAL012/017/025/030 | CCTT | HD/HD/NG/NG | βγδε |
| TAL012/018/021/026 | CGAA | HD/NK/NI/NI | βγδε |
| TAL012/018/021/027 | CGAC | HD/NK/NI/HD | βγδε |
| TAL012/018/021/028 | CGAG | HD/NK/NI/NK | βγδε |
| TAL012/018/021/029 | CGAG | HD/NK/NI/NN | βγδε |
| TAL012/018/021/030 | CGAT | HD/NK/NI/NG | βγδε |
| TAL012/018/022/026 | CGCA | HD/NK/HD/NI | βγδε |
| TAL012/018/022/027 | CGCC | HD/NK/HD/HD | βγδε |
| TAL012/018/022/028 | CGCG | HD/NK/HD/NK | βγδε |
| TAL012/018/022/029 | CGCG | HD/NK/HD/NN | βγδε |
| TAL012/018/022/030 | CGCT | HD/NK/HD/NG | βγδε |

| | | | |
|---|---|---|---|
| TAL012/018/023/026 | CGGA | HD/NK/NK/NI | βγδε |
| TAL012/018/023/027 | CGGC | HD/NK/NK/HD | βγδε |
| TAL012/018/023/028 | CGGG | HD/NK/NK/NK | βγδε |
| TAL012/018/023/029 | CGGG | HD/NK/NK/NN | βγδε |
| TAL012/018/023/030 | CGGT | HD/NK/NK/NG | βγδε |
| TAL012/018/024/026 | CGGA | HD/NK/NN/NI | βγδε |
| TAL012/018/024/027 | CGGC | HD/NK/NN/HD | βγδε |
| TAL012/018/024/028 | CGGG | HD/NK/NN/NK | βγδε |
| TAL012/018/024/029 | CGGG | HD/NK/NN/NN | βγδε |
| TAL012/018/024/030 | CGGT | HD/NK/NN/NG | βγδε |
| TAL012/018/025/026 | CGTA | HD/NK/NG/NI | βγδε |
| TAL012/018/025/027 | CGTC | HD/NK/NG/HD | βγδε |
| TAL012/018/025/028 | CGTG | HD/NK/NG/NK | βγδε |
| TAL012/018/025/029 | CGTG | HD/NK/NG/NN | βγδε |
| TAL012/018/025/030 | CGTT | HD/NK/NG/NG | βγδε |
| TAL012/019/021/026 | CGAA | HD/NN/NI/NI | βγδε |
| TAL012/019/021/027 | CGAC | HD/NN/NI/HD | βγδε |
| TAL012/019/021/028 | CGAG | HD/NN/NI/NK | βγδε |
| TAL012/019/021/029 | CGAG | HD/NN/NI/NN | βγδε |
| TAL012/019/021/030 | CGAT | HD/NN/NI/NG | βγδε |
| TAL012/019/022/026 | CGCA | HD/NN/HD/NI | βγδε |
| TAL012/019/022/027 | CGCC | HD/NN/HD/HD | βγδε |
| TAL012/019/022/028 | CGCG | HD/NN/HD/NK | βγδε |
| TAL012/019/022/029 | CGCG | HD/NN/HD/NN | βγδε |
| TAL012/019/022/030 | CGCT | HD/NN/HD/NG | βγδε |
| TAL012/019/023/026 | CGGA | HD/NN/NK/NI | βγδε |
| TAL012/019/023/027 | CGGC | HD/NN/NK/HD | βγδε |
| TAL012/019/023/028 | CGGG | HD/NN/NK/NK | βγδε |
| TAL012/019/023/029 | CGGG | HD/NN/NK/NN | βγδε |
| TAL012/019/023/030 | CGGT | HD/NN/NK/NG | βγδε |
| TAL012/019/024/026 | CGGA | HD/NN/NN/NI | βγδε |
| TAL012/019/024/027 | CGGC | HD/NN/NN/HD | βγδε |
| TAL012/019/024/028 | CGGG | HD/NN/NN/NK | βγδε |
| TAL012/019/024/029 | CGGG | HD/NN/NN/NN | βγδε |
| TAL012/019/024/030 | CGGT | HD/NN/NN/NG | βγδε |
| TAL012/019/025/026 | CGTA | HD/NN/NG/NI | βγδε |
| TAL012/019/025/027 | CGTC | HD/NN/NG/HD | βγδε |
| TAL012/019/025/028 | CGTG | HD/NN/NG/NK | βγδε |
| TAL012/019/025/029 | CGTG | HD/NN/NG/NN | βγδε |
| TAL012/019/025/030 | CGTT | HD/NN/NG/NG | βγδε |
| TAL012/020/021/026 | CTAA | HD/NG/NI/NI | βγδε |

| | | | |
|---|---|---|---|
| TAL012/020/021/027 | CTAC | HD/NG/NI/HD | βγδε |
| TAL012/020/021/028 | CTAG | HD/NG/NI/NK | βγδε |
| TAL012/020/021/029 | CTAG | HD/NG/NI/NN | βγδε |
| TAL012/020/021/030 | CTAT | HD/NG/NI/NG | βγδε |
| TAL012/020/022/026 | CTCA | HD/NG/HD/NI | βγδε |
| TAL012/020/022/027 | CTCC | HD/NG/HD/HD | βγδε |
| TAL012/020/022/028 | CTCG | HD/NG/HD/NK | βγδε |
| TAL012/020/022/029 | CTCG | HD/NG/HD/NN | βγδε |
| TAL012/020/022/030 | CTCT | HD/NG/HD/NG | βγδε |
| TAL012/020/023/026 | CTGA | HD/NG/NK/NI | βγδε |
| TAL012/020/023/027 | CTGC | HD/NG/NK/HD | βγδε |
| TAL012/020/023/028 | CTGG | HD/NG/NK/NK | βγδε |
| TAL012/020/023/029 | CTGG | HD/NG/NK/NN | βγδε |
| TAL012/020/023/030 | CTGT | HD/NG/NK/NG | βγδε |
| TAL012/020/024/026 | CTGA | HD/NG/NN/NI | βγδε |
| TAL012/020/024/027 | CTGC | HD/NG/NN/HD | βγδε |
| TAL012/020/024/028 | CTGG | HD/NG/NN/NK | βγδε |
| TAL012/020/024/029 | CTGG | HD/NG/NN/NN | βγδε |
| TAL012/020/024/030 | CTGT | HD/NG/NN/NG | βγδε |
| TAL012/020/025/026 | CTTA | HD/NG/NG/NI | βγδε |
| TAL012/020/025/027 | CTTC | HD/NG/NG/HD | βγδε |
| TAL012/020/025/028 | CTTG | HD/NG/NG/NK | βγδε |
| TAL012/020/025/029 | CTTG | HD/NG/NG/NN | βγδε |
| TAL012/020/025/030 | CTTT | HD/NG/NG/NG | βγδε |
| TAL013/016/021/026 | GAAA | NK/NI/NI/NI | βγδε |
| TAL013/016/021/027 | GAAC | NK/NI/NI/HD | βγδε |
| TAL013/016/021/028 | GAAG | NK/NI/NI/NK | βγδε |
| TAL013/016/021/029 | GAAG | NK/NI/NI/NN | βγδε |
| TAL013/016/021/030 | GAAT | NK/NI/NI/NG | βγδε |
| TAL013/016/022/026 | GACA | NK/NI/HD/NI | βγδε |
| TAL013/016/022/027 | GACC | NK/NI/HD/HD | βγδε |
| TAL013/016/022/028 | GACG | NK/NI/HD/NK | βγδε |
| TAL013/016/022/029 | GACG | NK/NI/HD/NN | βγδε |
| TAL013/016/022/030 | GACT | NK/NI/HD/NG | βγδε |
| TAL013/016/023/026 | GAGA | NK/NI/NK/NI | βγδε |
| TAL013/016/023/027 | GAGC | NK/NI/NK/HD | βγδε |
| TAL013/016/023/028 | GAGG | NK/NI/NK/NK | βγδε |
| TAL013/016/023/029 | GAGG | NK/NI/NK/NN | βγδε |
| TAL013/016/023/030 | GAGT | NK/NI/NK/NG | βγδε |
| TAL013/016/024/026 | GAGA | NK/NI/NN/NI | βγδε |
| TAL013/016/024/027 | GAGC | NK/NI/NN/HD | βγδε |

| | | | |
|---|---|---|---|
| TAL013/016/024/028 | GAGG | NK/NI/NN/NK | βγδε |
| TAL013/016/024/029 | GAGG | NK/NI/NN/NN | βγδε |
| TAL013/016/024/030 | GAGT | NK/NI/NN/NG | βγδε |
| TAL013/016/025/026 | GATA | NK/NI/NG/NI | βγδε |
| TAL013/016/025/027 | GATC | NK/NI/NG/HD | βγδε |
| TAL013/016/025/028 | GATG | NK/NI/NG/NK | βγδε |
| TAL013/016/025/029 | GATG | NK/NI/NG/NN | βγδε |
| TAL013/016/025/030 | GATT | NK/NI/NG/NG | βγδε |
| TAL013/017/021/026 | GCAA | NK/HD/NI/NI | βγδε |
| TAL013/017/021/027 | GCAC | NK/HD/NI/HD | βγδε |
| TAL013/017/021/028 | GCAG | NK/HD/NI/NK | βγδε |
| TAL013/017/021/029 | GCAG | NK/HD/NI/NN | βγδε |
| TAL013/017/021/030 | GCAT | NK/HD/NI/NG | βγδε |
| TAL013/017/022/026 | GCCA | NK/HD/HD/NI | βγδε |
| TAL013/017/022/027 | GCCC | NK/HD/HD/HD | βγδε |
| TAL013/017/022/028 | GCCG | NK/HD/HD/NK | βγδε |
| TAL013/017/022/029 | GCCG | NK/HD/HD/NN | βγδε |
| TAL013/017/022/030 | GCCT | NK/HD/HD/NG | βγδε |
| TAL013/017/023/026 | GCGA | NK/HD/NK/NI | βγδε |
| TAL013/017/023/027 | GCGC | NK/HD/NK/HD | βγδε |
| TAL013/017/023/028 | GCGG | NK/HD/NK/NK | βγδε |
| TAL013/017/023/029 | GCGG | NK/HD/NK/NN | βγδε |
| TAL013/017/023/030 | GCGT | NK/HD/NK/NG | βγδε |
| TAL013/017/024/026 | GCGA | NK/HD/NN/NI | βγδε |
| TAL013/017/024/027 | GCGC | NK/HD/NN/HD | βγδε |
| TAL013/017/024/028 | GCGG | NK/HD/NN/NK | βγδε |
| TAL013/017/024/029 | GCGG | NK/HD/NN/NN | βγδε |
| TAL013/017/024/030 | GCGT | NK/HD/NN/NG | βγδε |
| TAL013/017/025/026 | GCTA | NK/HD/NG/NI | βγδε |
| TAL013/017/025/027 | GCTC | NK/HD/NG/HD | βγδε |
| TAL013/017/025/028 | GCTG | NK/HD/NG/NK | βγδε |
| TAL013/017/025/029 | GCTG | NK/HD/NG/NN | βγδε |
| TAL013/017/025/030 | GCTT | NK/HD/NG/NG | βγδε |
| TAL013/018/021/026 | GGAA | NK/NK/NI/NI | βγδε |
| TAL013/018/021/027 | GGAC | NK/NK/NI/HD | βγδε |
| TAL013/018/021/028 | GGAG | NK/NK/NI/NK | βγδε |
| TAL013/018/021/029 | GGAG | NK/NK/NI/NN | βγδε |
| TAL013/018/021/030 | GGAT | NK/NK/NI/NG | βγδε |
| TAL013/018/022/026 | GGCA | NK/NK/HD/NI | βγδε |
| TAL013/018/022/027 | GGCC | NK/NK/HD/HD | βγδε |
| TAL013/018/022/028 | GGCG | NK/NK/HD/NK | βγδε |

| | | | |
|---|---|---|---|
| TAL013/018/022/029 | GGCG | NK/NK/HD/NN | βγδε |
| TAL013/018/022/030 | GGCT | NK/NK/HD/NG | βγδε |
| TAL013/018/023/026 | GGGA | NK/NK/NK/NI | βγδε |
| TAL013/018/023/027 | GGGC | NK/NK/NK/HD | βγδε |
| TAL013/018/023/028 | GGGG | NK/NK/NK/NK | βγδε |
| TAL013/018/023/029 | GGGG | NK/NK/NK/NN | βγδε |
| TAL013/018/023/030 | GGGT | NK/NK/NK/NG | βγδε |
| TAL013/018/024/026 | GGGA | NK/NK/NN/NI | βγδε |
| TAL013/018/024/027 | GGGC | NK/NK/NN/HD | βγδε |
| TAL013/018/024/028 | GGGG | NK/NK/NN/NK | βγδε |
| TAL013/018/024/029 | GGGG | NK/NK/NN/NN | βγδε |
| TAL013/018/024/030 | GGGT | NK/NK/NN/NG | βγδε |
| TAL013/018/025/026 | GGTA | NK/NK/NG/NI | βγδε |
| TAL013/018/025/027 | GGTC | NK/NK/NG/HD | βγδε |
| TAL013/018/025/028 | GGTG | NK/NK/NG/NK | βγδε |
| TAL013/018/025/029 | GGTG | NK/NK/NG/NN | βγδε |
| TAL013/018/025/030 | GGTT | NK/NK/NG/NG | βγδε |
| TAL013/019/021/026 | GGAA | NK/NN/NI/NI | βγδε |
| TAL013/019/021/027 | GGAC | NK/NN/NI/HD | βγδε |
| TAL013/019/021/028 | GGAG | NK/NN/NI/NK | βγδε |
| TAL013/019/021/029 | GGAG | NK/NN/NI/NN | βγδε |
| TAL013/019/021/030 | GGAT | NK/NN/NI/NG | βγδε |
| TAL013/019/022/026 | GGCA | NK/NN/HD/NI | βγδε |
| TAL013/019/022/027 | GGCC | NK/NN/HD/HD | βγδε |
| TAL013/019/022/028 | GGCG | NK/NN/HD/NK | βγδε |
| TAL013/019/022/029 | GGCG | NK/NN/HD/NN | βγδε |
| TAL013/019/022/030 | GGCT | NK/NN/HD/NG | βγδε |
| TAL013/019/023/026 | GGGA | NK/NN/NK/NI | βγδε |
| TAL013/019/023/027 | GGGC | NK/NN/NK/HD | βγδε |
| TAL013/019/023/028 | GGGG | NK/NN/NK/NK | βγδε |
| TAL013/019/023/029 | GGGG | NK/NN/NK/NN | βγδε |
| TAL013/019/023/030 | GGGT | NK/NN/NK/NG | βγδε |
| TAL013/019/024/026 | GGGA | NK/NN/NN/NI | βγδε |
| TAL013/019/024/027 | GGGC | NK/NN/NN/HD | βγδε |
| TAL013/019/024/028 | GGGG | NK/NN/NN/NK | βγδε |
| TAL013/019/024/029 | GGGG | NK/NN/NN/NN | βγδε |
| TAL013/019/024/030 | GGGT | NK/NN/NN/NG | βγδε |
| TAL013/019/025/026 | GGTA | NK/NN/NG/NI | βγδε |
| TAL013/019/025/027 | GGTC | NK/NN/NG/HD | βγδε |
| TAL013/019/025/028 | GGTG | NK/NN/NG/NK | βγδε |
| TAL013/019/025/029 | GGTG | NK/NN/NG/NN | βγδε |

| | | | |
|---|---|---|---|
| TAL013/019/025/030 | GGTT | NK/NN/NG/NG | βγδε |
| TAL013/020/021/026 | GTAA | NK/NG/NI/NI | βγδε |
| TAL013/020/021/027 | GTAC | NK/NG/NI/HD | βγδε |
| TAL013/020/021/028 | GTAG | NK/NG/NI/NK | βγδε |
| TAL013/020/021/029 | GTAG | NK/NG/NI/NN | βγδε |
| TAL013/020/021/030 | GTAT | NK/NG/NI/NG | βγδε |
| TAL013/020/022/026 | GTCA | NK/NG/HD/NI | βγδε |
| TAL013/020/022/027 | GTCC | NK/NG/HD/HD | βγδε |
| TAL013/020/022/028 | GTCG | NK/NG/HD/NK | βγδε |
| TAL013/020/022/029 | GTCG | NK/NG/HD/NN | βγδε |
| TAL013/020/022/030 | GTCT | NK/NG/HD/NG | βγδε |
| TAL013/020/023/026 | GTGA | NK/NG/NK/NI | βγδε |
| TAL013/020/023/027 | GTGC | NK/NG/NK/HD | βγδε |
| TAL013/020/023/028 | GTGG | NK/NG/NK/NK | βγδε |
| TAL013/020/023/029 | GTGG | NK/NG/NK/NN | βγδε |
| TAL013/020/023/030 | GTGT | NK/NG/NK/NG | βγδε |
| TAL013/020/024/026 | GTGA | NK/NG/NN/NI | βγδε |
| TAL013/020/024/027 | GTGC | NK/NG/NN/HD | βγδε |
| TAL013/020/024/028 | GTGG | NK/NG/NN/NK | βγδε |
| TAL013/020/024/029 | GTGG | NK/NG/NN/NN | βγδε |
| TAL013/020/024/030 | GTGT | NK/NG/NN/NG | βγδε |
| TAL013/020/025/026 | GTTA | NK/NG/NG/NI | βγδε |
| TAL013/020/025/027 | GTTC | NK/NG/NG/HD | βγδε |
| TAL013/020/025/028 | GTTG | NK/NG/NG/NK | βγδε |
| TAL013/020/025/029 | GTTG | NK/NG/NG/NN | βγδε |
| TAL013/020/025/030 | GTTT | NK/NG/NG/NG | βγδε |
| TAL014/016/021/026 | GAAA | NN/NI/NI/NI | βγδε |
| TAL014/016/021/027 | GAAC | NN/NI/NI/HD | βγδε |
| TAL014/016/021/028 | GAAG | NN/NI/NI/NK | βγδε |
| TAL014/016/021/029 | GAAG | NN/NI/NI/NN | βγδε |
| TAL014/016/021/030 | GAAT | NN/NI/NI/NG | βγδε |
| TAL014/016/022/026 | GACA | NN/NI/HD/NI | βγδε |
| TAL014/016/022/027 | GACC | NN/NI/HD/HD | βγδε |
| TAL014/016/022/028 | GACG | NN/NI/HD/NK | βγδε |
| TAL014/016/022/029 | GACG | NN/NI/HD/NN | βγδε |
| TAL014/016/022/030 | GACT | NN/NI/HD/NG | βγδε |
| TAL014/016/023/026 | GAGA | NN/NI/NK/NI | βγδε |
| TAL014/016/023/027 | GAGC | NN/NI/NK/HD | βγδε |
| TAL014/016/023/028 | GAGG | NN/NI/NK/NK | βγδε |
| TAL014/016/023/029 | GAGG | NN/NI/NK/NN | βγδε |
| TAL014/016/023/030 | GAGT | NN/NI/NK/NG | βγδε |

| | | | |
|---|---|---|---|
| TAL014/016/024/026 | GAGA | NN/NI/NN/NI | βγδε |
| TAL014/016/024/027 | GAGC | NN/NI/NN/HD | βγδε |
| TAL014/016/024/028 | GAGG | NN/NI/NN/NK | βγδε |
| TAL014/016/024/029 | GAGG | NN/NI/NN/NN | βγδε |
| TAL014/016/024/030 | GAGT | NN/NI/NN/NG | βγδε |
| TAL014/016/025/026 | GATA | NN/NI/NG/NI | βγδε |
| TAL014/016/025/027 | GATC | NN/NI/NG/HD | βγδε |
| TAL014/016/025/028 | GATG | NN/NI/NG/NK | βγδε |
| TAL014/016/025/029 | GATG | NN/NI/NG/NN | βγδε |
| TAL014/016/025/030 | GATT | NN/NI/NG/NG | βγδε |
| TAL014/017/021/026 | GCAA | NN/HD/NI/NI | βγδε |
| TAL014/017/021/027 | GCAC | NN/HD/NI/HD | βγδε |
| TAL014/017/021/028 | GCAG | NN/HD/NI/NK | βγδε |
| TAL014/017/021/029 | GCAG | NN/HD/NI/NN | βγδε |
| TAL014/017/021/030 | GCAT | NN/HD/NI/NG | βγδε |
| TAL014/017/022/026 | GCCA | NN/HD/HD/NI | βγδε |
| TAL014/017/022/027 | GCCC | NN/HD/HD/HD | βγδε |
| TAL014/017/022/028 | GCCG | NN/HD/HD/NK | βγδε |
| TAL014/017/022/029 | GCCG | NN/HD/HD/NN | βγδε |
| TAL014/017/022/030 | GCCT | NN/HD/HD/NG | βγδε |
| TAL014/017/023/026 | GCGA | NN/HD/NK/NI | βγδε |
| TAL014/017/023/027 | GCGC | NN/HD/NK/HD | βγδε |
| TAL014/017/023/028 | GCGG | NN/HD/NK/NK | βγδε |
| TAL014/017/023/029 | GCGG | NN/HD/NK/NN | βγδε |
| TAL014/017/023/030 | GCGT | NN/HD/NK/NG | βγδε |
| TAL014/017/024/026 | GCGA | NN/HD/NN/NI | βγδε |
| TAL014/017/024/027 | GCGC | NN/HD/NN/HD | βγδε |
| TAL014/017/024/028 | GCGG | NN/HD/NN/NK | βγδε |
| TAL014/017/024/029 | GCGG | NN/HD/NN/NN | βγδε |
| TAL014/017/024/030 | GCGT | NN/HD/NN/NG | βγδε |
| TAL014/017/025/026 | GCTA | NN/HD/NG/NI | βγδε |
| TAL014/017/025/027 | GCTC | NN/HD/NG/HD | βγδε |
| TAL014/017/025/028 | GCTG | NN/HD/NG/NK | βγδε |
| TAL014/017/025/029 | GCTG | NN/HD/NG/NN | βγδε |
| TAL014/017/025/030 | GCTT | NN/HD/NG/NG | βγδε |
| TAL014/018/021/026 | GGAA | NN/NK/NI/NI | βγδε |
| TAL014/018/021/027 | GGAC | NN/NK/NI/HD | βγδε |
| TAL014/018/021/028 | GGAG | NN/NK/NI/NK | βγδε |
| TAL014/018/021/029 | GGAG | NN/NK/NI/NN | βγδε |
| TAL014/018/021/030 | GGAT | NN/NK/NI/NG | βγδε |
| TAL014/018/022/026 | GGCA | NN/NK/HD/NI | βγδε |

| | | | |
|---|---|---|---|
| TAL014/018/022/027 | GGCC | NN/NK/HD/HD | βγδε |
| TAL014/018/022/028 | GGCG | NN/NK/HD/NK | βγδε |
| TAL014/018/022/029 | GGCG | NN/NK/HD/NN | βγδε |
| TAL014/018/022/030 | GGCT | NN/NK/HD/NG | βγδε |
| TAL014/018/023/026 | GGGA | NN/NK/NK/NI | βγδε |
| TAL014/018/023/027 | GGGC | NN/NK/NK/HD | βγδε |
| TAL014/018/023/028 | GGGG | NN/NK/NK/NK | βγδε |
| TAL014/018/023/029 | GGGG | NN/NK/NK/NN | βγδε |
| TAL014/018/023/030 | GGGT | NN/NK/NK/NG | βγδε |
| TAL014/018/024/026 | GGGA | NN/NK/NN/NI | βγδε |
| TAL014/018/024/027 | GGGC | NN/NK/NN/HD | βγδε |
| TAL014/018/024/028 | GGGG | NN/NK/NN/NK | βγδε |
| TAL014/018/024/029 | GGGG | NN/NK/NN/NN | βγδε |
| TAL014/018/024/030 | GGGT | NN/NK/NN/NG | βγδε |
| TAL014/018/025/026 | GGTA | NN/NK/NG/NI | βγδε |
| TAL014/018/025/027 | GGTC | NN/NK/NG/HD | βγδε |
| TAL014/018/025/028 | GGTG | NN/NK/NG/NK | βγδε |
| TAL014/018/025/029 | GGTG | NN/NK/NG/NN | βγδε |
| TAL014/018/025/030 | GGTT | NN/NK/NG/NG | βγδε |
| TAL014/019/021/026 | GGAA | NN/NN/NI/NI | βγδε |
| TAL014/019/021/027 | GGAC | NN/NN/NI/HD | βγδε |
| TAL014/019/021/028 | GGAG | NN/NN/NI/NK | βγδε |
| TAL014/019/021/029 | GGAG | NN/NN/NI/NN | βγδε |
| TAL014/019/021/030 | GGAT | NN/NN/NI/NG | βγδε |
| TAL014/019/022/026 | GGCA | NN/NN/HD/NI | βγδε |
| TAL014/019/022/027 | GGCC | NN/NN/HD/HD | βγδε |
| TAL014/019/022/028 | GGCG | NN/NN/HD/NK | βγδε |
| TAL014/019/022/029 | GGCG | NN/NN/HD/NN | βγδε |
| TAL014/019/022/030 | GGCT | NN/NN/HD/NG | βγδε |
| TAL014/019/023/026 | GGGA | NN/NN/NK/NI | βγδε |
| TAL014/019/023/027 | GGGC | NN/NN/NK/HD | βγδε |
| TAL014/019/023/028 | GGGG | NN/NN/NK/NK | βγδε |
| TAL014/019/023/029 | GGGG | NN/NN/NK/NN | βγδε |
| TAL014/019/023/030 | GGGT | NN/NN/NK/NG | βγδε |
| TAL014/019/024/026 | GGGA | NN/NN/NN/NI | βγδε |
| TAL014/019/024/027 | GGGC | NN/NN/NN/HD | βγδε |
| TAL014/019/024/028 | GGGG | NN/NN/NN/NK | βγδε |
| TAL014/019/024/029 | GGGG | NN/NN/NN/NN | βγδε |
| TAL014/019/024/030 | GGGT | NN/NN/NN/NG | βγδε |
| TAL014/019/025/026 | GGTA | NN/NN/NG/NI | βγδε |
| TAL014/019/025/027 | GGTC | NN/NN/NG/HD | βγδε |

| | | | |
|---|---|---|---|
| TAL014/019/025/028 | GGTG | NN/NN/NG/NK | βγδε |
| TAL014/019/025/029 | GGTG | NN/NN/NG/NN | βγδε |
| TAL014/019/025/030 | GGTT | NN/NN/NG/NG | βγδε |
| TAL014/020/021/026 | GTAA | NN/NG/NI/NI | βγδε |
| TAL014/020/021/027 | GTAC | NN/NG/NI/HD | βγδε |
| TAL014/020/021/028 | GTAG | NN/NG/NI/NK | βγδε |
| TAL014/020/021/029 | GTAG | NN/NG/NI/NN | βγδε |
| TAL014/020/021/030 | GTAT | NN/NG/NI/NG | βγδε |
| TAL014/020/022/026 | GTCA | NN/NG/HD/NI | βγδε |
| TAL014/020/022/027 | GTCC | NN/NG/HD/HD | βγδε |
| TAL014/020/022/028 | GTCG | NN/NG/HD/NK | βγδε |
| TAL014/020/022/029 | GTCG | NN/NG/HD/NN | βγδε |
| TAL014/020/022/030 | GTCT | NN/NG/HD/NG | βγδε |
| TAL014/020/023/026 | GTGA | NN/NG/NK/NI | βγδε |
| TAL014/020/023/027 | GTGC | NN/NG/NK/HD | βγδε |
| TAL014/020/023/028 | GTGG | NN/NG/NK/NK | βγδε |
| TAL014/020/023/029 | GTGG | NN/NG/NK/NN | βγδε |
| TAL014/020/023/030 | GTGT | NN/NG/NK/NG | βγδε |
| TAL014/020/024/026 | GTGA | NN/NG/NN/NI | βγδε |
| TAL014/020/024/027 | GTGC | NN/NG/NN/HD | βγδε |
| TAL014/020/024/028 | GTGG | NN/NG/NN/NK | βγδε |
| TAL014/020/024/029 | GTGG | NN/NG/NN/NN | βγδε |
| TAL014/020/024/030 | GTGT | NN/NG/NN/NG | βγδε |
| TAL014/020/025/026 | GTTA | NN/NG/NG/NI | βγδε |
| TAL014/020/025/027 | GTTC | NN/NG/NG/HD | βγδε |
| TAL014/020/025/028 | GTTG | NN/NG/NG/NK | βγδε |
| TAL014/020/025/029 | GTTG | NN/NG/NG/NN | βγδε |
| TAL014/020/025/030 | GTTT | NN/NG/NG/NG | βγδε |
| TAL015/016/021/026 | TAAA | NG/NI/NI/NI | βγδε |
| TAL015/016/021/027 | TAAC | NG/NI/NI/HD | βγδε |
| TAL015/016/021/028 | TAAG | NG/NI/NI/NK | βγδε |
| TAL015/016/021/029 | TAAG | NG/NI/NI/NN | βγδε |
| TAL015/016/021/030 | TAAT | NG/NI/NI/NG | βγδε |
| TAL015/016/022/026 | TACA | NG/NI/HD/NI | βγδε |
| TAL015/016/022/027 | TACC | NG/NI/HD/HD | βγδε |
| TAL015/016/022/028 | TACG | NG/NI/HD/NK | βγδε |
| TAL015/016/022/029 | TACG | NG/NI/HD/NN | βγδε |
| TAL015/016/022/030 | TACT | NG/NI/HD/NG | βγδε |
| TAL015/016/023/026 | TAGA | NG/NI/NK/NI | βγδε |
| TAL015/016/023/027 | TAGC | NG/NI/NK/HD | βγδε |
| TAL015/016/023/028 | TAGG | NG/NI/NK/NK | βγδε |

| | | | |
|---|---|---|---|
| TAL015/016/023/029 | TAGG | NG/NI/NK/NN | βγδε |
| TAL015/016/023/030 | TAGT | NG/NI/NK/NG | βγδε |
| TAL015/016/024/026 | TAGA | NG/NI/NN/NI | βγδε |
| TAL015/016/024/027 | TAGC | NG/NI/NN/HD | βγδε |
| TAL015/016/024/028 | TAGG | NG/NI/NN/NK | βγδε |
| TAL015/016/024/029 | TAGG | NG/NI/NN/NN | βγδε |
| TAL015/016/024/030 | TAGT | NG/NI/NN/NG | βγδε |
| TAL015/016/025/026 | TATA | NG/NI/NG/NI | βγδε |
| TAL015/016/025/027 | TATC | NG/NI/NG/HD | βγδε |
| TAL015/016/025/028 | TATG | NG/NI/NG/NK | βγδε |
| TAL015/016/025/029 | TATG | NG/NI/NG/NN | βγδε |
| TAL015/016/025/030 | TATT | NG/NI/NG/NG | βγδε |
| TAL015/017/021/026 | TCAA | NG/HD/NI/NI | βγδε |
| TAL015/017/021/027 | TCAC | NG/HD/NI/HD | βγδε |
| TAL015/017/021/028 | TCAG | NG/HD/NI/NK | βγδε |
| TAL015/017/021/029 | TCAG | NG/HD/NI/NN | βγδε |
| TAL015/017/021/030 | TCAT | NG/HD/NI/NG | βγδε |
| TAL015/017/022/026 | TCCA | NG/HD/HD/NI | βγδε |
| TAL015/017/022/027 | TCCC | NG/HD/HD/HD | βγδε |
| TAL015/017/022/028 | TCCG | NG/HD/HD/NK | βγδε |
| TAL015/017/022/029 | TCCG | NG/HD/HD/NN | βγδε |
| TAL015/017/022/030 | TCCT | NG/HD/HD/NG | βγδε |
| TAL015/017/023/026 | TCGA | NG/HD/NK/NI | βγδε |
| TAL015/017/023/027 | TCGC | NG/HD/NK/HD | βγδε |
| TAL015/017/023/028 | TCGG | NG/HD/NK/NK | βγδε |
| TAL015/017/023/029 | TCGG | NG/HD/NK/NN | βγδε |
| TAL015/017/023/030 | TCGT | NG/HD/NK/NG | βγδε |
| TAL015/017/024/026 | TCGA | NG/HD/NN/NI | βγδε |
| TAL015/017/024/027 | TCGC | NG/HD/NN/HD | βγδε |
| TAL015/017/024/028 | TCGG | NG/HD/NN/NK | βγδε |
| TAL015/017/024/029 | TCGG | NG/HD/NN/NN | βγδε |
| TAL015/017/024/030 | TCGT | NG/HD/NN/NG | βγδε |
| TAL015/017/025/026 | TCTA | NG/HD/NG/NI | βγδε |
| TAL015/017/025/027 | TCTC | NG/HD/NG/HD | βγδε |
| TAL015/017/025/028 | TCTG | NG/HD/NG/NK | βγδε |
| TAL015/017/025/029 | TCTG | NG/HD/NG/NN | βγδε |
| TAL015/017/025/030 | TCTT | NG/HD/NG/NG | βγδε |
| TAL015/018/021/026 | TGAA | NG/NK/NI/NI | βγδε |
| TAL015/018/021/027 | TGAC | NG/NK/NI/HD | βγδε |
| TAL015/018/021/028 | TGAG | NG/NK/NI/NK | βγδε |
| TAL015/018/021/029 | TGAG | NG/NK/NI/NN | βγδε |

| | | | |
|---|---|---|---|
| TAL015/018/021/030 | TGAT | NG/NK/NI/NG | βγδε |
| TAL015/018/022/026 | TGCA | NG/NK/HD/NI | βγδε |
| TAL015/018/022/027 | TGCC | NG/NK/HD/HD | βγδε |
| TAL015/018/022/028 | TGCG | NG/NK/HD/NK | βγδε |
| TAL015/018/022/029 | TGCG | NG/NK/HD/NN | βγδε |
| TAL015/018/022/030 | TGCT | NG/NK/HD/NG | βγδε |
| TAL015/018/023/026 | TGGA | NG/NK/NK/NI | βγδε |
| TAL015/018/023/027 | TGGC | NG/NK/NK/HD | βγδε |
| TAL015/018/023/028 | TGGG | NG/NK/NK/NK | βγδε |
| TAL015/018/023/029 | TGGG | NG/NK/NK/NN | βγδε |
| TAL015/018/023/030 | TGGT | NG/NK/NK/NG | βγδε |
| TAL015/018/024/026 | TGGA | NG/NK/NN/NI | βγδε |
| TAL015/018/024/027 | TGGC | NG/NK/NN/HD | βγδε |
| TAL015/018/024/028 | TGGG | NG/NK/NN/NK | βγδε |
| TAL015/018/024/029 | TGGG | NG/NK/NN/NN | βγδε |
| TAL015/018/024/030 | TGGT | NG/NK/NN/NG | βγδε |
| TAL015/018/025/026 | TGTA | NG/NK/NG/NI | βγδε |
| TAL015/018/025/027 | TGTC | NG/NK/NG/HD | βγδε |
| TAL015/018/025/028 | TGTG | NG/NK/NG/NK | βγδε |
| TAL015/018/025/029 | TGTG | NG/NK/NG/NN | βγδε |
| TAL015/018/025/030 | TGTT | NG/NK/NG/NG | βγδε |
| TAL015/019/021/026 | TGAA | NG/NN/NI/NI | βγδε |
| TAL015/019/021/027 | TGAC | NG/NN/NI/HD | βγδε |
| TAL015/019/021/028 | TGAG | NG/NN/NI/NK | βγδε |
| TAL015/019/021/029 | TGAG | NG/NN/NI/NN | βγδε |
| TAL015/019/021/030 | TGAT | NG/NN/NI/NG | βγδε |
| TAL015/019/022/026 | TGCA | NG/NN/HD/NI | βγδε |
| TAL015/019/022/027 | TGCC | NG/NN/HD/HD | βγδε |
| TAL015/019/022/028 | TGCG | NG/NN/HD/NK | βγδε |
| TAL015/019/022/029 | TGCG | NG/NN/HD/NN | βγδε |
| TAL015/019/022/030 | TGCT | NG/NN/HD/NG | βγδε |
| TAL015/019/023/026 | TGGA | NG/NN/NK/NI | βγδε |
| TAL015/019/023/027 | TGGC | NG/NN/NK/HD | βγδε |
| TAL015/019/023/028 | TGGG | NG/NN/NK/NK | βγδε |
| TAL015/019/023/029 | TGGG | NG/NN/NK/NN | βγδε |
| TAL015/019/023/030 | TGGT | NG/NN/NK/NG | βγδε |
| TAL015/019/024/026 | TGGA | NG/NN/NN/NI | βγδε |
| TAL015/019/024/027 | TGGC | NG/NN/NN/HD | βγδε |
| TAL015/019/024/028 | TGGG | NG/NN/NN/NK | βγδε |
| TAL015/019/024/029 | TGGG | NG/NN/NN/NN | βγδε |
| TAL015/019/024/030 | TGGT | NG/NN/NN/NG | βγδε |

| | | | |
|---|---|---|---|
| TAL015/019/025/026 | TGTA | NG/NN/NG/NI | βγδε |
| TAL015/019/025/027 | TGTC | NG/NN/NG/HD | βγδε |
| TAL015/019/025/028 | TGTG | NG/NN/NG/NK | βγδε |
| TAL015/019/025/029 | TGTG | NG/NN/NG/NN | βγδε |
| TAL015/019/025/030 | TGTT | NG/NN/NG/NG | βγδε |
| TAL015/020/021/026 | TTAA | NG/NG/NI/NI | βγδε |
| TAL015/020/021/027 | TTAC | NG/NG/NI/HD | βγδε |
| TAL015/020/021/028 | TTAG | NG/NG/NI/NK | βγδε |
| TAL015/020/021/029 | TTAG | NG/NG/NI/NN | βγδε |
| TAL015/020/021/030 | TTAT | NG/NG/NI/NG | βγδε |
| TAL015/020/022/026 | TTCA | NG/NG/HD/NI | βγδε |
| TAL015/020/022/027 | TTCC | NG/NG/HD/HD | βγδε |
| TAL015/020/022/028 | TTCG | NG/NG/HD/NK | βγδε |
| TAL015/020/022/029 | TTCG | NG/NG/HD/NN | βγδε |
| TAL015/020/022/030 | TTCT | NG/NG/HD/NG | βγδε |
| TAL015/020/023/026 | TTGA | NG/NG/NK/NI | βγδε |
| TAL015/020/023/027 | TTGC | NG/NG/NK/HD | βγδε |
| TAL015/020/023/028 | TTGG | NG/NG/NK/NK | βγδε |
| TAL015/020/023/029 | TTGG | NG/NG/NK/NN | βγδε |
| TAL015/020/023/030 | TTGT | NG/NG/NK/NG | βγδε |
| TAL015/020/024/026 | TTGA | NG/NG/NN/NI | βγδε |
| TAL015/020/024/027 | TTGC | NG/NG/NN/HD | βγδε |
| TAL015/020/024/028 | TTGG | NG/NG/NN/NK | βγδε |
| TAL015/020/024/029 | TTGG | NG/NG/NN/NN | βγδε |
| TAL015/020/024/030 | TTGT | NG/NG/NN/NG | βγδε |
| TAL015/020/025/026 | TTTA | NG/NG/NG/NI | βγδε |
| TAL015/020/025/027 | TTTC | NG/NG/NG/HD | βγδε |
| TAL015/020/025/028 | TTTG | NG/NG/NG/NK | βγδε |
| TAL015/020/025/029 | TTTG | NG/NG/NG/NN | βγδε |
| TAL015/020/025/030 | TTTT | NG/NG/NG/NG | βγδε |
| TAL011/016 | AA | NI/NI | βγ |
| TAL011/017 | AC | NI/HD | βγ |
| TAL011/018 | AG | NI/NK | βγ |
| TAL011/019 | AG | NI/NN | βγ |
| TAL011/020 | AT | NI/NG | βγ |
| TAL012/016 | CA | HD/NI | βγ |
| TAL012/017 | CC | HD/HD | βγ |
| TAL012/018 | CG | HD/NK | βγ |
| TAL012/019 | CG | HD/NN | βγ |
| TAL012/020 | CT | HD/NG | βγ |
| TAL013/016 | GA | NK/NI | βγ |

| | | | |
|---|---|---|---|
| TAL013/017 | GC | NK/HD | βγ |
| TAL013/018 | GG | NK/NK | βγ |
| TAL013/019 | GG | NK/NN | βγ |
| TAL013/020 | GT | NK/NG | βγ |
| TAL014/016 | GA | NN/NI | βγ |
| TAL014/017 | GC | NN/HD | βγ |
| TAL014/018 | GG | NN/NK | βγ |
| TAL014/019 | GG | NN/NN | βγ |
| TAL014/020 | GT | NN/NG | βγ |
| TAL015/016 | TA | NG/NI | βγ |
| TAL015/017 | TC | NG/HD | βγ |
| TAL015/018 | TG | NG/NK | βγ |
| TAL015/019 | TG | NG/NN | βγ |
| TAL015/020 | TT | NG/NG | βγ |
| TAL011/016/021 | AAA | NI/NI/NI | βγδ |
| TAL011/016/022 | AAC | NI/NI/HD | βγδ |
| TAL011/016/023 | AAG | NI/NI/NK | βγδ |
| TAL011/016/024 | AAG | NI/NI/NN | βγδ |
| TAL011/016/025 | AAT | NI/NI/NG | βγδ |
| TAL011/017/021 | ACA | NI/HD/NI | βγδ |
| TAL011/017/022 | ACC | NI/HD/HD | βγδ |
| TAL011/017/023 | ACG | NI/HD/NK | βγδ |
| TAL011/017/024 | ACG | NI/HD/NN | βγδ |
| TAL011/017/025 | ACT | NI/HD/NG | βγδ |
| TAL011/018/021 | AGA | NI/NK/NI | βγδ |
| TAL011/018/022 | AGC | NI/NK/HD | βγδ |
| TAL011/018/023 | AGG | NI/NK/NK | βγδ |
| TAL011/018/024 | AGG | NI/NK/NN | βγδ |
| TAL011/018/025 | AGT | NI/NK/NG | βγδ |
| TAL011/019/021 | AGA | NI/NN/NI | βγδ |
| TAL011/019/022 | AGC | NI/NN/HD | βγδ |
| TAL011/019/023 | AGG | NI/NN/NK | βγδ |
| TAL011/019/024 | AGG | NI/NN/NN | βγδ |
| TAL011/019/025 | AGT | NI/NN/NG | βγδ |
| TAL011/020/021 | ATA | NI/NG/NI | βγδ |
| TAL011/020/022 | ATC | NI/NG/HD | βγδ |
| TAL011/020/023 | ATG | NI/NG/NK | βγδ |
| TAL011/020/024 | ATG | NI/NG/NN | βγδ |
| TAL011/020/025 | ATT | NI/NG/NG | βγδ |
| TAL012/016/021 | CAA | HD/NI/NI | βγδ |
| TAL012/016/022 | CAC | HD/NI/HD | βγδ |

| | | | |
|---|---|---|---|
| TAL012/016/023 | CAG | HD/NI/NK | βγδ |
| TAL012/016/024 | CAG | HD/NI/NN | βγδ |
| TAL012/016/025 | CAT | HD/NI/NG | βγδ |
| TAL012/017/021 | CCA | HD/HD/NI | βγδ |
| TAL012/017/022 | CCC | HD/HD/HD | βγδ |
| TAL012/017/023 | CCG | HD/HD/NK | βγδ |
| TAL012/017/024 | CCG | HD/HD/NN | βγδ |
| TAL012/017/025 | CCT | HD/HD/NG | βγδ |
| TAL012/018/021 | CGA | HD/NK/NI | βγδ |
| TAL012/018/022 | CGC | HD/NK/HD | βγδ |
| TAL012/018/023 | CGG | HD/NK/NK | βγδ |
| TAL012/018/024 | CGG | HD/NK/NN | βγδ |
| TAL012/018/025 | CGT | HD/NK/NG | βγδ |
| TAL012/019/021 | CGA | HD/NN/NI | βγδ |
| TAL012/019/022 | CGC | HD/NN/HD | βγδ |
| TAL012/019/023 | CGG | HD/NN/NK | βγδ |
| TAL012/019/024 | CGG | HD/NN/NN | βγδ |
| TAL012/019/025 | CGT | HD/NN/NG | βγδ |
| TAL012/020/021 | CTA | HD/NG/NI | βγδ |
| TAL012/020/022 | CTC | HD/NG/HD | βγδ |
| TAL012/020/023 | CTG | HD/NG/NK | βγδ |
| TAL012/020/024 | CTG | HD/NG/NN | βγδ |
| TAL012/020/025 | CTT | HD/NG/NG | βγδ |
| TAL013/016/021 | GAA | NK/NI/NI | βγδ |
| TAL013/016/022 | GAC | NK/NI/HD | βγδ |
| TAL013/016/023 | GAG | NK/NI/NK | βγδ |
| TAL013/016/024 | GAG | NK/NI/NN | βγδ |
| TAL013/016/025 | GAT | NK/NI/NG | βγδ |
| TAL013/017/021 | GCA | NK/HD/NI | βγδ |
| TAL013/017/022 | GCC | NK/HD/HD | βγδ |
| TAL013/017/023 | GCG | NK/HD/NK | βγδ |
| TAL013/017/024 | GCG | NK/HD/NN | βγδ |
| TAL013/017/025 | GCT | NK/HD/NG | βγδ |
| TAL013/018/021 | GGA | NK/NK/NI | βγδ |
| TAL013/018/022 | GGC | NK/NK/HD | βγδ |
| TAL013/018/023 | GGG | NK/NK/NK | βγδ |
| TAL013/018/024 | GGG | NK/NK/NN | βγδ |
| TAL013/018/025 | GGT | NK/NK/NG | βγδ |
| TAL013/019/021 | GGA | NK/NN/NI | βγδ |
| TAL013/019/022 | GGC | NK/NN/HD | βγδ |
| TAL013/019/023 | GGG | NK/NN/NK | βγδ |

| | | | |
|---|---|---|---|
| TAL013/019/024 | GGG | NK/NN/NN | βγδ |
| TAL013/019/025 | GGT | NK/NN/NG | βγδ |
| TAL013/020/021 | GTA | NK/NG/NI | βγδ |
| TAL013/020/022 | GTC | NK/NG/HD | βγδ |
| TAL013/020/023 | GTG | NK/NG/NK | βγδ |
| TAL013/020/024 | GTG | NK/NG/NN | βγδ |
| TAL013/020/025 | GTT | NK/NG/NG | βγδ |
| TAL014/016/021 | GAA | NN/NI/NI | βγδ |
| TAL014/016/022 | GAC | NN/NI/HD | βγδ |
| TAL014/016/023 | GAG | NN/NI/NK | βγδ |
| TAL014/016/024 | GAG | NN/NI/NN | βγδ |
| TAL014/016/025 | GAT | NN/NI/NG | βγδ |
| TAL014/017/021 | GCA | NN/HD/NI | βγδ |
| TAL014/017/022 | GCC | NN/HD/HD | βγδ |
| TAL014/017/023 | GCG | NN/HD/NK | βγδ |
| TAL014/017/024 | GCG | NN/HD/NN | βγδ |
| TAL014/017/025 | GCT | NN/HD/NG | βγδ |
| TAL014/018/021 | GGA | NN/NK/NI | βγδ |
| TAL014/018/022 | GGC | NN/NK/HD | βγδ |
| TAL014/018/023 | GGG | NN/NK/NK | βγδ |
| TAL014/018/024 | GGG | NN/NK/NN | βγδ |
| TAL014/018/025 | GGT | NN/NK/NG | βγδ |
| TAL014/019/021 | GGA | NN/NN/NI | βγδ |
| TAL014/019/022 | GGC | NN/NN/HD | βγδ |
| TAL014/019/023 | GGG | NN/NN/NK | βγδ |
| TAL014/019/024 | GGG | NN/NN/NN | βγδ |
| TAL014/019/025 | GGT | NN/NN/NG | βγδ |
| TAL014/020/021 | GTA | NN/NG/NI | βγδ |
| TAL014/020/022 | GTC | NN/NG/HD | βγδ |
| TAL014/020/023 | GTG | NN/NG/NK | βγδ |
| TAL014/020/024 | GTG | NN/NG/NN | βγδ |
| TAL014/020/025 | GTT | NN/NG/NG | βγδ |
| TAL015/016/021 | TAA | NG/NI/NI | βγδ |
| TAL015/016/022 | TAC | NG/NI/HD | βγδ |
| TAL015/016/023 | TAG | NG/NI/NK | βγδ |
| TAL015/016/024 | TAG | NG/NI/NN | βγδ |
| TAL015/016/025 | TAT | NG/NI/NG | βγδ |
| TAL015/017/021 | TCA | NG/HD/NI | βγδ |
| TAL015/017/022 | TCC | NG/HD/HD | βγδ |
| TAL015/017/023 | TCG | NG/HD/NK | βγδ |
| TAL015/017/024 | TCG | NG/HD/NN | βγδ |

| | | | |
|---|---|---|---|
| TAL015/017/025 | TCT | NG/HD/NG | βγδ |
| TAL015/018/021 | TGA | NG/NK/NI | βγδ |
| TAL015/018/022 | TGC | NG/NK/HD | βγδ |
| TAL015/018/023 | TGG | NG/NK/NK | βγδ |
| TAL015/018/024 | TGG | NG/NK/NN | βγδ |
| TAL015/018/025 | TGT | NG/NK/NG | βγδ |
| TAL015/019/021 | TGA | NG/NN/NI | βγδ |
| TAL015/019/022 | TGC | NG/NN/HD | βγδ |
| TAL015/019/023 | TGG | NG/NN/NK | βγδ |
| TAL015/019/024 | TGG | NG/NN/NN | βγδ |
| TAL015/019/025 | TGT | NG/NN/NG | βγδ |
| TAL015/020/021 | TTA | NG/NG/NI | βγδ |
| TAL015/020/022 | TTC | NG/NG/HD | βγδ |
| TAL015/020/023 | TTG | NG/NG/NK | βγδ |
| TAL015/020/024 | TTG | NG/NG/NN | βγδ |
| TAL015/020/025 | TTT | NG/NG/NG | βγδ |
| TAL011/031 | AA | NI/NI | βγ' |
| TAL011/032 | AC | NI/HD | βγ' |
| TAL011/033 | AG | NI/NK | βγ' |
| TAL011/034 | AG | NI/NN | βγ' |
| TAL011/035 | AT | NI/NG | βγ' |
| TAL012/031 | CA | HD/NI | βγ' |
| TAL012/032 | CC | HD/HD | βγ' |
| TAL012/033 | CG | HD/NK | βγ' |
| TAL012/034 | CG | HD/NN | βγ' |
| TAL012/035 | CT | HD/NG | βγ' |
| TAL013/031 | GA | NK/NI | βγ' |
| TAL013/032 | GC | NK/HD | βγ' |
| TAL013/033 | GG | NK/NK | βγ' |
| TAL013/034 | GG | NK/NN | βγ' |
| TAL013/035 | GT | NK/NG | βγ' |
| TAL014/031 | GA | NN/NI | βγ' |
| TAL014/032 | GC | NN/HD | βγ' |
| TAL014/033 | GG | NN/NK | βγ' |
| TAL014/034 | GG | NN/NN | βγ' |
| TAL014/035 | GT | NN/NG | βγ' |
| TAL015/031 | TA | NG/NI | βγ' |
| TAL015/032 | TC | NG/HD | βγ' |
| TAL015/033 | TG | NG/NK | βγ' |
| TAL015/034 | TG | NG/NN | βγ' |
| TAL015/035 | TT | NG/NG | βγ' |

| | | | |
|---|---|---|---|
| TAL021/036 | AA | NI/NI | δε' |
| TAL021/037 | AC | NI/HD | δε' |
| TAL021/038 | AG | NI/NK | δε' |
| TAL021/039 | AG | NI/NN | δε' |
| TAL021/040 | AT | NI/NG | δε' |
| TAL022/036 | CA | HD/NI | δε' |
| TAL022/037 | CC | HD/HD | δε' |
| TAL022/038 | CG | HD/NK | δε' |
| TAL022/039 | CG | HD/NN | δε' |
| TAL022/040 | CT | HD/NG | δε' |
| TAL023/036 | GA | NK/NI | δε' |
| TAL023/037 | GC | NK/HD | δε' |
| TAL023/038 | GG | NK/NK | δε' |
| TAL023/039 | GG | NK/NN | δε' |
| TAL023/040 | GT | NK/NG | δε' |
| TAL024/036 | GA | NN/NI | δε' |
| TAL024/037 | GC | NN/HD | δε' |
| TAL024/038 | GG | NN/NK | δε' |
| TAL024/039 | GG | NN/NN | δε' |
| TAL024/040 • | GT | NN/NG | δε' |
| TAL025/036 | TA | NG/NI | δε' |
| TAL025/037 | TC | NG/HD | δε' |
| TAL025/038 | TG | NG/NK | δε' |
| TAL025/039 | TG | NG/NN | δε' |
| TAL025/040 | TT | NG/NG | δε' |
| TAL011 | A | NI | β |
| TAL012 | C | HD | β |
| TAL013 | G | NK | β |
| TAL014 | G | NN | β |
| TAL015 | T | NG | β |

To prepare DNA fragments encoding α units for use in assembly, 20 rounds of PCR were performed with each α unit plasmid as a template using primers oJS2581 (5'-Biotin–TCTAGAGAAGACAAGAACCTGACC–3' (SEQ ID NO:237)) and oJS2582 (5'-GGATCCGGTCTCTTAAGGCCGTGG–3' (SEQ ID NO:238)). The resulting PCR products were biotinylated on the 5' end. Each α PCR product was then digested with 40 units of BsaI-HF restriction enzyme to generate 4bp overhangs, purified using the

QIAquick PCR purification kit (QIAGEN) according to manufacturer's instructions except that the final product was eluted in 50 µl of 0.1X EB.

To prepare DNA fragments encoding β, βγδε, βγδ, βγ, βγ*, and δε* repeats, 10 µg of each of these plasmids was digested with 50 units of BbsI restriction enzyme in NEBuffer 2 for 2 hours at 37 °C followed by serial restriction digests performed in NEBuffer 4 at 37 °C using 100 units each of XbaI, BamHI-HF, and SalI-HF enzymes that were added at 5 minute intervals. The latter set of restriction digestions were designed to cleave the plasmid backbone to ensure that this larger DNA fragment does not interfere with subsequent ligations performed during the assembly process. These restriction digest reactions were then purified using the QIAquick PCR purification kit (QIAGEN) according to manufacturer's instructions except that the final product was eluted in 180 µl of 0.1X EB.

All assembly steps were performed using a Sciclone G3 liquid handling workstation (Caliper) in 96-well plates and using a SPRIplate 96-ring magnet (Beckman Coulter Genomics) and a DynaMag-96 Side magnet (Life Technologies). In the first assembly step, a biotinylated α unit fragment was ligated to the first βγδε fragment and then the resulting αβγδε fragments are bound to Dynabeads MyOne C1 streptavidin-coated magnetic beads (Life Technologies) in 2X B&W Buffer (Life Technologies). Beads were then drawn to the side of the well by placing the plate on the magnet and then washed with 100 µl B&W buffer with 0.005% Tween 20™ (Sigma) and again with 100 µl 0.1 mg/ml bovine serum albumin (BSA) (New England Biolabs). Additional βγδε fragments were ligated by removing the plate from the magnet, resuspending the beads in solution in each well, digesting the bead bound fragment with BsaI-HF restriction enzyme, placing the plate on the magnet, washing with 100 µl B&W/Tween20™ followed by 100µl of 0.1 mg/ml BSA, and then ligating the next fragment. This process was repeated multiple times with additional βγδε units to extend the bead-bound fragment. The last fragment to be ligated was always a β, βγ*, βγδ, or δε* unit to enable cloning of the full-length fragment into expression vectors (note that fragments that end with a δε* unit are always preceded by ligation of a βγ unit).

The final full-length bead-bound fragment was digested with 40 units of BsaI-HF restriction enzyme followed by 25 units of BbsI restriction enzyme (New England Biolabs). Digestion with BbsI releaseedthe fragment from the beads and generated a unique 5' overhang for cloning of the fragment. Digestion with BsaI-HF resulted in creation of a unique 3' overhang for cloning.

DNA fragments encoding the assembled TALE repeat arrays were subcloned into one of four TALEN expression vectors. Each of these vectors included a CMV promoter, a translational start codon optimized for mammalian cell expression, a triple FLAG epitope tag, a nuclear localization signal, amino acids 153 to 288 from the TALE 13 protein (Miller et al., 2011, Nat. Biotechnol., 29:143-148), two unique and closely positioned Type IIS BsmBI restriction sites, a 0.5 TALE repeat domain encoding one of four possible RVDs (NI, HD, NN, or NG for recognition of an A, C, G, or T nucleotide, respectively), amino acids 715 to 777 from the TALE 13 protein, and the wild-type FokI cleavage domain. All DNA fragments possessed overhangs that enable directional cloning into any of the four TALEN expression vectors that has been digested with BsmBI.

To prepare a TALEN expression vector for subcloning, 5 µg of plasmid DNA were digested with 50 units of BsmBI restriction enzyme (New England Biolabs) in NEBuffer 3 for 8 hours at 55 degrees C. Digested DNA was purified using 90 µl of Ampure™ XP beads (Agencourt) according to manufacturer's instructions and diluted to a final concentration of 5ng/µl in 1 mM TrisHCl. The assembled TALE repeat arrays were ligated into TALEN expression vectors using 400 U of T4 DNA Ligase (New England Biolabs). Ligation products were transformed into chemically competent XL-1 Blue cells. Six colonies were picked for each ligation and plasmid DNA isolated by an alkaline lysis miniprep procedure. Simultaneously, the same six colonies were screened by PCR using primers oSQT34 (5'-GACGGTGGCTGTCAAATACCAAGATATG-3' (SEQ ID NO:239)) and oSQT35 (5'- TCTCCTCCAGTTCACTTTTGACTAGTTGGG - 3' (SEQ ID NO:240)). PCR products were analyzed on a QIAxcel capillary electrophoresis system (Qiagen). Miniprep DNA from clones that contained correctly sized PCR products were sent for DNA sequence confirmation with primers oSQT1 (5'-

AGTAACAGCGGTAGAGGCAG-3' (SEQ ID NO:241)), oSQT3 (5'-ATTGGGCTACGATGGACTCC-3' (SEQ ID NO:242)), and oJS2980 (5-TTAATTCAATATATTCATGAGGCAC-3' (SEQ ID NO:243)).

Because the final fragment ligated can encode one, two, or three TALE repeats, the methods disclosed herein can be used to assemble arrays consisting of any desired number of TALE repeats. Assembled DNA fragments encoding the final full-length TALE repeat array are released from the beads by restriction enzyme digestion and can be directly cloned into a desired expression vector of choice.

The methods can be efficiently practiced in 96-well format using a robotic liquid handling workstation. With automation, DNA fragments encoding 96 different TALE repeat arrays of variable lengths can be assembled in less than one day. Medium-throughput assembly of fragments can be performed in one to two days using multi-channel pipets and 96-well plates. Fragments assembled using either approach can then be cloned into expression vectors (e.g., for expression as a TALEN) to generate sequence-verified plasmids in less than one week. Using the automated assembly approach, sequence-verified TALE repeat array expression plasmids can be made quickly and inexpensively.

Example 6. Large-scale Testing of Assembled TALENs Using a Human Cell-based Reporter Assay

To perform a large-scale test of the robustness of TALENs for genome editing in human cells, the method described in Example 5 was used to construct a series of plasmids encoding 48 TALEN pairs targeted to different sites scattered throughout the EGFP reporter gene. Monomers in each of the TALEN pairs contained the same number of repeats (ranging from 8.5 to 19.5 in number), and these pairs were targeted to sites possessing a fixed length "spacer" sequence (16 bps) between the "half-sites" bound by each TALEN monomer (Table 6).

**Table 6. EGFP reporter gene sequences targeted by 48 pairs of TALENs**

| TALEN pair # | Position within EGFP of the first nucleotide in the binding site | Target site (half-sites in CAPS, spacer in lowercase) | SEQ ID NO: | # of repeat domains in Left TALEN | # of repeat domains in Right TALEN |
|---|---|---|---|---|---|
| 1 | -8 | TCGCCACCATggtgagcaaggg cgagGAGCTGTTCA | 93 | 8.5 | 8.5 |
| 2 | 35 | TGGTGCCCATcctggtcgagct ggacGGCGACGTAA | 94 | 8.5 | 8.5 |
| 3 | 143 | TCTGCACCACcggcaagctgcc cgtgCCCTGGCCCA | 95 | 8.5 | 8.5 |
| 4 | 425 | TGGAGTACAActacaacagcca caacGTCTATATCA | 96 | 8.5 | 8.5 |
| 5 | 82 | TTCAGCGTGTCcggcgagggcg agggcGATGCCACCTA | 97 | 9.5 | 9.5 |
| 6 | 111 | TGCCACCTACGgcaagctgacc ctgaaGTTCATCTGCA | 98 | 9.5 | 9.5 |
| 7 | 172 | TGGCCCACCCTcgtgaccaccc tgaccTACGGCGTGCA | 99 | 9.5 | 9.5 |
| 8 | 496 | TTCAAGATCCGccacaacatcg aggacGGCAGCGTGCA | 100 | 9.5 | 9.5 |
| 9 | -23 | TAGAGGATCCACcggtcgccac catggtGAGCAAGGGCGA | 101 | 10.5 | 10.5 |
| 10 | 91 | TCCGGCGAGGGCgagggcgatg ccacctACGGCAAGCTGA | 102 | 10.5 | 10.5 |
| 11 | 194 | TGACCTACGGCGtgcagtgctt cagccgCTACCCCGACCA | 103 | 10.5 | 10.5 |
| 12 | 503 | TCCGCCACAACAtcgaggacgg cagcgtGCAGCTCGCCGA | 104 | 10.5 | 10.5 |
| 13 | 44 | TCCTGGTCGAGCTggacggcga cgtaaacGGCCACAAGTTCA | 105 | 11.5 | 11.5 |
| 14 | 215 | TCAGCCGCTACCCcgaccacat gaagcagCACGACTTCTTCA | 106 | 11.5 | 11.5 |
| 15 | 251 | TCTTCAAGTCCGCcatgcccga aggctacGTCCAGGAGCGCA | 107 | 11.5 | 11.5 |
| 16 | 392 | TCAAGGAGGACGGcaacatcct ggggcacAAGCTGGAGTACA | 108 | 11.5 | 11.5 |
| 17 | 485 | TCAAGGTGAACTTcaagatccg ccacaacATCGAGGACGGCA | 109 | 11.5 | 11.5 |
| 18 | -16 | TCCACCGGTCGCCAccatggtg agcaagggCGAGGAGCTGTTCA | 110 | 12.5 | 12.5 |
| 19 | 82 | TTCAGCGTGTCCGGcgagggcg agggcgatGCCACCTACGGCAA | 111 | 12.5 | 12.5 |
| 20 | 214 | TTCAGCCGCTACCCcgaccaca tgaagcagCACGACTTCTTCAA | 112 | 12.5 | 12.5 |
| 21 | 436 | TACAACAGCCACAAcgtctata tcatggccGACAAGCAGAAGAA | 113 | 12.5 | 12.5 |

| | | | | | |
|---|---|---|---|---|---|
| 22 | 35 | TGGTGCCCATCCTGGtcgagctggacggcgaCGTAAACGGCCACAA | 114 | 13.5 | 13.5 |
| 23 | 266 | TGCCCGAAGGCTACGtccaggagcgcaccatCTTCTTCAAGGACGA | 115 | 13.5 | 13.5 |
| 24 | 362 | TGAACCGCATCGAGCtgaaggcatcgacttCAAGGAGGACGGCAA | 116 | 13.5 | 13.5 |
| 25 | 497 | TCAAGATCCGCCACAacatcgaggacggcagCGTGCAGCTCGCCGA | 117 | 13.5 | 13.5 |
| 26 | 23 | TGTTCACCGGGGTGGTgcccatcctggtcgagCTGGACGGCGACGTAA | 118 | 14.5 | 14.5 |
| 27 | 38 | TGCCCATCCTGGTCGAgctggacggcgacgtaAACGGCCACAAGTTCA | 119 | 14.5 | 14.5 |
| 28 | 89 | TGTCCGGCGAGGGCGAgggcgatgccacctacGGCAAGCTGACCCTGA | 120 | 14.5 | 14.5 |
| 29 | 140 | TCATCTGCACCACCGGcaagctgcccgtgcccTGGCCCACCCTCGTGA | 121 | 14.5 | 14.5 |
| 30 | 452 | TCTATATCATGGCCGAcaagcagaagaacggcATCAAGGTGAACTTCA | 122 | 14.5 | 14.5 |
| 31 | 199 | TACGGCGTGCAGTGCTTcagccgctaccccgacCACATGAAGCAGCACGA | 123 | 15.5 | 15.5 |
| 32 | 223 | TACCCCGACCACATGAAgcagcacgacttcttcAAGTCCGCCATGCCCGA | 124 | 15.5 | 15.5 |
| 33 | 259 | TCCGCCATGCCCGAAGGctacgtccaggagcgcACCATCTTCTTCAAGGA | 125 | 15.5 | 15.5 |
| 34 | 391 | TTCAAGGAGGACGGCAAcatcctggggcacaagCTGGAGTACAACTACAA | 126 | 15.5 | 15.5 |
| 35 | 430 | TACAACTACAACAGCCAcaacgtctatatcatgGCCGACAAGCAGAAGAA | 127 | 15.5 | 15.5 |
| 36 | 26 | TCACCGGGGTGGTGCCCAtcctggtcgagctggaCGGCGACGTAAACGGCCA | 128 | 16.5 | 16.5 |
| 37 | 68 | TAAACGGCCACAAGTTCAgcgtgtccggcgagggCGAGGGCGATGCCACCTA | 129 | 16.5 | 16.5 |
| 38 | 206 | TGCAGTGCTTCAGCCGCTaccccgaccacatgaaGCAGCACGACTTCTTCAA | 130 | 16.5 | 16.5 |
| 39 | 83 | TCAGCGTGTCCGGCGAGGGcgagggcgatgccaccTACGGCAAGCTGACCCTGA | 131 | 17.5 | 17.5 |

| | | | | | |
|---|---|---|---|---|---|
| 40 | 134 | TGAAGTTCATCTGCACCACcgg caagctgcccgtgCCCTGGCCC ACCCTCGTGA | 132 | 17.5 | 17.5 |
| 41 | 182 | TCGTGACCACCCTGACCTAcgg cgtgcagtgcttcAGCCGCTAC CCCGACCACA | 133 | 17.5 | 17.5 |
| 42 | 458 | TCATGGCCGACAAGCAGAAgaa cggcatcaaggtgAACTTCAAG ATCCGCCACA | 134 | 17.5 | 17.5 |
| 43 | 25 | TTCACCGGGGTGGTGCCCATcc tggtcgagctggacGGCGACGT AAACGGCCACAA | 135 | 18.5 | 18.5 |
| 44 | 145 | TGCACCACCGGCAAGCTGCCcg tgccctggcccaccCTCGTGAC CACCCTGACCTA | 136 | 18.5 | 18.5 |
| 45 | 253 | TTCAAGTCCGCCATGCCCGAag gctacgtccaggagCGCACCAT CTTCTTCAAGGA | 137 | 18.5 | 18.5 |
| 46 | 454 | TATATCATGGCCGACAAGCAga agaacggcatcaagGTGAACTT CAAGATCCGCCA | 138 | 18.5 | 18.5 |
| 47 | 139 | TTCATCTGCACCACCGGCAAGc tgcccgtgccctggcCCACCCT CGTGACCACCCTGA | 139 | 19.5 | 19.5 |
| 48 | 338 | TGAAGTTCGAGGGCGACACCCt ggtgaaccgcatcgaGCTGAAG GGCATCGACTTCAA | 140 | 19.5 | 19.5 |

Each of the 48 TALEN pairs was tested in human cells for its ability to disrupt the coding sequence of a chromosomally integrated EGFP reporter gene. In this assay, NHEJ-mediated repair of TALEN-induced breaks within the EGFP coding sequence led to loss of EGFP expression, which was quantitatively assessed using flow cytometry 2 and 5 days following transfection. (To ensure that activities of each active TALEN pair could be detected, we only targeted sites located at or upstream of nucleotide position 503 in the gene, a position we had previously shown would disrupt EGFP function when mutated with a zinc finger nuclease (ZFN) (Maeder et al., 2008, Mol. Cell 31:294-301).) Strikingly, all 48 TALEN pairs showed significant EGFP gene-disruption activities in this assay (FIG. 19A). The net percentage of EGFP-disrupted cells induced by TALENs on day 2 post-transfection ranged from 9.4% to 68.0%, levels comparable to the percentage disruption observed with four EGFP-targeted ZFN pairs originally made by the Oligomerized Pool Engineering (OPEN) method (Fig. 19A). These results demonstrate that TALENs containing as few as 8.5 TALE repeats possess significant nuclease

activities and provide a large-scale demonstration of the robustness of TALENs in human cells.

Interestingly, re-quantification of the percentage of EGFP-negative cells at day 5 post-transfection revealed that cells expressing shorter-length TALENs (such as those composed of 8.5 to 10.5 repeats) showed significant reductions in the percentage of EGFP-disrupted cells whereas those expressing longer TALENs did not (FIGs. 19A-B and 20A). One potential explanation for this effect is cellular toxicity associated with expression of shorter-length TALENs. Consistent with this hypothesis, in cells transfected with plasmids encoding shorter-length TALENs, greater reductions in the percentage of tdTomato-positive cells were observed from day 2 to day 5 post-transfection (FIG. 20D) (a tdTomato-encoding plasmid was co-transfected together with the TALEN expression plasmids on day 0). Taken together, our results suggest that although shorter-length TALENs are as active as longer-length TALENs, the former can cause greater cytotoxicity in human cells.

Our EGFP experiments also provided an opportunity to assess four of five computationally-derived design guidelines (Cermak et al., 2011, Nucleic Acids Res., 39:e82). The guidelines proposed by Cermak are as follows:

1. The nucleotide just 5' to the first nucleotide of the half-site should be a thymine.

2. The first nucleotide of the half-site should not be a thymine.

3. The second nucleotide of the half-site should not be an adenosine.

4. The 3' most nucleotide in the target half-site should be a thymine.

5. The composition of each nucleotide within the target half-site should not vary from the observed percentage composition of naturally occurring binding sites by more than 2 standard deviations. The percentage composition of all naturally occurring TALE binding sites is: A = 31±16%, C = 37±13%, G = 9±8%, T = 22±10%. Hence, the nucleotide composition of potential TALE binding sites should be: A = 0% to 63%, C = 11% to 63%, G = 0% to 25% and T = 2% to 42%.

These guidelines have been implemented in the TALE-NT webserver (boglabx.plp.iastate.edu/TALENT/TALENT/) to assist users in identifying potential

TALEN target sites. All 48 of the sequences we targeted in EGFP did not meet one or more of these guidelines (however, note that all of our sites did meet the requirement for a 5' T). The ~100% success rate observed for these 48 sites demonstrates that TALENs can be readily obtained for target sequences that do not follow these guidelines. In addition, for each of the four design guidelines, we did not find any statistically significant correlation between guideline violation and the level of TALEN-induced mutagenesis on either day 2 or day 5 post-transfection. We also failed to find a significant correlation between the total number of guideline violations and the level of mutagenic TALEN activity. Thus, our results show that failure to meet four of the five previously described design guidelines when identifying potential TALEN target sites does not appear to adversely affect success rates or nuclease efficiencies.

Example 7. High-throughput Alteration of Endogenous Human Genes Using Assembled TALENs

Having established the robustness of the TALEN platform with a chromosomally integrated reporter gene, it was next determined whether this high success rate would also be observed with endogenous genes in human cells. To test this, the assembly method described in Example 5 was used to engineer TALEN pairs targeted to 96 different human genes: 78 genes implicated in human cancer (Vogelstein and Kinzler, 2004, Nat. Med., 10:789-799) and 18 genes involved in epigenetic regulation of gene expression (Table 7). For each gene, a TALEN pair was designed to cleave near the amino-terminal end of the protein coding sequence, although in a small number of cases the presence of repetitive sequences led us to target alternate sites in neighboring downstream exons or introns (Table 7). Guided by the results with the EGFP TALENs, TALENs composed of 14.5, 15.5, or 16.5 repeats were constructed that cleaved sites with 16, 17, 18, 19 or 21 bp spacer sequences. All of the target sites had a T at the 5' end of each half-site.

**Table 7. Endogenous human gene sequences targeted by 96 pairs of TALENs**

| Target gene name | % NHEJ | Target site (half-sites in CAPS, spacer in lowercase, ATG underlined) | SEQ ID NO: | Length of LEFT half site (include 5' T) | Length of spacer | Length of RIGHT half site (include 5' T) | Gene Type |
|---|---|---|---|---|---|---|---|
| ABL1 | 22.5 ± 7.1 | TACCTATTATTACT TTATggggcagcagcctgg aaAAGTACTTGGGG ACCAA | 141. | 16.5 | 17 | 15.5 | Cancer |
| AKT2 | 14.1 ± 7.3 | TGTGTCTTGGGATG AGTGggtcagtgttctggtg CTCACAGGATGGCT GGCA | 142. | 16.5 | 16 | 16.5 | Cancer |
| ALK | 12.7 ± 2.9 | TCCTGTGGCTCCTG CCGCtgctgctttccacggc AGCTGTGGGCTCCG GGA | 143. | 16.5 | 16 | 15.5 | Cancer |
| APC | 48.8 ± 9.8 | TATGTACGCCTCCC TGGGctcgggtccggtcgcc CCTTTGCCCGCTTC TGTA | 144. | 16.5 | 16 | 16.5 | Cancer |
| ATM | 35.5 ± 15.6 | TGAATTGGGATGCT GTTTttaggtattctattcaaa TTTATTTTACTGTCT TTA | 145. | 16.5 | 18 | 16.5 | Cancer |
| AXIN2 | 2.5 ± 0.6 | TCCCTCACCATGAG TAGCgctatgttggtgacttG CCTCCCGGACCCCA GCA | 146. | 16.5 | 16 | 16.5 | Cancer |
| BAX | 14.7 ± 11.6 | TGTGCGATCTCCAA GCACtgaggggcagaaact cCCGGATCGGGCGC TGCCA | 147. | 16.5 | 16 | 16.5 | Cancer |
| BCL6 | 14.9 ± 5.9 | TTTTCAAGTGAAGA CAAAatggcctcgccggct gACAGCTGTATCCA GTTCA | 148. | 16.5 | 16 | 16.5 | Cancer |
| BMPR1A | 50.4 ± 16.4 | TACAATTGAACAAT GCCTcagctatacatttacat CAGATTATTGGGAG CCTA | 149. | 16.5 | 17 | 16.5 | Cancer |
| BRCA1 | 44.5 ± 15.5 | TCCGAAGCTGACAG ATGGgtattctttgacgggg GGTAGGGGCGGAA CCTGA | 150. | 16.5 | 16 | 16.5 | Cancer |
| BRCA2 | 41.6 ± 10.5 | TTAGACTTAGGTAA GTAAtgcaatatggtagact GGGGAGAACTACA AACTA | 151. | 16.5 | 16 | 16.5 | Cancer |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CBX3 | 35.2 ± 22.6 | TCTGCAATAAAAAA TGGCctccaacaaaactaca TTGGTAAGTTAATG AAAA | 152. | 16.5 | 16 | 16.5 | Epigen etic |
| CBX8 | 13.5 ± 3.4 | TGGAGCTTTCAGCG GTGGgggagcgggtgttcg cgGCCGAAGCCCTC CTGAA | 153. | 16.5 | 17 | 15.5 | Epigen etic |
| CCND1 | 40.5 ± 2.2 | TGGAACACCAGCTC CTGTgctgcgaagtggaaac catCCGCCGCGCGTA CCCCGA | 154. | 16.5 | 19 | 16.5 | Cancer |
| CDC73 | 36.3 ± 7.7 | TGCTTAGCGTCCTG CGACagtacaacatccagaa GAAGGAGATTGTG GTGAA | 155. | 16.5 | 16 | 16.5 | Cancer |
| CDH1 | none | TGCTGCAGGTACCC CGGAtcccctgacttgcgag GGACGCATTCGGGC CGCA | 156. | 16.5 | 16 | 16.5 | Cancer |
| CDK4 | 21.5 ± 17.4 | TCCCTTGATCTGAG AAtggctacctctcgataTG AGCCAGTGGCTGA AA | 157. | 14.5 | 16 | 15.5 | Cancer |
| CHD4 | 9.6 ± 0.1 | TGGCGTCGGGCCTG GGCtccccgtcccccctgctc GGCGGGCAGTGAG GAGGA | 158. | 15.5 | 17 | 16.5 | Epigen etic |
| CHD7 | 11.4 ± 2.7 | TGTGTTGGAAGAAG ATGGcagatccaggaatgat GAGTCTTTTTGGCG AGGA | 159. | 16.5 | 16 | 16.5 | Epigen etic |
| CTNNB 1 | 26.0 ± 8.1 | TCCAGCGTGGACAA TGGctactcaaggtttgtgTC ATTAAATCTTTAGT TA | 160. | 15.5 | 16 | 16.5 | Cancer |
| CYLD | 24.7 ± 2.3 | TAATATCACAATGA GTTCaggcttatggagccaa gaAAAAGTCACTTC ACCCTA | 161. | 16.5 | 18 | 16.5 | Cancer |
| DDB2 | 15.8 ± 7.2 | TCACACGGAGGAC GCGatggctcccaagaaac GCCCAGAAACCCA GAAGA | 162. | 14.5 | 16 | 16.5 | Cancer |
| ERCC2 | 55.8 ± 12.7 | TCCGGCCGGCGCCA TGAagtgagaaggggggctg GGGGTCGCGCTCGC TA | 163. | 15.5 | 16 | 14.5 | Cancer |
| ERCC5 | none | TCCGGGATCGCCAT GGGAactcaatagaaaatcc tcaTCTTCTCACTTTG TTTCA | 164. | 16.5 | 19 | 16.5 | Cancer |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| EWSR1 | 14.3 ± 8.2 | TGGCGTCCACGGGTGAGTatggtggaactgcggtcGCGCCGGCGGTAGCCGGA | 165. | 16.5 | 17 | 16.5 | Cancer |
| EXT1 | 9.5 ± 3.0 | TGACCCAGGCAGGACACAtgcaggccaaaaaacgcTATTTCATCCTGCTCTCA | 166. | 16.5 | 17 | 16.5 | Cancer |
| EXT2 | none | TTCCTCCCAGGGGGATGTcctgcgcctcagggtcCGGTGGTGGCCTGCGGCA | 167. | 16.5 | 16 | 16.5 | Cancer |
| EZH2 | 41.3 ± 2.6 | TGCTTTTAGAATAATCATgggccagactgggaagAAATCTGAGAAGGGACCA | 168. | 16.5 | 16 | 16.5 | Epigenetic |
| FANCA | 9.7 ± 5.0 | TAGGCGCCAAGGCCATGTccgactcgtgggtcccGAACTCCGCCTCGGGCCA | 169. | 16.5 | 16 | 16.5 | Cancer |
| FANCC | 23.7 ± 17.8 | TGAAGGGACATCACCTTTtcgcttttccaagatgGCTCAAGATTCAGTAGA | 170. | 16.5 | 17 | 15.5 | Cancer |
| FANCE | none | TGCCCCGGCATGGCGACAccggacgcggggctcccTGGGGCTGAGGGCGTGGA | 171. | 16.5 | 17 | 16.5 | Cancer |
| FANCF | 46.0 ± 7.7 | TTCGCGCACCTCATGGaatcccttctgcagcaCCTGGATCGCTTTTCCGA | 172. | 14.5 | 16 | 16.5 | Cancer |
| FANCG | 26.9 ± 16.2 | TCGGCCACCATGTCCCgccagaccacctctgtGGGCTCCAGCTGCCTGGA | 173. | 14.5 | 16 | 16.5 | Cancer |
| FES | 12.6 ± 10.6 | TCCCCAGAACAGCACTATgggcttctcttccgagctGTGCAGCCCCCAGGGCCA | 174. | 16.5 | 18 | 16.5 | Cancer |
| FGFR1 | 17.4 ± 6.2 | TCTGCTCCCCACCGAGGAcctctgcatgcaggcaTGAATCCCAGGAGCCTA | 175. | 16.5 | 16 | 15.5 | Cancer |
| FH | 20.9 ± 11.8 | TGTACCGAGCACTTCGGCtcctcgcgcgctcgcgtCCCCTCGTGCGGGCTCCA | 176. | 16.5 | 17 | 16.5 | Cancer |
| FLCN | 11.1 ± 4.4 | TCTCCAAGGCACCATGAAtgccatcgtggctctctgCCACTTCTGCGAGCTCCA | 177. | 16.5 | 18 | 16.5 | Cancer |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| FLT3 | none | TCCGGAGGCCATGCCGGCgttggcgcgcgacggcggccaGCTGCCGCTGCTCGGTA | 178. | 16.5 | 21 | 15.5 | Cancer |
| FLT4 | 9.9 ± 5.0 | TGCAGCGGGGCGCCGCGCtgtgcctgcgactggctCTGCCTGGGACTCCTGGA | 179. | 16.5 | 19 | 16.5 | Cancer |
| FOXO1 | 8.5 ± 1.1 | TCACCATGGCCGAGGCGcctcaggtggtggagaTCGACCCGGACTTCGA | 180. | 15.5 | 16 | 14.5 | Cancer |
| FOXO3 | 7.3 ± 2.3 | TCTCCGCTCGAAGTGGAGctggacccggagttcgagCCCCAGAGCCGTCCGCGA | 181. | 16.5 | 18 | 16.5 | Cancer |
| GLI1 | 21.5 ± 12.4 | TCCTCTGAGACGCCATGTtcaactcgatgaccccACCACCAATCAGTAGCTA | 182. | 16.5 | 16 | 16.5 | Cancer |
| HDAC1 | 10.8 ± 3.0 | TGGCGCAGACGCAGGGCacccggaggaaagtctgTTACTACTACGACGGTGA | 183. | 15.5 | 17 | 16.5 | Epigenetic |
| HDAC2 | 4.2 ± 0.9 | TGCGCTCACCTCCCTGCGgcctcctgaggtggtttgGTGGCCCCCTCCTCGCGA | 184. | 16.5 | 18 | 16.5 | Epigenetic |
| HDAC6 | 21.4 ± 2.1 | TCCTCAACTATGACCTCAaccggccaggattccaCCACAACCAGGCAGCGAA | 185. | 16.5 | 16 | 16.5 | Epigenetic |
| HMGA2 | 3.0 ± 1.5 | TGAGCGCACGCGGTGAGGgcgcggggcagccgtcCACTTCAGCCCAGGGACA | 186. | 16.5 | 16 | 16.5 | Cancer |
| HOXA13 | 7.6 ± 3.1 | TCCGTGCTCCTCCACCCCcgctggatcgagcccacCGTCATGTTTCTCTACGA | 187. | 16.5 | 17 | 16.5 | Cancer |
| HOXA9 | 6.4 ± 2.7 | TGGGCACGGTGATGGCcaccactggggccctgGGCAACTACTACGTGGA | 188. | 14.5 | 16 | 15.5 | Cancer |
| HOXC13 | 10.5 ± 0.3 | TCCAGCAGATCATGTCATgacgacttcgctgctcctGCATCCACGCTGGCCGGA | 189. | 16.5 | 18 | 16.5 | Cancer |
| HOXD11 | none | TTGACGAGTGCGGCCAGagcgcagccagcatgtaCCTGCCGGGCTGCGCCTA | 190. | 15.5 | 17 | 16.5 | Cancer |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HOXD13 | none | TGCGGGCAGACGGCGGGGgcgccggtggcgccccgGCCTCTTCCTCCTCCTCA | 191. | 16.5 | 17 | 16.5 | Cancer |
| JAK2 | 44.9 ± 16.9 | TCTGAAAAAGACTCTGCAtgggaatggcctgcctTACGATGACAGAAATGGA | 192. | 16.5 | 16 | 16.5 | Cancer |
| KIT | none | TACCGCGATGAGAGGCGCtcgcggcgcctggattttCTCTGCGTTCTGCTCCTA | 193. | 16.5 | 19 | 16.5 | Cancer |
| KRAS | 9.4 ± 0.9 | TGAAAATGACTGAATATAaacttgtggtagttggaGCTGGTGGCGTAGGCAA | 194. | 16.5 | 17 | 15.5 | Cancer |
| MAP2K4 | 11.9 ± 7.1 | TAGGGTCCCCGGCGCCAGgccaccccggccgtcagCAGCATGCAGGGTAAGGA | 195. | 16.5 | 16 | 16.5 | Cancer |
| MDM2 | 33.0 ± 20.2 | TCCAAGCGCGAAAACCCCggatggtgaggagcaggTACTGGCCCGGCAGCGA | 196. | 16.5 | 17 | 15.5 | Cancer |
| MET | 40.4 ± 10.7 | TTATTATTACATGGCTTTgccttactgaggcttcATCTTGTCCTCTGGTCCA | 197. | 16.5 | 16 | 16.5 | Cancer |
| MLH1 | 44.9 ± 6.3 | TCTGGCGCCAAAATGTCGttcgtggcaggggttaTTCGGCGGCTGGACGAGA | 198. | 16.5 | 16 | 16.5 | Cancer |
| MSH2 | 27.5 ± 10.4 | TGAGGAGGTTTCGACATGgcggtgcagccgaaggAGACGCTGCAGTTGGAGA | 199. | 16.5 | 16 | 16.5 | Cancer |
| MUTYH | 24.9 ± 8.4 | TCACTGTCGGCGGCCATGacaccgctcgtctcccgcCTGAGTCGTCTGTGGGTA | 200. | 16.5 | 18 | 16.5 | Cancer |
| MYC | 13.4 ± 4.0 | TGCTTAGACGCTGGATTTttttcgggtagtggaaAACCAGGTAAGCACCGAA | 201. | 16.5 | 16 | 16.5 | Cancer |
| MYCL1 | 17.3 ± 0.6 | TCCCGCAGGGAGCGGACAtggactacgactcgtaCCAGCACTATTTCTACGA | 202. | 16.5 | 16 | 16.5 | Cancer |
| MYCN | 16.3 ± 11.6 | TGCCGAGCTGCTCCACgtccaccatgccgggcATGATCTGCAAGAACCCA | 203. | 14.5 | 16 | 16.5 | Cancer |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NBN | 46.3 ± 15.5 | TGAGGAGCCGGAC·CGAtgtggaaactgctgccCGCCGCGGGCCCGGCA | 204. | 14.5 | 16 | 14.5 | Cancer |
| NCOR1 | 29.6 ± 13.1 | TCTTTACTGATAATGTCAagttcatgttaccctcCCAACCAAGGAGCATTCA | 205. | 16.5 | 16 | 16.5 | Epigenetic |
| NCOR2 | 3.3 ± 0.6 | TGGAGGGCCACTGAGCcccgctacccgccccaCAGCCTTTCCTACCCA | 206. | 14.5 | 16 | 14.5 | Epigenetic |
| NTRK1 | none | TCGGCGCATGAAGGAGGTactcctcattttcgttCTCTCTCTCTGTGCCCCA | 207. | 16.5 | 16 | 16.5 | Cancer |
| PDGFRA | 16.0 ± 4.3 | TTGCGCTCGGGGCGGCCAtgtcggccggcgaggtCGAGCGCCTAGTGTCGGA | 208. | 16.5 | 16 | 16.5 | Cancer |
| PDGFRB | 16.0 ± 3.2 | TCTGCAGGACACCATGCGgcttccgggtgcgatgCCAGCTCTGGCCCTCAAA | 209. | 16.5 | 16 | 16.5 | Cancer |
| PHF8 | 22.2 ± 6.1 | TGAGTACTCCGCCTCTACcccggctgaagccccgcCCCCGCCGCCACCTATTA | 210. | 16.5 | 16 | 16.5 | Epigenetic |
| PMS2 | 26.9 ± 9.5 | TCGGGTGTTGCATCCATGgagcgagctgagagctcgAGGTGAGCGGGGCTCGCA | 211. | 16.5 | 18 | 16.5 | Cancer |
| PTCH1 | 27.5 ± 15.9 | TGGAACTGCTTAATAGaaacaggcttgtaattGTGAGTCCGCGCTGCA | 212. | 14.5 | 16 | 14.5 | Cancer |
| PTEN | 31.5 ± 11.7 | TCCCAGACATGACAGCCatcatcaaagagatcgTTAGCAGAAACAAAAGGA | 213. | 15.5 | 16 | 16.5 | Cancer |
| RARA | 13.4 ± 6.1 | TGGCATGGCCAGCAACAGcagctcctgcccgacacCTGGGGGCGGGCACCTCA | 214. | 16.5 | 17 | 16.5 | Cancer |
| RBBP5 | 15.7 ± 9.5 | TGCTGGGTGAGAAGGGCtgtggctgcgttttagaGAAGCGTTGGGTACTGGA | 215. | 15.5 | 17 | 16.5 | Epigenetic |
| RECQL4 | 22.1 ± 16.2 | TGCGGGACGTGCGGGAGCggctgcaggcgtggaGCGCGCGTTCCGACGGCA | 216. | 16.5 | 16 | 16.5 | Cancer |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| REST | none | TCAGAATACAGTTATGGCcacccaggtaatgggGCAGTCTTCTGGAGGAGGA | 217. | 16.5 | 16 | 16.5 | Epigenetic |
| RET | 5.4 ± 1.8 | TGAGTTCTGCCGGCCGCCggctcccgcaggggccaGGGCGAAGTTGGCGCCGA | 218. | 16.5 | 17 | 16.5 | Cancer |
| RNF2 | none | TTCTTTATTTCCAGCAATgtctcaggctgtgcagACAAACGGAACTCAACCA | 219. | 16.5 | 16 | 16.5 | Epigenetic |
| RUNX1 | 25.1 ± 6.9 | TTCAGGAGGAAGCGATGGcttcagacagcatattTGAGTCATTTCCTTCGTA | 220. | 16.5 | 16 | 16.5 | Epigenetic |
| SDHB | 36.4 ± 19.2 | TCTCCTTGAGGCGCCGGTtgccggccacaaccctTGGCGGAGCCTGCCTGCA | 221. | 16.5 | 16 | 16.5 | Cancer |
| SDHC | 13.7 ± 3.4 | TGTTGCTGAGGTGACTTCagtgggactgggagttggtGCCTGCGGCCCTCCGGA | 222. | 16.5 | 19 | 15.5 | Cancer |
| SDHD | 42.0 ± 7.8 | TCAGGAACGAGATGGCGGttctctggaggctgagtGCCGTTTGCGGTGCCCTA | 223. | 16.5 | 17 | 16.5 | Cancer |
| SETDB1 | 33.5 ± 6.1 | TGCAGAGGACAAAAGCATgtcttcccttcctgggTGCATTGGTTTGGATGCA | 224. | 16.5 | 16 | 16.5 | Epigenetic |
| SIRT6 | 43.3 ± 3.1 | TTACGCGGCGGGGCTGTCgccgtacgcggacaagggCAAGTGCGGCCTCCCGGA | 225. | 16.5 | 18 | 16.5 | Epigenetic |
| SMAD2 | 3.9 ± 1.6 | TTTGGTAAGAACATGTCGtccatcttgccattcacGCCGCCAGTTGTGAAGA | 226. | 16.5 | 17 | 15.5 | Cancer |
| SS18 | 31.4 ± 7.9 | TGGTGACGGCGGCAACATgtctgtggctttcggcggCCCCGAGGCAGCGAGGCA | 227. | 16.5 | 17 | 16.5 | Cancer |
| SUZ12 | 13.1 ± 0.4 | TGGCGCCTCAGAAGCAcggcggtgggggagggGGCGGCTCGGGGCCCA | 228. | 14.5 | 16 | 14.5 | Epigenetic |
| TFE3 | 17.3 ± 2.4 | TCATGTCTCATGCGGCCGaaccagctcgggatggCGTAGAGGCCAGCGCGGA | 229. | 16.5 | 16 | 16.5 | Cancer |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| TGFBR2 | none | TCGGGGGCTGCTCAGGGGcctgtggccgctgcacaTCGTCCTGTGGACGCGTA | 230. | 16.5 | 17 | 16.5 | Cancer |
| TLX3 | none | TTCCGCCCGCCCAGGATGgaggcgcccgccagcgcGCAGACCCCGCACCCGCA | 231. | 16.5 | 17 | 16.5 | Cancer |
| TP53 | 19.9 ± 3.6 | TTGCCGTCCCAAGCAATGgatgatttgatgctgtcCCCGGACGATATTGAACA | 232. | 16.5 | 17 | 16.5 | Cancer |
| TSC2 | 30.7 ± 22.7 | TCCTGGTCCACCATGGCcaaaccaacaagcaaagATTCAGGCTTGAAGGAGA | 233. | 15.5 | 17 | 16.5 | Cancer |
| VHL | 19.4 ± 1.1 | TCTGGATCGCGGAGGGAAtgccccggagggcggaGAACTGGGACGAGGCCGA | 234. | 16.5 | 16 | 16.5 | Cancer |
| XPA | 12.9 ± 2.2 | TGGGCCAGAGATGGCGGCggccgacgggggcttttgCCGGAGGCGGCGGCTTTA | 235. | 16.5 | 16 | 16.5 | Cancer |
| XPC | 31.4 ± 4.2 | TGCCCAGACAAGCAACATggctcggaaacgcgcggccGGCGGGGAGCCGCGGGGA | 236. | 16.5 | 19 | 16.5 | Cancer |

The abilities of the 96 TALEN pairs to introduce NHEJ-mediated insertion or deletion (indel) mutations at their intended endogenous gene targets were tested in cultured human cells using a slightly modified version of a previously described T7 Endonuclease I (T7EI) assay (Mussolino et al., 2011, Nucleic Acids Res., 39:9283-93; Kim et al., 2009, Genome Res., 19:1279-88). With this T7EI assay, 83 of the 96 TALEN pairs showed evidence of NHEJ-mediated mutagenesis at their intended endogenous gene target sites, an overall success rate of ~86% (Table 7). The efficiencies of TALEN-induced mutagenesis we observed ranged from 2.5% to 55.8% with a mean of 22.5%. To provide molecular confirmation of the mutations we identified by T7EI assay, we sequenced target loci for 11 different TALEN pairs that induced varying efficiencies of mutagenesis (FIGs. 21A-D). As expected, this sequencing revealed indels at the expected target gene sites with frequencies similar to those determined by the T7EI assays.

The nucleotide and amino acid sequences for 14 of the 96 pairs of TALENs targeted to the endogenous human genes in Table 7 are presented below. Each TALEN monomer is presented as follows:

(1) A header with information presented in the format: Gene target_Left or Right monomer_Target DNA site shown 5' to 3'_TALE repeat monomers and 0.5 repeat plasmid used with code as shown in Table 4.

(2) DNA sequence encoding the N-terminal part of the TALE required for activity, the TALE repeat array, the C-terminal 0.5 TALE repeat domain, and the C-terminal 63 amino acids required for activity from a NheI site to a BamHI site. This sequence is present in the "Vector Sequence" plasmid shown below, taking the place of the underlined X's flanked by NheI and BamHI sites

(3) Amino acid sequences the N-terminal part of the TALE required for activity, the TALE repeat array, the C-terminal 0.5 TALE repeat domain, and the C-terminal 63 amino acids required for activity shown from the start of translation (located just 3' to the NheI site and including an N-terminal FLAG epitope tag) to a Gly-Ser sequence (encoded by the BamHI site) that serves as a linker from the TALE repeat array to the FokI cleavage domain.

**VECTOR SEQUENCE**

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTG**GCTAGCXXXXXXXXXXGGATCC**CAACTAG
TCAAAAGTGAACTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATT
GAATTAATTGAAATTGCCAGAAATTCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTAT
GAAAGTTTATGGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCG
GATCTCCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGC
CAAGCAGATGAAATGCAACGATATGTCGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATG

GTGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACT
ACAAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTT
TTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGACGGAAATTTAATAACGG
CGAGATAAACTTTTAAGGGCCCTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGC
GTACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC
CAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTC
CTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCT
GAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGC
GGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCT
TCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTC
CGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATC
GCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAA
CTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTAT
TGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGG
TGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA
GGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACC
ATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGG
CTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAG
GAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGAT
CAGCACGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAA
ACCATGGCCAAGCCTTTGTCTCAAGAAGAATCCACCCTCATTGAAAGAGCAACGGCTACAATCAACAGCAT
CCCCATCTCTGAAGACTACAGCGTCGCCAGCGCAGCTCTCTCTAGCGACGGCCGCATCTTCACTGGTGTCA
ATGTATATCATTTTACTGGGGGACCTTGTGCAGAACTCGTGGTGCTGGGCACTGCTGCTGCTGCGGCAGCT
GGCAACCTGACTTGTATCGTCGCGATCGGAAATGAGAACAGGGGCATCTTGAGCCCCTGCGGACGGTGTCG
ACAGGTGCTTCTCGATCTGCATCCTGGGATCAAAGCGATAGTGAAGGACAGTGATGGACAGCCGACGGCAG
TTGGGATTCGTGAATTGCTGCCCTCTGGTTATGTGTGGGAGGGCTAAGCACTTCGTGGCCGAGGAGCAGGA
CTGACACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCC
GGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTT
ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACT
GCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCT
AGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAAC
ATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTT
GCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG
GGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGG
AAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCC
ATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTAC
CGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCA
GTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCC
TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGG
TAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT
ACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAG
AAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCAC
GTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGT
TTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACC

TATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATAC
GGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTA
TCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCA
GTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA
TTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCA
AGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG
AAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCG
AGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCAT
TGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
CTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGG
CAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC
AAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC SEQ ID NO: 244

**TALE REPEAT SEQUENCES**

| Target | SEQEUNCE | SEQ ID NO: | SEQUENCE | SEQ ID NO: |
|---|---|---|---|---|
| >APC_Left_TATGTACGCCTCCCTGGG_TAL/006/015/019/025/026/012/019/022/027/015/017/022/027/015/019/024/JDS74/('TATGTACGCCTCCCTGGG' disclosed as SEQ ID NO: 412) | GCTAGCaccATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAG | 245. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQADHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS | 246. |

| | | | | |
|---|---|---|---|---|
| | GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCATCCCACGACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGAATAA<br>CAATGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAACAACAACGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT | | | |

| | | | | |
|---|---|---|---|---|
| | GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTAATAATAAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >APC_Right_TACAGAAGCGGGCAAAGG_TAL/006/012/016/024/026/011/019/022/029/014/019/022/026/011/016/024/JDS74/ ('TACAGAAGCGGGCAAAGG' disclosed as SEQ ID NO: 413) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCGAACATT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGAACATTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAACAACAACGGCGGTAAGCAGGC | 247. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>IGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | 248. |

| | | | | |
|---|---|---|---|---|
| | GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAACATCGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGAATAA<br>CAATGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAATAATAACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGAATAACAATGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CGAACATTGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAACATCGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGAACATTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC | | | |

| | | | | |
|---|---|---|---|---|
| | AAGTGGTCGCCATCGCCAACAACAACGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTAATAATAAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >BRCA1_Left_TCCGAAGCTGACAGATGG_TAL/007/012/019/021/026/014/017/025/029/011/017/021/029/011/020/024/JDS74/ ('TCCGAAGCTGACAGATGG' disclosed as SEQ ID NO: 414) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCACATGAC<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGAATAACAATGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC | 249. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN<br>NGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | 250. |

| | | | | |
|---|---|---|---|---|
| | AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCCAATATTGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCGAATGGCGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAACATCGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCCAATA<br>TTGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGA<br>ACAATAATGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAACATCGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCCAACGGTGGAGGGAAACAA | | | |

| | | | | |
|---|---|---|---|---|
| | GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAACAACAACGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTAATAATAAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >BRCA1_Right_TCAGGTTCCGCCCCTACC_TAL/007/011/019/024/030/015/017/022/029/012/017/022/027/015/016/022/JDS71/('TCAGGTTCGCCCCTACC' disclosed as SEQ ID N: 415) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCACATGAC<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAGC<br>AACATCGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT | 251. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN<br>NGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | 252. |

| | | | | |
|---|---|---|---|---|
| | GCGAATAACAATGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAACAACAACGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCAAACGGAGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCATCCCACGACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG | | | |

| | | | | |
|---|---|---|---|---|
| | TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGAACATTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAGCCATGATGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCCCACGAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >BRCA2_Left_TTAGACTTAGGTAAGTAA_TAL/010/011/019/021/027/015/020/021/029/014/020/021/026/014/020/021/JDS70/('TTAGACTTAGGTAAGTAA' disclosed as SEQ ID NO: 416) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCAAACGGA<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAGC<br>AACATCGGTGGCAAACAGGCTCTTGAGACGGT | 253. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN<br>NGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | 254. |

| | | | | |
|---|---|---|---|---|
| | TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGAATAACAATGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCCAATATTGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCACATGACGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCCAA<br>CGGTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAATAATAACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCCAATA<br>TTGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CGAACATTGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA | | | |

| | | | | |
|---|---|---|---|---|
| | TTGCAAATAATAACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCCAACGGTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCCAATATTGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAACATC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >BRCA2_Right_TAGTTTGTAGTTCTCCCC_TAL/006/014/020/025/030/014/020/021/029/015/020/022/030/012/017/022/JDS71/('TAGTTTGTAGTTCTCCCC'<br><br>disclosed as SEQ ID NO: 417) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCGAACATT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC | 255. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>GGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRP | 256. |

| | | | | |
|---|---|---|---|---|
| | TTACACCGGAGCAAGTCGTGGCCATTGCAAAT<br>AATAACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCCAACGGTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCGAATGGCGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCAAACGGAGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCCAA<br>CGGTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CAAACGGAGGGGGAAAGCAAGCCCTGGAAACC | | ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGCATGACGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAGCCATGATGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCCCACGAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >ERCC2_Left_TCCGGCCGGCGCCATGA_TAL/007/012/019/024/027/012/019/024/027/014/017/022/026/015/034/JDS70/('TCCGGCCGGCGCCATGA' disclosed as SEQ ID NO: 418) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCACATGAC | 257. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN<br>NGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN | 258. |

| | | | | |
|---|---|---|---|---|
| | GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGAATAACAATGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAACAACAACGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCACATGACGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCATCCCACGACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGAATAA<br>CAATGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAACAACAACGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAATAATAACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG | | NNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIG<br>GRPALESIVAQLSRPDPALAALTNDHLVALACLGGRP<br>ALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CGAACATTGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTCTGACACCCGAAC<br>AGGTGGTCGCCATTGCTTCTAACATCGGAGGA<br>CGGCCAGCCTTGGAGTCCATCGTAGCCCAATT<br>GTCCAGGCCCGATCCCGCGTTGGCTGCGTTAA<br>CGAATGACCATCTGGTGGCGTTGGCATGTCTT<br>GGTGGACGACCCGCGCTCGATGCAGTCAAAAA<br>GGGTCTGCCTCATGCTCCCGCATTGATCAAAA<br>GAACCAACCGGCGGATTCCCGAGAGAACTTCC<br>CATCGAGTCGCGGGATCC | | | |
| >ERCC2_Right_TAGCGAGCGCGACCCC_TAL/006/014/017/024/026/014/017/024/027/014/016/022/027/012/JDS71/ ('TAGCGAGCGCGACCCC' disclosed as SEQ ID NO: 419) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCGAACATT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG | 259. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH<br>DGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS<br>HDGGRPALESIVAQLSRPDPALAALTNDHLVALACLG | 260. |

| | | | | |
|---|---|---|---|---|
| | GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAAT<br>AATAACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGCATGACGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAACAACAACGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAACAACAACGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAATAATAACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGAACATTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT | | GRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACACCCGAACAGGTGG<br>TCGCCATTGCTTCCCACGACGGAGGACGGCCA<br>GCCTTGGAGTCCATCGTAGCCCAATTGTCCAG<br>GCCCGATCCCGCGTTGGCTGCGTTAACGAATG<br>ACCATCTGGTGGCGTTGGCATGTCTTGGTGGA<br>CGACCCGCGCTCGATGCAGTCAAAAAGGGTCT<br>GCCTCATGCTCCCGCATTGATCAAAAGAACCA<br>ACCGGCGGATTCCCGAGAGAACTTCCCATCGA<br>GTCGCGGGATCC | | | |
| >FANCA_Left_TAGGCGCCAAGGCCATGT_TAL/006/014/019/022/029/012/017/021/026/014/019/022/027/011/020/024/JDS78/ ('TAGGCGCCAAGGCCATGT' disclosed as SEQ ID NO: 420) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCGAACATT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAAT<br>AATAACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC | 261. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN<br>NGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | 262. |

| | | | | |
|---|---|---|---|---|
| | ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGAATAACAATGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGAACAATAATGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCATCCCACGACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCGAACATTGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAATAATAACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGAATAACAATGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAACATCGGTGGCAAACAGGCTCTT | | | |

| | | | | |
|---|---|---|---|---|
| | GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCCAACGGTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAACAACAACGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAATGGG<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >FANCA_Right_TGGCCCGAGGCGGAGTTC_TAL/009/014/017/022/027/014/016/024/029/012/019/024/026/014/020/025/JDS71/('TGGCCCGAGGCGGAGTTC'<br><br>disclosed as SEQ ID NO: 421) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGAACAATAAT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAAT | 263. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH<br>DGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD | 264. |

| | | | | |
|---|---|---|---|---|
| | AATAACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGCATGACGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCACATGACGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA<br>CATTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAACAACAACGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCATCCCACGACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGAATAACAATGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAACAACA<br>ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CGAACATTGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA | | AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAATAATAACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCCAACGGTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCCCACGAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >FANCC_Left_TGAAGGGACATCACCTTT_TAL/009/011/016/024/029/014/016/022/026/015/017/021/027/012/020/025/JDS78/ ('TGAAGGGACATCACCTTT' disclosed as SEQ ID NO: 422) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGAACAATAAT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG | 265. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>IGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG | 266. |

| | | | | |
|---|---|---|---|---|
| | GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAGC<br>AACATCGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGAACATTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAACAACAACGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGAACAATAATGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA<br>CATTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCGAACATTGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCCAATA<br>TTGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT | | GKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCCAACGGTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAATGGG<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >FANCC_Right_TCTACTGAATCTTGAGC_TAL/007/015/016/022/030/014/016/021/030/012/020/025/029/011/034/JDS71/ ('TCTACTGAATCTTGAGC' disclosed as SEQ ID NO: 423) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC | 267. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>IGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVA | 268. |

| | | | | |
|---|---|---|---|---|
| | CAGACCAGGTAGTCGCAATCGCGTCACATGAC<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAGC<br>AATGGGGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGAACATTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCAAACGGAGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA<br>CATTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCAAACGGAGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCATCCCACGACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCGAATG<br>GCGGCGGTAAGCAGGCGCTGGAAACAGTACAG | | IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN<br>NNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDG<br>GRPALESIVAQLSRPDPALAALTNDHLVALACLGGRP<br>ALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGA<br>ACAATAATGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAACATCGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTCTGACACCCGAAC<br>AGGTGGTCGCCATTGCTTCCCACGACGGAGGA<br>CGGCCAGCCTTGGAGTCCATCGTAGCCCAATT<br>GTCCAGGCCCGATCCCGCGTTGGCTGCGTTAA<br>CGAATGACCATCTGGTGGCGTTGGCATGTCTT<br>GGTGGACGACCCGCGCTCGATGCAGTCAAAAA<br>GGGTCTGCCTCATGCTCCCGCATTGATCAAAA<br>GAACCAACCGGCGGATTCCCGAGAGAACTTCC<br>CATCGAGTCGCGGGATCC | | | |
| >FANCG_Left_TCGGCCACCATGTCCC_TAL/007/014/019/022/027/011/017/022/026/015/019/025/027/012/JDS71/ ('TCGGCCACCATGTCCC' disclosed as SEQ ID NO: 424) | GCTAGCaccATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGTCACATGAC | 269. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS | 270. |

| | | | | |
|---|---|---|---|---|
| | GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG | | HDGGRPALESIVAQLSRPDPALAALTNDHLVALACLG | |
| | GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC | | GRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |
| | TTACACCGGAGCAAGTCGTGGCCATTGCAAAT | | | |
| | AATAACGGTGGCAAACAGGCTCTTGAGACGGT | | | |
| | TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC | | | |
| | ACGGGCTGACTCCCGATCAAGTTGTAGCGATT | | | |
| | GCGAATAACAATGGAGGGAAACAAGCATTGGA | | | |
| | GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC | | | |
| | AAGCCCACGGTTTGACGCCTGCACAAGTGGTC | | | |
| | GCCATCGCCAGCCATGATGGCGGTAAGCAGGC | | | |
| | GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC | | | |
| | TGTGCCAGGATCATGGACTGACCCCAGACCAG | | | |
| | GTAGTCGCAATCGCGTCACATGACGGGGGAAA | | | |
| | GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC | | | |
| | CGGTCCTTTGTCAAGACCACGGCCTTACACCG | | | |
| | GAGCAAGTCGTGGCCATTGCAAGCAACATCGG | | | |
| | TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC | | | |
| | TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG | | | |
| | ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA | | | |
| | TGACGGAGGGAAACAAGCATTGGAGACTGTCC | | | |
| | AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC | | | |
| | GGTTTGACGCCTGCACAAGTGGTCGCCATCGC | | | |
| | CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA | | | |
| | CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG | | | |
| | GATCATGGACTGACCCCAGACCAGGTAGTCGC | | | |
| | AATCGCGTCGAACATTGGGGGAAAGCAAGCCC | | | |
| | TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT | | | |
| | TGTCAAGACCACGGCCTTACACCGGAGCAAGT | | | |
| | CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC | | | |
| | AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA | | | |
| | GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA | | | |
| | TCAAGTTGTAGCGATTGCGAATAACAATGGAG | | | |
| | GGAAACAAGCATTGGAGACTGTCCAACGGCTC | | | |
| | CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC | | | |
| | GCCTGCACAAGTGGTCGCCATCGCCTCGAATG | | | |
| | GCGGCGGTAAGCAGGCGCTGGAAACAGTACAG | | | |
| | CGCCTGCTGCCTGTACTGTGCCAGGATCATGG | | | |

| | | | | |
|---|---|---|---|---|
| | ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACACCCGAACAGGTGG<br>TCGCCATTGCTTCCCACGACGGAGGACGGCCA<br>GCCTTGGAGTCCATCGTAGCCCAATTGTCCAG<br>GCCCGATCCCGCGTTGGCTGCGTTAACGAATG<br>ACCATCTGGTGGCGTTGGCATGTCTTGGTGGA<br>CGACCCGCGCTCGATGCAGTCAAAAAGGGTCT<br>GCCTCATGCTCCCGCATTGATCAAAAGAACCA<br>ACCGGCGGATTCCCGAGAGAACTTCCCATCGA<br>GTCGCGGGATCC | | | |
| >FANCG_Right_TCCAGGCAGCTGGAGCCC_TAL/007/012/016/024/029/012/016/024/027/015/019/024/026/014/017/022/JDS71/('TCCAGGCAGCTGGAGCCC' disclosed as SEQ ID NO: 425) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCACATGAC<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT | 271. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>IGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | 272. |

| | | | | |
|---|---|---|---|---|
| | TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGAACATTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAACAACAACGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGAACAATAATGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCATCCCACGACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA<br>CATTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAACAACAACGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGAATAACAATGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAACAACA<br>ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CGAACATTGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA | | | |

| | | | | |
|---|---|---|---|---|
| | TTGCAAATAATAACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGCATGACGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAGCCATGATGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCCCACGAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >JAK2_Left_TCTGAAAAA_GACTCTGCA_TAL/007/015/019/021/026/011/016/021/029/011/017/025/027/015/019/022/JDS70/('TCTGAAAAAGACTCTGCA' disclosed as SEQ ID NO: 426) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCACATGAC<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC | 273. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN<br>NGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN<br>NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGRP | 274. |

| | | | | |
|---|---|---|---|---|
| | TTACACCGGAGCAAGTCGTGGCCATTGCAAGC<br>AATGGGGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGAATAACAATGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCCAATATTGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAACATCGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA<br>CATTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAACATCGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCGAATG<br>GCGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC | | ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCAGCCATGATGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAACATC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAAGAACCAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC | | | |
| >JAK2_Right_TCCATTTCTGTCATCGTA_TAL/007/012/016/025/030/015/017/025/029/015/017/021/030/012/019/025/JDS70/('TCCATTTCTGTCATCGTA' disclosed as SEQ ID NO: 427) | GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCCAAGAAGAAG AGGAAGGTGGGCATTCACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGCACCCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCGAAGCCACGCACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCACATGAC | 275. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN IGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQA LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR LLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLP VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN | 276. |

| | | | | |
|---|---|---|---|---|
| | GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGAACATTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCGAATGGCGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCAAACGGAGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCGAATGGCGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCCAATA<br>TTGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG | | NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CAAACGGAGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAACATC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >KRAS_Left<br>_TGAAAATGA<br>CTGAATATA_<br>TAL/009/01<br>1/016/021/<br>026/015/01<br>9/021/027/<br>015/019/02<br>1/026/015/<br>016/025/JD<br>S70/<br>('TGAAAATG<br>ACTGAATATA<br>'<br>disclosed<br>as SEQ ID | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA | 277. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>IGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ | 278. |

| | | | | |
|---|---|---|---|---|
| NO: 428) | TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGAACAATAAT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAGC<br>AACATCGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGAACATTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCCAATATTGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGAATAA<br>CAATGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGAATAACAATGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCCAATA | | VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | TTGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CGAACATTGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGAACATTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAACATC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >KRAS_Right_TTGCCTACGCCACCAGC_TAL/010/014/017/022/030/011/017/024/027/012/016/022/027/011/034/JDS71/('TTGCCTACGCCACCAGC' disclosed as SEQ ID | GCTAGCaccATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACCAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGC | 279. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH | 280. |

| | | | | |
|---|---|---|---|---|
| NO: 429) | AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCAAACGGA<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAAT<br>AATAACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGCATGACGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCAAACGGAGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAACATCGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAACAACAACGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCATCCCACGACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGAACATTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC | | GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN<br>NNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDG<br>GRPALESIVAQLSRPDPALAALTNDHLVALACLGGRP<br>ALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAACATCGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTCTGACACCCGAAC<br>AGGTGGTCGCCATTGCTTCCCACGACGGAGGA<br>CGGCCAGCCTTGGAGTCCATCGTAGCCCAATT<br>GTCCAGGCCCGATCCCGCGTTGGCTGCGTTAA<br>CGAATGACCATCTGGTGGCGTTGGCATGTCTT<br>GGTGGACGACCCGCGCTCGATGCAGTCAAAAA<br>GGGTCTGCCTCATGCTCCCGCATTGATCAAAA<br>GAACCAACCGGCGGATTCCCGAGAGAACTTCC<br>CATCGAGTCGCGGGATCC | | | |
| >MYC_Left_<br>TGCTTAGACG<br>CTGGATTT_T<br>AL/009/012<br>/020/025/0<br>26/014/016<br>/022/029/0<br>12/020/024<br>/029/011/0<br>20/025/JDS<br>78/<br>('TGCTTAGA<br>CGCTGGATTT<br>'<br>disclosed | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA | 281. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>GGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLT | 282. |

| | | | | |
|---|---|---|---|---|
| as SEQ ID NO: 430) | ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGAACAATAAT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCCAACGGTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCGAATGGCGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA<br>CATTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCATCCCACGACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC | | PAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | GCCTGCACAAGTGGTCGCCATCGCCAACAACA<br>ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGA<br>ACAATAATGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAACATCGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCCAACGGTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAATGGG<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >MYC_Right<br>_TTCGGTGCT<br>_TACCTGGTT_<br>TAL/010/01<br>2/019/024/<br>030/014/01<br>7/025/030/<br>011/017/02<br>2/030/014/<br>019/025/JD<br>S78/<br>('TTCGGTGC<br>TTACCTGGTT | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT | 283. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN<br>NGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC | 284. |

| | | | | |
|---|---|---|---|---|
| disclosed as SEQ ID NO: 431) | TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCAAACGGA<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGAATAACAATGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAACAACAACGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCAAACGGAGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCGAATGGCGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCAAACGGAGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAACATCGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG | | QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN<br>NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CAAACGGAGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAATAATAACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAATGGG<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >PTEN_Left<br>_TCCCAGACA<br>TGACAGCC_T<br>AL/007/012<br>/017/021/0<br>29/011/017<br>/021/030/0<br>14/016/022<br>/026/014/0<br>32/JDS71/<br>('TCCCAGAC | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG | 285. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH<br>DGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR | 286. |

| | | | | |
|---|---|---|---|---|
| ATGACAGCC' disclosed as SEQ ID NO: 432) | GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCACATGAC<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGCATGACGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCCAATATTGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGAACAATAATGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAACATCGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCAAACGGAGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAATAATAACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA | | LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>HDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDG<br>GRPALESIVAQLSRPDPALAALTNDHLVALACLGGRP<br>ALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGAACATTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CGAACATTGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAATAATAACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGCATGACGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTCTGACACCCGAAC<br>AGGTGGTCGCCATTGCTTCCCACGACGGAGGA<br>CGGCCAGCCTTGGAGTCCATCGTAGCCCAATT<br>GTCCAGGCCCGATCCCGCGTTGGCTGCGTTAA<br>CGAATGACCATCTGGTGGCGTTGGCATGTCTT<br>GGTGGACGACCCGCGCTCGATGCAGTCAAAAA<br>GGGTCTGCCTCATGCTCCCGCATTGATCAAAA<br>GAACCAACCGGCGGATTCCCGAGAGAACTTCC<br>CATCGAGTCGCGGGATCC | | | |
| >PTEN_Righ<br>t_TCCTTTG<br>TTTCTGCTAA<br>_TAL/007/0<br>12/020/025<br>/030/015/0<br>19/025/030<br>/015/017/0<br>25/029/012<br>/020/021/J<br>DS70/<br>('TCCTTTG | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA | 287. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>GGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLP | 288. |

| | | | | |
|---|---|---|---|---|
| TTTCTGCTAA<br>'<br>disclosed<br>as SEQ ID<br>NO: 433) | CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCACATGAC<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCCAACGGTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCTCGAATGGCGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCAAACGGAGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGAATAA<br>CAATGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCGAATGGCGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCAAACGGAGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA | | VLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCGAATG<br>GCGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGA<br>ACAATAATGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCCAACGGTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCCAATATTGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAACATC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >TP53_Left<br>_TTGCCGTCC<br>CAAGCAATG_<br>TAL/010/01<br>4/017/022/<br>029/015/01<br>7/022/027/<br>011/016/02<br>4/027/011/<br>016/025/JD | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC | 289. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH<br>DGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET | 290. |

| | | | | |
|---|---|---|---|---|
| S74/ ('TTGCCGTC CCAAGCAATG ' disclosed as SEQ ID NO: 434) | CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCAAACGGA<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAAT<br>AATAACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGCATGACGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGAACAATAATGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA<br>TGACGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAACATCGGTGGCAAAC | | VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>NIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGAACATTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAACAACA<br>ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAACATCGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGAACATTGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTAATAATAAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >TP53_Right_TGTTCAATATCGTCCGGG_TAL/009/015/020/022/026/011/020/021/030/012/019/0 | GCTAGCaccATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCG | 291. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIANNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG | 292. |

| | | | | |
|---|---|---|---|---|
| 25/027/012 /019/024/J DS74/ ('TGTTCAAT ATCGTCCGGG ' disclosed as SEQ ID NO: 435) | CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGAACAATAAT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAGC<br>AATGGGGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCCAACGGTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAACATCGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCCAA<br>CGGTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCAAACGGAGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT | | KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN<br>NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCATCCCACGACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGAATAACAATGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCTCGAATG<br>GCGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAACAACAACGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTAATAATAAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >XPA_Left_<br>TGGGCCAGAG<br>ATGGCGGC_T<br>AL/009/014<br>/019/022/0<br>27/011/019 | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG | 293. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI | 294. |

| | | | | |
|---|---|---|---|---|
| /021/029/011/020/024/029/012/019/024/JDS71/ ('TGGGCCAGAGATGGCGGC' disclosed as SEQ ID NO: 436) | AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGAACAATAAT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAAT<br>AATAACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGAATAACAATGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCACATGACGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAGCAACATCGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGAATAA<br>CAATGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC | | ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN<br>NGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN<br>NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | AATCGCGAACAATAATGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAACATCGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAACAACA<br>ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGA<br>ACAATAATGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAACAACAACGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCCCACGAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >XPA_Right_TAAAGCCGC<br>CGCCTCCGG_<br>TAL/006/01 | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT | 295. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT | 296. |

| | | | | |
|---|---|---|---|---|
| 1/016/024/027/012/019/022/027/014/017/022/030/012/017/024/JDS74/ ('TAAAGCCGCCGCCTCCGG' disclosed as SEQ ID NO: 437) | GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCGAACATT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCAAGC<br>AACATCGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGAACATTGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAACAACAACGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCACATGACGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCATCCCACGACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGAATAA<br>CAATGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA | | GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN<br>IGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAATAATAACGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CAAACGGAGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGCATGACGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAACAACAACGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTAATAATAAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |
| >XPC_Left_<br>TGCCCAGACA | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG | 297. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL | 298. |

| | | | | |
|---|---|---|---|---|
| AGCAACAT_T<br>AL/009/012<br>/017/022/0<br>26/014/016<br>/022/026/0<br>11/019/022<br>/026/011/0<br>17/021/JDS<br>78/<br>('TGCCCAGA<br>CAAGCAACAT<br>'<br>disclosed<br>as SEQ ID<br>NO: 438) | ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCCCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGAACAATAAT<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGCATGACGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGTCGAACATTGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCAAATAATAACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA<br>CATTGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC | | VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH<br>DGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS<br>HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | |

| | | | | |
|---|---|---|---|---|
| | GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCGAACATTGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAACATCGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGAATAACAATGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CGAACATTGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCAAGCAACATCGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGTCGCATGACGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCTCCAATATTGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCTAATGGG<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA<br>TCAAAAGAACCAACCGGCGGATTCCCGAGAGA<br>ACTTCCCATCGAGTCGCGGGATCC | | | |

| | | | | |
|---|---|---|---|---|
| >XPC_Right_TCCCCGCGG_CTCCCCGCC_TAL/007/012/017/022/029/012/019/024/027/015/017/022/027/012/019/022/JDS71/ ('TCCCCGCGGCTCCCCGCC' disclosed as SEQ ID NO: 439) | GCTAGCaccATGGACTACAAAGACCATGACGG<br>TGATTATAAAGATCATGACATCGATTACAAGG<br>ATGACGATGACAAGATGGCCCCCAAGAAGAAG<br>AGGAAGGTGGGCATTCACCGCGGGGTACCTAT<br>GGTGGACTTGAGGACACTCGGTTATTCGCAAC<br>AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG<br>AGCACCGTCGCGCAACACCACGAGGCGCTTGT<br>GGGGCATGGCTTCACTCATGCGCATATTGTCG<br>CGCTTTCACAGCACCCTGCGGCGCTTGGGACG<br>GTGGCTGTCAAATACCAAGATATGATTGCGGC<br>CCTGCCCGAAGCCACGCACGAGGCAATTGTAG<br>GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA<br>CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT<br>TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC<br>AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA<br>ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA<br>TGCGCTCACCGGGGCCCCCTTGAACCTGACCC<br>CAGACCAGGTAGTCGCAATCGCGTCACATGAC<br>GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG<br>GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC<br>TTACACCGGAGCAAGTCGTGGCCATTGCATCC<br>CACGACGGTGGCAAACAGGCTCTTGAGACGGT<br>TCAGAGACTTCTCCCAGTTCTCTGTCAAGCCC<br>ACGGGCTGACTCCCGATCAAGTTGTAGCGATT<br>GCGTCGCATGACGGAGGGAAACAAGCATTGGA<br>GACTGTCCAACGGCTCCTTCCCGTGTTGTGTC<br>AAGCCCACGGTTTGACGCCTGCACAAGTGGTC<br>GCCATCGCCAGCCATGATGGCGGTAAGCAGGC<br>GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC<br>TGTGCCAGGATCATGGACTGACCCCAGACCAG<br>GTAGTCGCAATCGCGAACAATAATGGGGGAAA<br>GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC<br>CGGTCCTTTGTCAAGACCACGGCCTTACACCG<br>GAGCAAGTCGTGGCCATTGCATCCCACGACGG<br>TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC<br>TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG<br>ACTCCCGATCAAGTTGTAGCGATTGCGAATAA | 299. | ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGI<br>HRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL<br>VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA<br>THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT<br>GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV<br>VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI<br>ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH<br>DGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG<br>KQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQA<br>LETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALET<br>VQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR<br>LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLP<br>VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC<br>QDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH<br>GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT<br>PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ<br>VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA<br>IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN<br>NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG<br>GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRP<br>ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD<br>AVKKGLPHAPALIKRTNRRIPERTSHRVAGS | 300. |

| | | | | |
|---|---|---|---|---|
| | CAATGGAGGGAAACAAGCATTGGAGACTGTCC<br>AACGGCTCCTTCCCGTGTTGTGTCAAGCCCAC<br>GGTTTGACGCCTGCACAAGTGGTCGCCATCGC<br>CAACAACAACGGCGGTAAGCAGGCGCTGGAAA<br>CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG<br>GATCATGGACTGACCCCAGACCAGGTAGTCGC<br>AATCGCGTCACATGACGGGGGAAAGCAAGCCC<br>TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT<br>TGTCAAGACCACGGCCTTACACCGGAGCAAGT<br>CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC<br>AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA<br>GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA<br>TCAAGTTGTAGCGATTGCGTCGCATGACGGAG<br>GGAAACAAGCATTGGAGACTGTCCAACGGCTC<br>CTTCCCGTGTTGTGTCAAGCCCACGGTTTGAC<br>GCCTGCACAAGTGGTCGCCATCGCCAGCCATG<br>ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG<br>CGCCTGCTGCCTGTACTGTGCCAGGATCATGG<br>ACTGACCCCAGACCAGGTAGTCGCAATCGCGT<br>CACATGACGGGGGAAAGCAAGCCCTGGAAACC<br>GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA<br>CCACGGCCTTACACCGGAGCAAGTCGTGGCCA<br>TTGCATCCCACGACGGTGGCAAACAGGCTCTT<br>GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG<br>TCAAGCCCACGGGCTGACTCCCGATCAAGTTG<br>TAGCGATTGCGAATAACAATGGAGGGAAACAA<br>GCATTGGAGACTGTCCAACGGCTCCTTCCCGT<br>GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC<br>AAGTGGTCGCCATCGCCAGCCATGATGGCGGT<br>AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT<br>GCCTGTACTGTGCCAGGATCATGGACTGACAC<br>CCGAACAGGTGGTCGCCATTGCTTCCCACGAC<br>GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC<br>CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG<br>CGTTAACGAATGACCATCTGGTGGCGTTGGCA<br>TGTCTTGGTGGACGACCCGCGCTCGATGCAGT<br>CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA | | | |

| | | | | |
|---|---|---|---|---|
| | TCAAAAGAACCAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC | | | |

## OTHER EMBODIMENTS

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

SEQUENCE LISTING IN ELECTRONIC FORM

In accordance with Section 111(1) of the Patent Rules, this description contains a sequence listing in electronic form in ASCII text format (file: 60412-4745 Seq 05-FEB-14 v1.txt).

A copy of the sequence listing in electronic form is available from the Canadian Intellectual Property Office.

CLAIMS:

1. A process for assembling nucleic acids encoding custom TALE repeat array proteins comprising:

(a) providing a first nucleic acid comprising a sequence encoding a first set comprising one or more transcription activator-like effector (TALE) repeat domains;

(b) contacting the first nucleic acid with a first enzyme, wherein the first enzyme creates a first ligatable end;

(c) providing a second nucleic acid comprising a sequence encoding a second set comprising one or more TALE repeat domains;

(d) contacting the second nucleic acid with a second enzyme, wherein the second enzyme creates a second ligatable end, and wherein the first and second ligatable ends are compatible; and

(e) ligating the first and second nucleic acids through the first and second ligatable ends to produce a first ligated nucleic acid, wherein the first ligated nucleic acid is linked to a solid support, and wherein the first ligated nucleic acid encodes a polypeptide comprising said first and second sets comprising one or more TALE repeat domains.

2. The process of claim 1, wherein the first and second enzymes are a first and second restriction endonuclease, wherein the first restriction endonuclease cleaves at a site within the first nucleic acid and creates a first cut end, and the second restriction endonuclease cleaves at a site within the second nucleic acid and creates a second cut end, and wherein the first and second ligatable ends are the first and second cut ends.

3. The process of claim 2, wherein the first ligated nucleic acid does not comprise a restriction site recognized by the first restriction endonuclease.

4. The process of any one of claims 1 to 3, further comprising:

(f) contacting the first ligated nucleic acid with a third enzyme, wherein the third enzyme creates a third ligatable end;

(g) providing a third nucleic acid comprising a sequence encoding a third set comprising one or more TALE repeat domains;

(h) contacting the third nucleic acid with a fourth enzyme, wherein the fourth enzyme creates a fourth ligatable end, and wherein the third and fourth ligatable ends are compatible; and

(i) ligating the first ligated and third nucleic acids through the third and fourth ligatable ends to produce a second ligated nucleic acid linked to the solid support, wherein the second ligated nucleic acid encodes a polypeptide comprising said first, second, and third sets comprising one or more TALE repeat domains.

5. The process of claim 4, wherein the third and fourth enzymes are a third and fourth restriction endonuclease, wherein the third restriction endonuclease cleaves at a site within the first ligated nucleic acid and creates a third cut end, and the fourth restriction endonuclease cleaves at a site within the third nucleic acid and creates a fourth cut end, and wherein the third and fourth ligatable ends are the third and fourth cut ends.

6. The process of claim 5, wherein the ligated nucleic acid does not comprise a restriction site recognized by the first endonuclease, and wherein the first and third restriction endonucleases are the same.

7. The process of claim 5 or 6, wherein the second and fourth restriction endonucleases are the same.

8. The process of any one of claims 4 to 7, further comprising:

(j) contacting the second ligated nucleic acid with a fifth enzyme, wherein the fifth enzyme creates a fifth ligatable end;

(k) providing a fourth nucleic acid comprising a sequence encoding a fourth set comprising one or more TALE repeat domains;

(l) contacting the fourth nucleic acid with a sixth enzyme, wherein the sixth enzyme creates a sixth ligatable end, and wherein the fifth and sixth ligatable ends are compatible; and

(m) ligating the second ligated and fourth nucleic acids through the fifth and sixth ligatable ends to produce a third ligated nucleic acid linked to the solid support, wherein the third ligated nucleic acid encodes a polypeptide comprising said first, second, third, and fourth sets comprising one or more TALE repeat domains.

9. The process of claim 8, wherein the fifth and sixth enzymes are a fifth and sixth restriction endonuclease, wherein the fifth restriction endonuclease cleaves at a site within the second ligated nucleic acid and creates a fifth cut end, and the sixth restriction endonuclease cleaves at a site within the fourth nucleic acid and creates a sixth cut end, and wherein the fifth and sixth ligatable ends are the fifth and sixth cut ends.

10. The process of claim 9, wherein the second ligated nucleic acid does not comprise a restriction site recognized by the first endonuclease, and wherein the first, third, and fifth restriction endonucleases are the same.

11. The process of claim 9 or 10, wherein the second, fourth, and sixth restriction endonucleases are the same.

12. The process of claim 1, wherein the second set comprises one to four TALE repeat domains.

13. The process of any one of claims 1-11, wherein the first and second ligatable ends each comprise an overhang of 1-10 nucleotides; or the first enzyme is a type IIS restriction endonuclease.

14. The process of any one of claims 1-11, further comprising unlinking the first ligated nucleic acid from the solid support and inserting the first ligated nucleic acid into a vector; unlinking the second ligated nucleic acid from the solid support and inserting the second ligated nucleic acid into a vector; unlinking the third ligated nucleic acid from the solid support and inserting the third ligated nucleic acid into a vector.

15. The process of claim 14, wherein the vector is an expression vector.

16. The process of claim 15, wherein the expression vector includes a sequence encoding an effector domain, and wherein the first, second, or third ligated nucleic acid is inserted into the vector such that the vector comprises a sequence encoding a fusion protein of the polypeptide and the effector domain.

17. The process of claim 16, wherein the effector domain is a nuclease domain.

18. The process of claim 16 or 17, further comprising inserting the expression vector into a cell.

19. The process of any one of claims 1 to 18, wherein the first nucleic acid comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 58-62; and the second nucleic acid comprises a β unit, a γ unit, a δ unit, and a ε unit, wherein the β unit comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 63-67, the γ unit comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 68-72, the δ unit comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 73-77, and the ε unit comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 78-82.

20. The process of any one of claims 1 to 18, wherein the first nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 58-62; and the second nucleic acid comprises a β unit, a γ unit, a δ unit, and a ε unit, wherein the β unit comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63-67, the γ unit comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 68-72, the δ unit comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 73-77, and the ε unit comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 78-82.

21. The process of any one of claims 4 to 11 and 13 to 20, wherein the third nucleic acid comprises a β unit, a γ unit, a δ unit, and a ε unit, wherein the β unit comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 63-67, the γ unit comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 68-72, the δ unit comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 73-77, and the ε unit comprises a nucleotide sequence encoding a protein encoded by a sequence selected from the group consisting of SEQ ID NOs: 78-82.

22. The process of any one of claims 4 to 11 and 13 to 20, wherein the third nucleic acid comprises a β unit, a γ unit, a δ unit, and a ε unit, wherein the β unit comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63-67, the γ unit comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 68-72, the δ unit comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 73-77, and the ε unit comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 78-82.

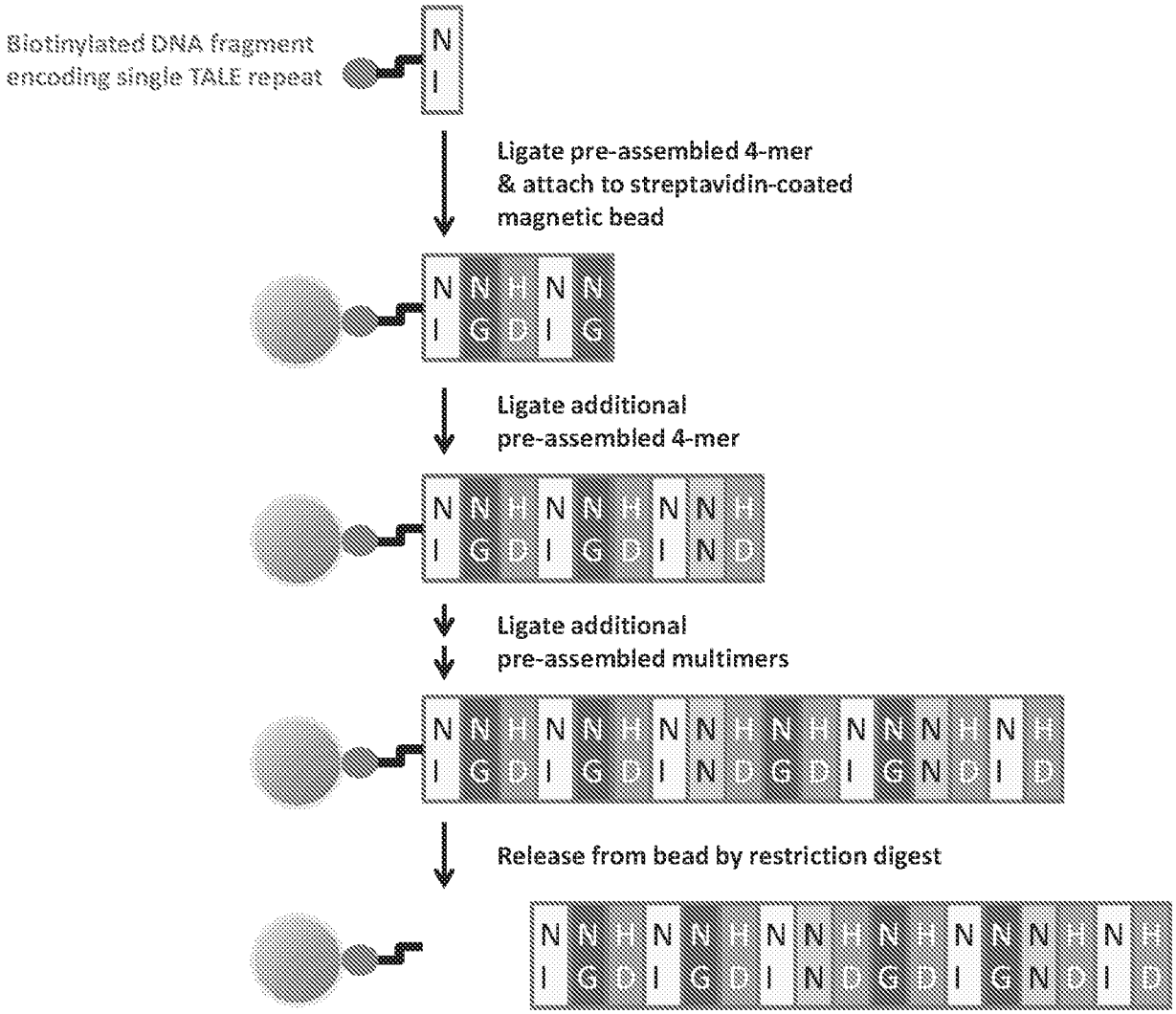

Biotinylated DNA fragment encoding single TALE repeat
Ligate pre-assembled 4-mer & attach to streptavidin-coated magnetic bead
Ligate additional pre-assembled 4-mer
Ligate additional pre-assembled multimers
Release from bead by restriction digest

FIG. 1

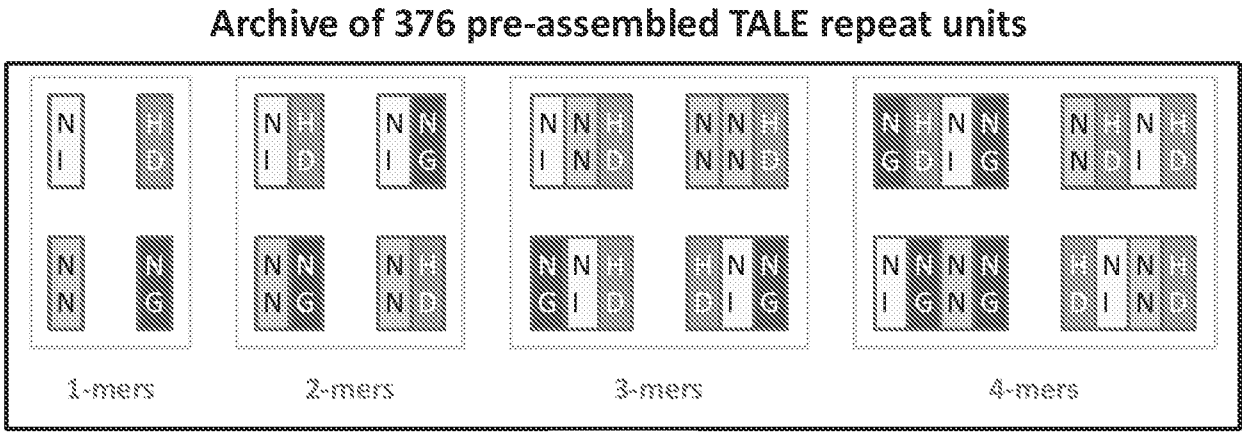

Archive of 376 pre-assembled TALE repeat units
1-mers
2-mers
3-mers
4-mers

FIG. 2

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGT
CTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGG
CTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGC
ACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAA
GGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATT
AAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCG
GTACCTCGCGAATGCATCTAGATATCGGATCCCGGGCCCGTCGACTGCAGAGGCCTGCATGCAAGCTT
GGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGC
GGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCG
TTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCT
GGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGC
GAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC
CCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGA
AGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTT
TTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAG
TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCG
TTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCC
CCAGTGCTGCAATGATACCGCGAGAgCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCA
TCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTT
ACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAA
GTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCG
AGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCAT
CATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGT
AACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAA
ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTT
CCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTA
TTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAA
ACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC
(SEQ ID NO:1)

FIG. 3

## FIG 4A

| Type | SEQUENCE | SEQ ID NO: |
|---|---|---|
| α/ε | LTP***D***QVVAIASNIGGKQALETVQRLLPVLCQ***D***HG | 2 |
| β | LTP***E***QVVAIASNIGGKQALETVQRLLPVLCQ***A***HG | 3 |
| γ | LTP***D***QVVAIASNIGGKQALETVQRLLPVLCQ***A***HG | 4 |
| δ | LTP***A***QVVAIASNIGGKQALETVQRLLPVLCQ***D***HG | 5 |

## FIG 4B

| Type | SEQUENCE | SEQ ID NO: |
|---|---|---|
| α/ε | CTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGG GGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCG GTCCTTTGTCAAGACCACGGC | 6 |
| Type β: | CTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGG | 7 |
| Type γ: | CTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGA GGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGT | 8 |
| Type δ: | TTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCG GTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACT GTGCCAGGATCATGGA | 9 |

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTA
TCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGAC
CGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTG
ACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTA
CATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACAT
CAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACA
TGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG
TACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGT
GTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAAT
ACGACTCACTATAGGGAGACCCAAGCTGGCTAGCaccATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGA
CATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCT
ATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGC
GCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCT
TGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCG
GTAAACAGTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCA
GCTCGACACCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCG
CAATGCGCTCACCGGGGCCCCCTTGAACAGAGACGATTAATGCGTCTCGCTGACACCCGAACAGGTGGTCGCCATTGC
T**XXXXXXXXX**GGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTT
AACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCA
TGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCAACTAGT
CAAAAGTGAACTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTAAT
TGAAATTGCCAGAAATTCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATAT
AGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTG
ATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTCGAA
GAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAGT
TTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAATGGA
GCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGA
CGGAAATTTAATAACGGCGAGATAAACTTTTAAGGGCCCTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCG
ATTCTACGCGTACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC
CAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAAT
GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGG
AGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTG
GGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTG
ACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCC
CGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATT
AGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAA
TAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGA
TTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTA
GGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGT
GTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTC
(SEQ ID NO:10)

FIG. 5A

FIG. 5B

AGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGC
TGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTT
TTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGTTGACA
ATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGCCTTTGTCTC
AAGAAGAATCCACCCTCATTGAAAGAGCAACGGCTACAATCAACAGCATCCCCATCTCTGAAGACTACAGCGTCGCC
AGCGCAGCTCTCTCTAGCGACGGCCGCATCTTCACTGGTGTCAATGTATATCATTTTACTGGGGGACCTTGTGCAGAA
CTCGTGGTGCTGGGCACTGCTGCTGCTGCGGCAGCTGGCAACCTGACTTGTATCGTCGCGATCGGAAATGAGAACA
GGGGCATCTTGAGCCCCTGCGGACGGTGTCGACAGGTGCTTCTCGATCTGCATCCTGGGATCAAAGCGATAGTGAA
GGACAGTGATGGACAGCCGACGGCAGTTGGGATTCGTGAATTGCTGCCCTCTGGTTATGTGTGGGAGGGCTAAGCA
CTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCT
TCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCC
AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCAC
TGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTT
GGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAG
CATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA
GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG
AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTC
CTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACG
CTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACC
GCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACAC
TAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAA
GAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG
ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAA
ACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTG
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGA
GACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCC
AACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCA
GAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAA
GATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGG
CGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCA
TCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGG
ATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGT
C (SEQ ID NO:11)

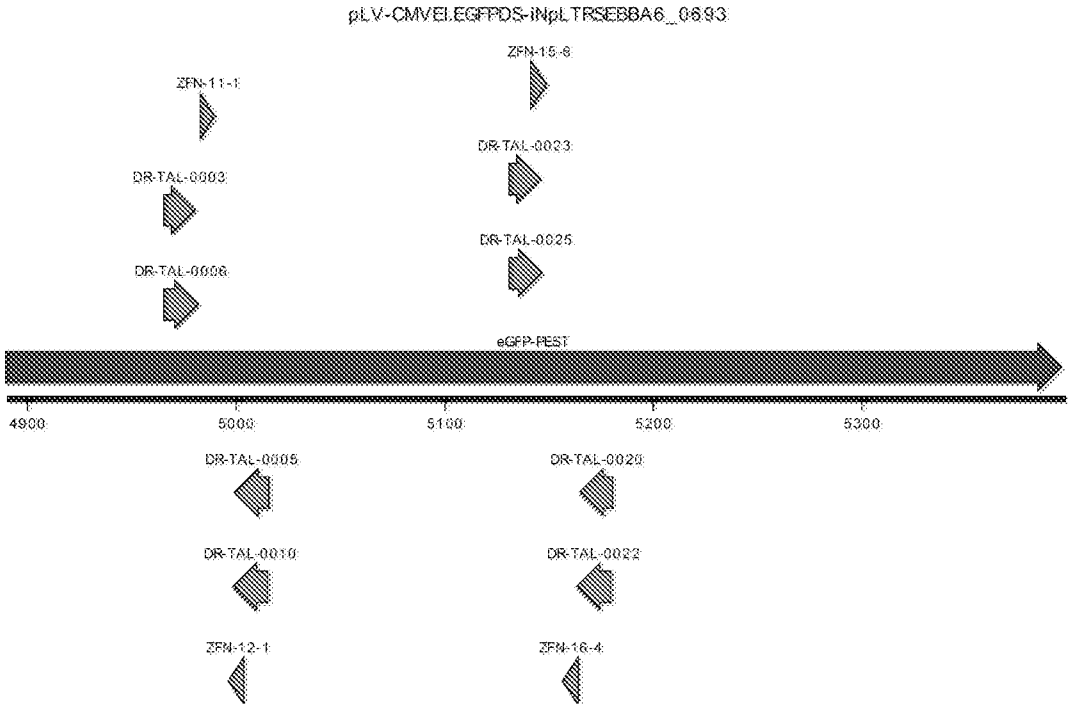

pLV-CMVELEGFPDS-iNpLTRSEBBA6_0693
ZFN-15-6
ZFN-11-1
DR-TAL-0023
DR-TAL-0003
DR-TAL-0025
DR-TAL-0006
eGFP-PEST
4900
5000
5100
5200
5300
DR-TAL-0005
DR-TAL-0020
DR-TAL-0010
DR-TAL-0022
ZFN-12-1
ZFN-16-4

FIG. 6

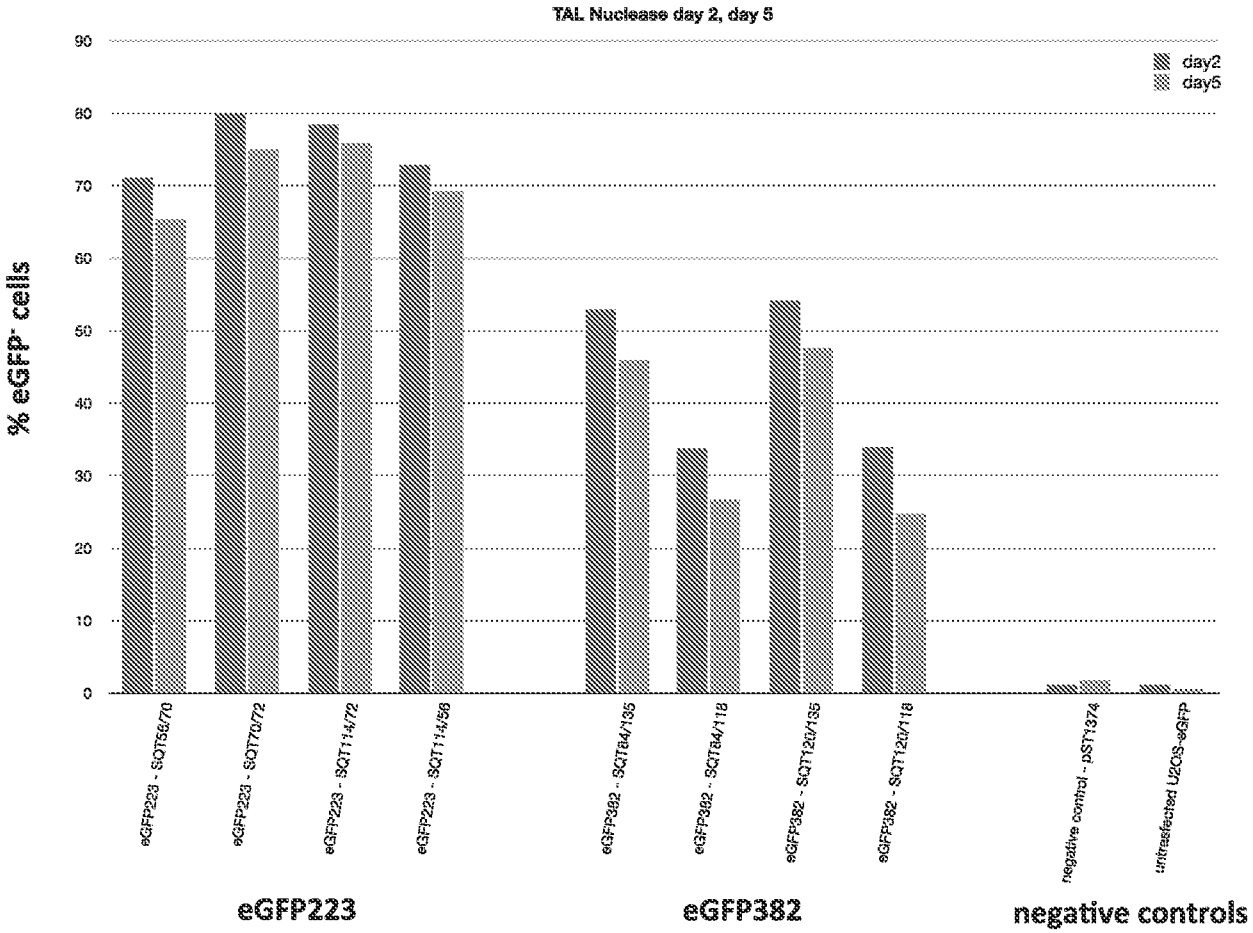

TAL Nuclease day 2, day 5
day2
day5
% eGFP- cells
90
80
70
60
50
40
30
20
10
0
eGFP223 - SQT56/70
eGFP223 - SQT70/72
eGFP223 - SQT114/72
eGFP223 - SQT114/56
eGFP382 - SQT84/135
eGFP382 - SQT84/118
eGFP382 - SQT120/135
eGFP382 - SQT120/118
negative control - pST1374
untrasfected U2OS-eGFP
eGFP223
eGFP382
negative controls

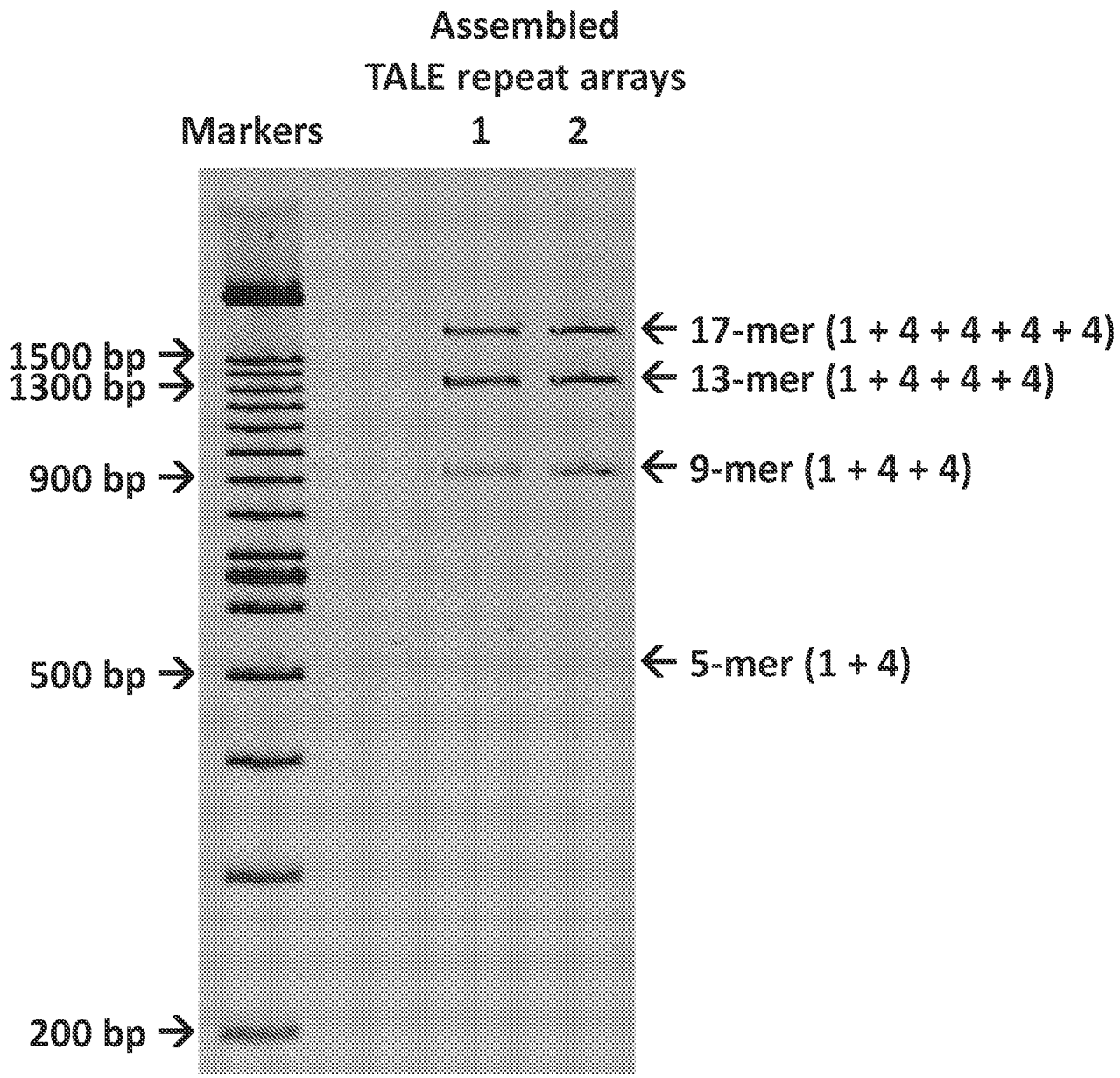
FIG. 7

14 of 30 sequences mutated. (46%)

| Sequence | Δ SIZE | SEQ ID NO: |
|---|---|---|
| CTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | WT | 12 |
| CTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA---CATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ3 (3X) | 13 |
| CTACGGCGTGCAGTGCTTCAGCCG----------AAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ16 | 14 |
| CTACGGCGTGCAGTGCTTCAGCCGCTAC----------GACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ21 | 15 |
| CTACGGCGTGCAGTGCTTCAGCCGC----------ACGACTTCTTCCAGTCCGCCATGCCCGAAGGCTACNTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ22 | 16 |
| CTACGGCGTGCAGTGCTTCAGC----------ACTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ30 (Δ31 and +1) | 17 |
| CTACGGCGTGCAGTGCTTCAG----------CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ33 | 18 |
| CTACGGCGTGCAG----------CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ33 | 19 |
| CTACGGCGTGCAGTGCTT----------GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ51 | 20 |
| CTACGGCGTGCAGTGCTTCAGCC----------GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ54 | 21 |
| CTACG----------CCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | Δ60 | 22 |
| CTACGGCGTGCAGTGCTT----------CTTCAAGGA | Δ84 | 23 |
| CTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA | +2 | 24 |

FIG. 8

FIG. 9


Assembled
TALE repeat arrays
Markers
1
2
1500 bp
1300 bp
900 bp
500 bp
200 bp
17-mer (1 + 4 + 4 + 4 + 4)
13-mer (1 + 4 + 4 + 4)
9-mer (1 + 4 + 4)
5-mer (1 + 4)

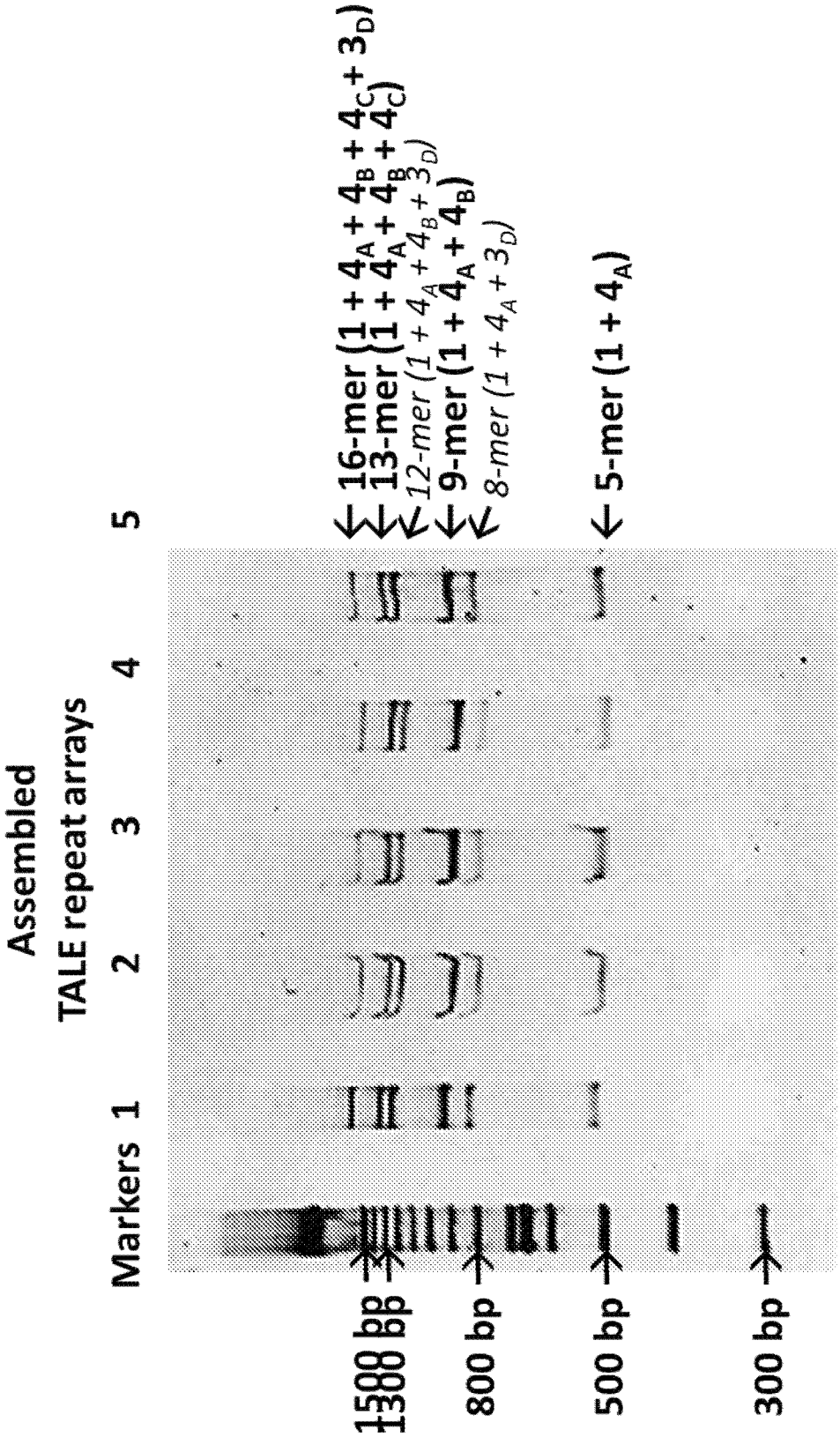

Markers
1
2
3
4
5
Assembled TALE repeat arrays
16-mer (1 + 4A + 4B + 4C + 3D)
13-mer (1 + 4A + 4B + 4C)
12-mer (1 + 4A + 4B + 3D)
9-mer (1 + 4A + 4B)
8-mer (1 + 4A + 3D)
5-mer (1 + 4A)
1500 bp
1300 bp
800 bp
500 bp
300 bp

FIG. 10

## FIG. 11A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCATCCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTC
GCAATCGCGTCAAACGGAGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCA
GAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGCATGACGGA
GGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAA
GTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAGCCCTGGAAAC
CGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCCAC
GACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCG
ATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGT
TGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGG
AAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGT
CACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTA
CACCGGAGCAAGTCGTGGCCATTGCATCCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCC
CAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAG
CATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTCTGACACCCGAACAGGTGGTCGCCAT
TGCTTCCCACGACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCT
GCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGT
CTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGAT
CC

(SEQ ID NO:25)

## FIG. 11B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGG
GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPEQVVAI
ASHDGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS

(SEQ ID NO:26)

FIG. 12A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCATCCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTC
GCAATCGCGTCAAACGGAGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCA
GAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGCATGACGGA
GGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAA
GTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAGCCCTGGAAAC
CGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCCAC
GACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCG
ATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGT
TGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGG
AAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGT
CACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTA
CACCGGAGCAAGTCGTGGCCATTGCATCCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCC
CAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAG
CATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCAT
CGCCAGCCATGATGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATG
GACTGACACCCGAACAGGTGGTCGCCATTGCTTCTAATGGGGGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCC
AATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGAC
GACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTC
CCGAGAGAACTTCCCATCGAGTCGCGGGATCC
(SEQ ID NO:27)

FIG. 12B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGG
GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAI
ASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPAL
DAVKKGLPHAPALIKRTNRRIPERTSHRVAGS
(SEQ ID NO:28)

## FIG. 13A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTC
GCAATCGCGAACAATAATGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGAC
CACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCCACGA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTG
TGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGAAC
AATAATGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACA
CCGGAGCAAGTCGTGGCCATTGCATCCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCA
GTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCG
CCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGAC
TGACACCCGAACAGGTGGTCGCCATTGCTTCCCACGACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAAT
TGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGAC
CCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCG
AGAGAACTTCCCATCGAGTCGCGGGATCC
(SEQ ID NO:29)

## FIG. 13B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQV
VAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGL
TPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIAN
NNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDA
VKKGLPHAPALIKRTNRRIPERTSHRVAGS
(SEQ ID NO:30)

FIG. 14A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTC
GCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGAC
CACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTG
TGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCA
CATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACA
CCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCA
GTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCG
CCAGCCATGATGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGAC
TGACACCCGAACAGGTGGTCGCCATTGCTTCTAATGGGGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAAT
TGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGAC
CCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCG
AGAGAACTTCCCATCGAGTCGCGGGATCC
(SEQ ID NO:31)

FIG. 14B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET
VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVV
AIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
QALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRL
LPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG
LTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIA
SHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD
AVKKGLPHAPALIKRTNRRIPERTSHRVAGS
(SEQ ID NO:32)

FIG. 15A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGTCACATGACGGGGGAAAGCAAGC
CCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATT
GCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGG
CTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTC
CTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGC
AGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCG
CAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACC
ACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTG
TGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCA
CATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACA
CCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCA
GTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTCTGACACCCGAACAGGTGGTCGCCATTG
CTTCCCACGACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTG
CGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTC
TGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATC
C

(SEQ ID NO: 33)

FIG. 15B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR
LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIA
SHDGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS
(SEQ ID NO:34)

FIG. 16A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGTCACATGACGGGGGAAAGCAAGC
CCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATT
GCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGG
CTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTC
CTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGC
AGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCG
CAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACC
ACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTG
TGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCA
CATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACA
CCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCA
GTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCG
CCAGCCATGATGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGAC
TGACACCCGAACAGGTGGTCGCCATTGCTAATAATAACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAAT
TGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGAC
CCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCG
AGAGAACTTCCCATCGAGTCGCGGGATCC
(SEQ ID NO:35)

FIG. 16B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR
LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIA
SHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD
AVKKGLPHAPALIKRTNRRIPERTSHRVAGS (SEQ ID NO:36)

FIG. 17A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTC
GCAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGAC
CACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGCATGACGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTCGCCATCGCCAGCCATGATGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTG
TGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCA
AACGGAGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTAC
ACCGGAGCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCC
AGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGC
ATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATC
GCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGG
ACTGACACCCGAACAGGTGGTCGCCATTGCTAATAATAACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCA
ATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACG
ACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCC
GAGAGAACTTCCCATCGAGTCGCGGGATCC (SEQ ID NO:37)

FIG. 17B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR
LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHG
LTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIA
SNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD
AVKKGLPHAPALIKRTNRRIPERTSHRVAGS (SEQ ID NO:38)

FIG. 18A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACG
GGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGC
TCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAGCCATGATGGCGGTAA
GCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGT
CGCAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCA
GAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGCATGACGGA
GGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAA
GTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAGCCCTGGAAAC
CGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAAT
AACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCG
ATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGT
TGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGG
AAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGA
ACAATAATGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTA
CACCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCC
CAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAG
CATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCAT
CGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATG
GACTGACACCCGAACAGGTGGTCGCCATTGCTTCCCACGACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCC
AATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGAC
GACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTC
CCGAGAGAACTTCCCATCGAGTCGCGGGATCC (SEQ ID NO:39)

FIG. 18B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRL
LPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHG
LTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIA
NNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD
AVKKGLPHAPALIKRTNRRIPERTSHRVAGS (SEQ ID NO:40)

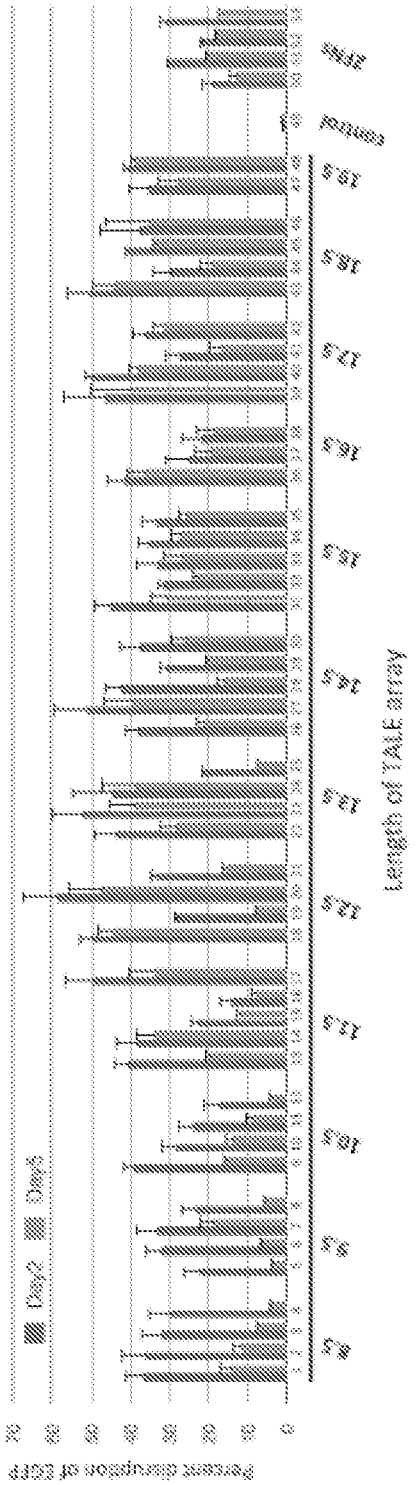



FIG. 19A

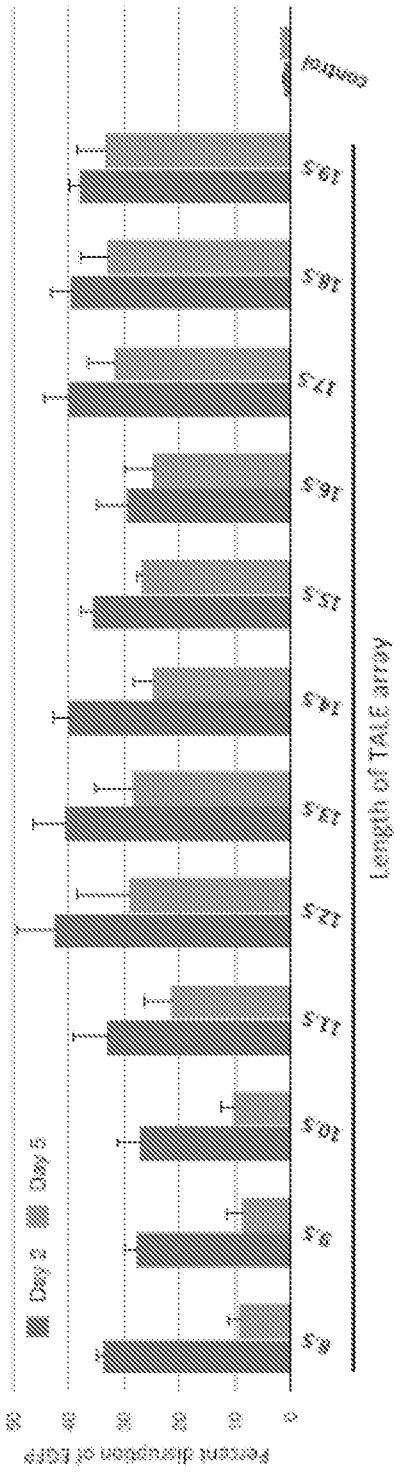



FIG. 19B

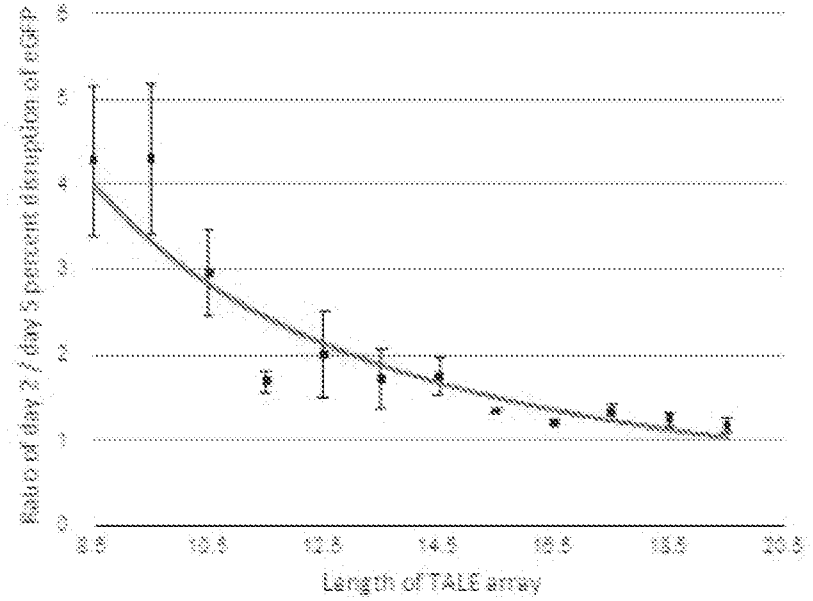



FIG. 20A

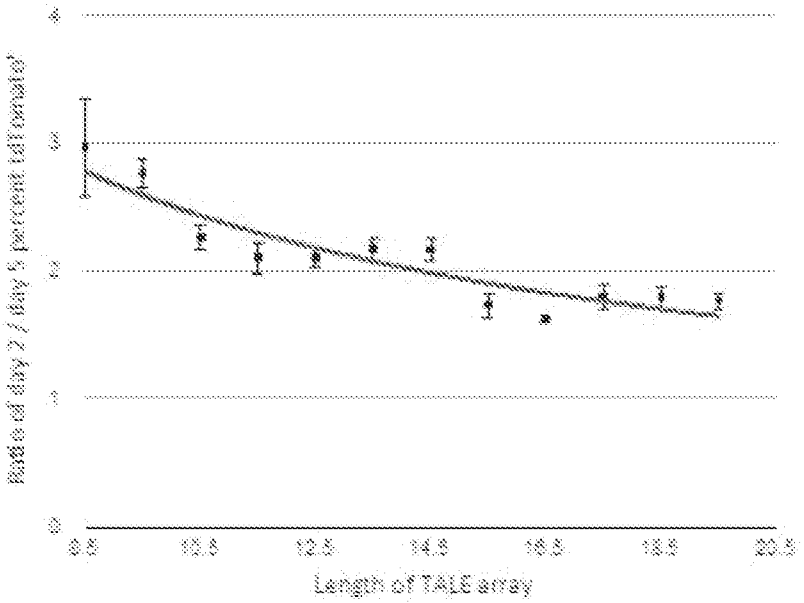



FIG. 20B

| SEQUENCE | Δ SIZE | SEQ ID NO: |
|---|---|---|
| ***AXIN2*** | | |
| **TAL2200/TAL2201** | | |
| **Mutations in 2 of 83 sequences ≈ 2.4%** | | |
| TTCCCAGACTCAGTGGGAAGAGCTCCCTCACCATGAGTAGGCGCTATGTTGGTGACTTGCCTCCCGGACCCCAGCAGCAGCTTCCGTGA | WT | 301. |
| TTCCCAGACTCAGTGGGAAGAGCTCCCTCACCATGAGTAGGCGCTATG----------TTGCCTCCCGGACCCCAGCAGCAGCTTCCGTGA | Δ8 | 302. |
| TTCCCAGACTCAGTGGGAAGAGCTCCCTCA----------CCCGGACCCCAGCAGCAGCTTCCGTGA | Δ31 | 303. |
| | | |
| ***BRCA1*** | | |
| **TAL2384/TAL2385** | | |
| **Mutations in 7 of 14 sequences ≈ 50.0%** | | |
| GGCGTGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTTTGACGGGGGGTAGGGGCGGAACCTGAGAGGCGTAAGGCGTTGTG | WT | 304. |
| GGCGTGGGAGAGTGGATTTCCGAAGCTGACAGAT----------GGGTAGGGGCGGAACCTGAGAGGCGTAAGGCGTTGTG | Δ17 | 305. |
| GGCGTGGGAGAGTGGATTTCCGAAGCTGACAGA----------GGGGCGGAACCTGAGAGGCGTAAGGCGTTGTG | Δ23 | 306. |
| GGCGTGGGAGAGTGGATTTCCGAAGCTGACAGAT----------GGAACCTGAGAGGCGTAAGGCGTTGTG | Δ27 | 307. |
| GGCGTGGGAGAGTGGATTTCCGAAGCTGACAGATG----------GGGTAAGGCGTTGTG | Δ38 | 308. |
| GGCGTGGGAGAGTGGATTTCCGAAGCTGACAGATG----------CGTAAGGCGTTGTG | Δ39 | 309. |
| | | |
| AGTGGATTTCCGAAGCTGACAGATG----------/ /-------tggaagaagaGGCTGGTCATGAGGTCAGGAGTTCC | Δ183 (Δ193 +10) | 310. |
| CGCTCAGGAGGCCTTCACCCTCTGCTCTGGGTAAAGG------/ /----------AACTGGAATATGCCTTGAGGGGG | Δ235 | 311. |
| | | |
| ***CHD7*** | | |
| **TAL2238/TAL2239** | | |
| **Mutations in 4 of 81 sequences ≈ 4.9%** | | |
| CCTGAGCTGTGGTTTGGAGGAGCCGTGTGTTGGAAGAAGATGGCAGATCCAGGAATGATGAGTCTTTTTGGCCAGGATGGCAATATTT | WT | 312. |
| CCTGAGCTGTGGTTTGGAGGAGCCGTGTGTTGCAAGAAGATca----------GAATGATGAGTCTTTTTGGCGAGGATGGCAATATTT | Δ9 (Δ11 +2) | 313. |
| CCTGAGCTGTGGTTTGGAGGAGCCGTGTGTTGGAAGAAGATG----------GAATGATGAGTCTTTTTGGCGAGGATGGCAATATTT | Δ10 | 314. |
| CCTGAGCTGTGGTTTGGAGGAGCCGTGTGTTGGAAGAAGAT----------GGCGAGGATGGCAATATTT | Δ28 | 315. |
| | | |
| TAAAGAAATATGCAATGACAT----------/ /----------AAAATCCAGTAAATCCTATG | Δ199 | 316. |

FIG. 21A

| SEQUENCE | Δ SIZE | SEQ ID NO: |
|---|---|---|
| *CDC73*<br>**TAL2202/TAL2203**<br>**Mutations in 9 of 37 sequences ≈ 24.3%** | | |
| GAGGGGGGGGAAGATGGCGGACGTGCTTAGCGTCCTGCGACAGTACAACATCCAGAAGAAGGAGATTGTGGTGAAGGGAGACGAAGTG | WT | 317. |
| GAGGGGGGGGAAGATGGCGGACGTGCTTAGCGTCCTGCGACAGTACA-------GAACAAGGAGATTGTGGTGAAGGGAGACGAAGTG | Δ7 | 318. |
| GAGGGGGGGGAAGATGGCGGACGTGCTTAGCGTCCTGCGACAGTACAAC--------GAAGGAGATTGTGGTGAAGGGAGACGAAGTG | Δ8 | 319. |
| GAGGGGGGGGAAGATGGCGGACGTGCTTAGCGTCCTGCGACA--------TCCAGAAGAAGGAGATTGTGGTGAAGGGAGACGAAGTG | Δ8 | 320. |
| GAGGGGGGGGAAGATGGCGGACGTGCTTAGCGTCCTGCGACAG----------AAGAAGGAGATTGTGGTGAAGGGAGACGAAGTG | Δ12 (2x) | 321. |
| GAGGGGGGGGAAGATGGCGGACGTGCTTAGCGTCCTGCGACAGT----------ATTGTGGTGAAGGGAGACGAAGTG | Δ20 | 322. |
| GAGGGGGGGGAAGATGGCGGACGTGCTTAGCGTCCTGCGACAG----------TTGTGGTGAAGGGAGACGAAGTG | Δ22 | 323. |
| GAGGGGGGGGAAGATGGCGGACGTGCTTAGCGTCCTGCGACAGTACAACAACATCCAGAAGAAGGAGATTGTGGTGAAGGGAGACGAA | +3 | 324. |
| AGGGCGAGGCGACAAGAGAAGAAGGAGGCAGG----------/ /----------ATTGTGGTGAAGGGAGACGAAGTG | Δ101 | 325. |
| | | |
| *CYLD*<br>**TAL2386/TAL2387**<br>**Mutations in 10 of 44 ≈ 22.7%** | | |
| TTAGTATTTTGAAGTTAATATCACAATGAGTTCAGGCTTATGGAGCCAAGAAAAAGTCACTTCACCCTACTGGGAAGAGCGGATTTTT | WT | 326. |
| TTAGTATTTTGAAGTTAATATCACAATGAGTTCAGGCTTATGGA---------AAGTCACTTCACCCTACTGGGAAGAGCGGATTTTT | Δ9 | 327. |
| TTAGTATTTTGAAGTTAATATCACAA----------GTCACTTCACCCTACTGGGAAGAGCGGATTTTT | Δ29 | 328. |
| TTAGTATTTTGAAGTTAATA----------TCACCCTACTGGGAAGAGCGGATTTTT | Δ41 | 329. |
| TTAGTATTTTGAAGTT----------CCCTACTGGGAAGAGCGGATTTTT | Δ48 | 330. |
| T----------AAGTCACTTCACCCTACTGGGAAGAGCGGATTTTT | Δ52 | 331. |
| TTAGTATTT----------CCCTACTGGGAAGAGCGGATTTTT | Δ55 | 332. |
| TTAGTATTTTGAAGTT----------GAAGAGCGGATTTTT | Δ57 | 333. |
| TTAGTATTTTGAA----------CGGATTTTT | Δ66 | 334. |
| TTAG---------- | Δ84 | 335. |
| | | |
| TAATATCACAATGACTTCtcaggggcacaccaggccccaggggaacaccaggccctgggtaagcatgcagtcccaggtggacatcaggtgccaggaggAAAAGTCACTTCACCCTA | +62 (Δ18 + 80) | 336. |

FIG. 21B

| SEQUENCE | Δ SIZE | SEQ ID NO: |
|---|---|---|
| ***NCOR2*** **TAL2284/TAL2285** **Mutations in 7 of 88 ≈ 8.0%** | | |
| TCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGCCCCGCTACCCGCCCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGA | WT | 337. |
| TCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGCCCCGCTA-CCGCCCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGA | Δ1 | 338. |
| TCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGCCCCGC-------CCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGA | Δ7 | 339. |
| TCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGC----------CCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGA | Δ12 | 340. |
| TCCACAC----------AGATCGCCCGGA | Δ69 | 341. |
| | | |
| CCCCCACCACC----------/ /----------TTTCCTACCCAGTGCAGATCGCCCGGA | Δ70 | 342. |
| GCTTATTGCGGC----------/ /----------CTTTCTCCTGGGAGCCAGGCT | Δ213 | 343. |
| TCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGCCCCGCTACCaccCGCCCCACAGCCTTTCCTACCCAGTGCAGATCGCCC | +3 | 344. |
| | | |
| ***JAK2*** **TAL2406/TAL2407** **Mutations in 13 of 21 sequences ≈ 61.9%** | | |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTGCATGGGAATGGCCTGCCTTACGATGACAGAAATGGAGGGAACATCCACCTCT | WT | 345. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTGCATG----GGGCCTGCCTTACGATGACAGAAATGGAGGGAACATCCACCTCT | Δ4 | 346. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTGCATG~~~~GGCCTGCCTTACGATGACAGAAATGGAGGGAACATCCACCTCT | Δ5 | 347. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTGCATGGGAATGG------TTACGATGACAGAAATGGAGGGAACATCCACCTCT | Δ6 | 348. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTGCA--------GCCTGCCTTACGATGACAGAAATGGAGGGAACATCCACCTCT | Δ8 | 349. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTGC----------CTTACGATGACAGAAATGGAGGGAACATCCACCTCT | Δ15 (2x) | 350. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTGCATGGG----------AATGGAGGGAACATCCACCTCT | Δ24 | 351. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGAC----------GATGACAGAAATGGAGGGAACATCCACCTCT | Δ25 | 352. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTG----------CAGAAATGGAGGGAACATCCACCTCT | Δ26 | 353. |
| TTTCTCTTACAGGC----------ATGGAGGGAACATCCACCTCT | Δ53 | 354. |
| T----------ACGATGACAGAAATGGAGGGAACATCCACCTCT | Δ54 | 355. |
| TTTCTCTTACAGGCAAATGTTCTGAAAAAGACTCTGCATGG----agtaagtCTTACGATGACAGAAATGGAGGGAACATCCACCTCT | Δ4 (Δ11 +7) | 356. |
| CGGAGGTTTGCTGCAaacagagaaattt----------/ /----------CGATGACAGAAATGGAGGGAACATCCACCTCT | Δ288 (Δ301 +13) | 357. |

FIG. 21C

| SEQUENCE | Δ SIZE | SEQ ID NO: |
|---|---|---|
| ***MYCN***<br>**TAL2280/TAL2281**<br>**Mutations in 12 of 35 sequences ≈ 34.3%** | | |
| CGGGAGGCGAGCCGATGCCGAGCTGCTCCACGTCCACCATGCCGGGCATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA | WT | 358. |
| CGGGAGGCGAGCCGATGCCGAGCTGCTCCACGTCCtCC------ccCATGATCTGCAGGAACCCAGACCTCGAGTTTGACTCGCTACA | Δ6 (Δ8 +2) | 359. |
| CGGGAGGCGAGCCGATGCCGAGCTGCTCCACGTC--------CGGGCATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA | Δ8 | 360. |
| CGGGAGGCGAGCCGATGCCGAGCTGCTCCACGTCCACCAT----------ATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA | Δ10 | 361. |
| CGGGAGGCGAGCCGATGCCGAGCTGCTCCACGTCCACCATG----------CTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA | Δ11 | 362. |
| CGGGAGGCGAGCCGATGCCGAGCTGCTCCACGTCCA----------TGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA | Δ12 | 363. |
| CGGGAGGCGAGCCGATGCCGAGCTGCTCCACGTCCAC----------CAAGAACCCAGACCTCGAGTTTGACTCGCTACA | Δ18 | 364. |
| CGGGAGGCGAGCCGATGCCGAGCTGCTC----------ATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA | Δ19 | 365. |
| | | |
| AAGGAAGCACCCCCGGTCTTAA----------/ /----------AGACCTCGAGTTTGACTCGCTACA | Δ100 | 366. |
| AGTGTTGGAGGTCGGCGCCGGCCCCCG----------/ /----------GGCATGATCTGCAAGAACCCAGACCTCGAGTTT | Δ122 | 367. |
| AAGGAAGCACCCCCGGTATTAA----------/ /----------ACGAAGATGACT | Δ141 | 368. |
| GAGGCGAGCCGATGCCCGTGATCTGGGTCACGGCTGCTCCAGCTTGGAGGAGAGGCGGCTCTCCCGGCGACCCTCCTCGCGCGGGCGC<br>CCCTGCCATTCCCGGGAACAGGGGCTCAGCCCCCTTTGTTAGTGCTCGTATGTCTTGGCCTGGGGAGCATTTTGGAGGCAGTGCTAGG<br>GGCAGAGAGGTCCTGTTTCCCCCAAGTCTTGATCTGCAAGAACCCA | +160(Δ29 +189)(2x) | 369. |
| | | |
| ***NBN***<br>**TAL2408/TAL2409**<br>**Mutations in 12 of 20 sequences ≈ 60%** | | |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCGATGTGGAAACTGCTGCCCGCCGCGGGCCCGGCAGGAGGTAAGGGCAGAAGGGAA | WT | 370. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCGATGTGG-----GCTGCCCGCCGCGGGCCCGGCAGGAGGTAAGGGCAGAAGGGAA | Δ5 | 371. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCGATGTG-------CTGCCCGCCGCGGGCCCGGCAGGAGGTAAGGGCAGAAGGGAA | Δ7 | 372. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCGATGTGGAAACTG----------GCCCGGCAGGAGGTAAGGGCAGAAGGGAA | Δ13 | 373. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCGATGTGGAct----------CCCGGCAGGAGGTAAGGGCAGAAGGGAA | Δ17 (Δ19 and +2) | 374. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCGATGTGGA----------CCCGGCAGGAGGTAAGGGCAGAAGGGAA | Δ19 | 375. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCGATG----------CAGGAGGTAAGGGCAGAAGGGAA | Δ28 | 376. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCGATGTGGAAA----------GTAAGGGCAGAAGGGAA | Δ28 | 377. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGACCG----------GCAGGAGGTAAGGGCAGAAGGGAA | Δ30 | 378. |
| TGCACGTCGGCCCCAGCCCTGAGGAGCCGGA----------AGGAGGTAAGGGCAGAAGGGAA | Δ35 | 379. |
| TGCACGTCGGCCC----------CGCGCGGGCCCGGCAGGAGGTAAGGGCAGAAGGGAA | Δ39 | 380. |
| TGCACGTCGGCCCCAGCC----------GGCAGGAGGTAAGGGCAGAAGGGAA | Δ45 | 381. |
| | | |
| GCTCCCGGGAGCGCGCACGTCCCGGAGCCCAT----------/ /----------GCCTGCGGGTGATTCCTGCG | Δ183 | 382. |

FIG. 21D

| SEQUENCE | Δ SIZE | SEQ ID NO: |
|---|---|---|
| ***XPC*** **- TAL2350/TAL2351**<br>**Mutations in 10 of 36 ≈ 27.8%** | | |
| TCGCGAAGTGGAATTTGCCCAGACAAGCAACATGGCTCGGAAACGCGCGGCCGGCGGGGAGCCGCGGGGACGCGAACTGCGCAGCCAG | WT | 383. |
| TCGCGAAGTGGAATTTGCCCAGACAAGCAACATGGCTCGGAGG--AACGGCCGGCGGGGAGCCGCGGGGACGCGAACTGCGCAGCCAG | Δ 2 (Δ6 and +4) | 384. |
| TCGCGAAGTGGAATTTGCCCAGACAAGCAACATGGCTCGGAAA----------GGGGAGCCGCGGGGACGCGAACTGCGCAGCCAG | Δ12 | 385. |
| TCGCGAAGTGGAATTTGCCCAGACAAGCAACATGGC----------CGGCGGGGAGCCGCGGGGACGCGAACTGCGCAGCCAG | Δ15 | 386. |
| TCGCGAAGTGGAATTTGCCCAGACAAGCAACATGGCTCG----------GGGAGCCGCGGGGACGCGAACTGCGCAGCCAG | Δ17 (x2) | 387. |
| TCGCGAAGTGGAATTTGCCCAGACAAGCAACATGGC----------GGGGAGCCGCGGGGACGCGAACTGCGCAGCCAG | Δ19 | 388. |
| TCGCGAAGTGGAATTTGCCCAGACAAGCAACATGTT----------GCGAACTGCGCAGCCAG | Δ35 (Δ37 and +2) | 389. |
| TCGCGAAGTGGAATTTGCCCAGACAAGCAACATGGCTCG----------GCGCAGCCAG | Δ39 | 390. |
| TCGCG----------CAGCCAG | Δ76 | 391. |
| | | |
| TCGCGAAGTGG----------/ /----------AATCCAAGGCCAAGAGCAAG | Δ78 | 392. |
| | | |
| ***ERCC2*** **- TAL2306/TAL2307**<br>**Mutations in 19 of 20 sequences ≈ 95.0%** | | |
| GTCGACCCCGCTGCACAGTCCGGCCGGCGCCATGAAGTGAGAAGGGGGCTGGGGGTCGCGCTCGCTAGCGGGCGCGGGGGGGTCTTGAA | WT | 393. |
| GTCGACCCCGCTGCACAGTCCGGCCGGCGCCATGAAGT-------GGGCTGGGGGTCGCGCTCGCTAGCGGGCGCGGGGGGGTCTTGAA | Δ7 | 394. |
| GTCGACCCCGCTGCACAGTCCGGCCGGCGCCATGAA----------GCTGGGGGGTCGCGCTCGCTAGCGGGCGCGGGGGGGTCTTGAA | Δ11 | 395. |
| GTCGACCCCGCTGCACAGTCCGGCCGGCGC----------TCGCTAGCGGGCGCGGGGGGGTCTTGAA | Δ31 | 396. |
| GTCGACCCCGCTGCACAGTCCGGCCGGCGCCATGAAGT----------GGGGTCTTGAA | Δ39 | 397. |
| | | |
| GGCCTCTCGCTGAATATTCATG----------/ /----------ATGGGGTCATCGGTGGGCGCG | Δ99 | 398. |
| GTCGACCCCGCTGCACAGTCCGGCCGGCGCCATGAAG-----/ /----------GGTGAAGGGGTGGGAC | Δ100 | 399. |
| GTCGACCCCGCTGCACAGTCCGGCCGGCGCCATGAAGTG---/ /----------CGTCCGGGGGTGGGACGGGGGCAGCCG | Δ99 (Δ104 + 5) | 400. |
| GTCGACCCCGCTGCACAGTCCGGCCGGCGCCATGAAGT----/ /----------GGCAGCCGCAGGGAGCAGCAGT | Δ118 | 401. |
| ATGGTCTCGTAATATAGGTGGAGCGAGCCCTCGAGG------/ /----------GGGGTCTTGAAGATG | Δ122 | 402. |
| GTCGACCCCGCTG----------/ /----------CTGAGGACCTGAGGGTTACC | Δ200 | 403. |
| GGCCTCTCGCTGAATATTCA----------/ /----------GCTCCTGAGGACCTGAGGGTTA | Δ221 | 404. |
| GTCGACCCCGCTGCACAGTCCGGC----------/ /----------CCGTCACCCTTCTCTGGGCT | Δ222 (2x) | 405. |
| GTCGACCCCGCTGCACAGTCCGGCC----------/ /----------GCTCGACGACCGGGC | Δ238 | 406. |
| GTCGACCCCGCTGCACAGTCCGGCCGGC----------/ /----------GACCGGGCACTGTGGAGG | Δ242 | 407. |
| GTCGACCCCGCTGCACAGTCCGGC----------/ /----------CGGGCACTGTGGAGG | Δ249 | 408. |
| GTCGAC----------/ /----------CCAGCAGCCCCTG | Δ313 | 409. |
| GTCGACCC----------/ /----------CCCTGCCGCCA | Δ319 | 410. |
| GGCCTCTCGCTGAATATTCATGAG----------/ /----------CCGCCAGGCTCAACGTGGA | Δ340 | 411. |

FIG. 21E

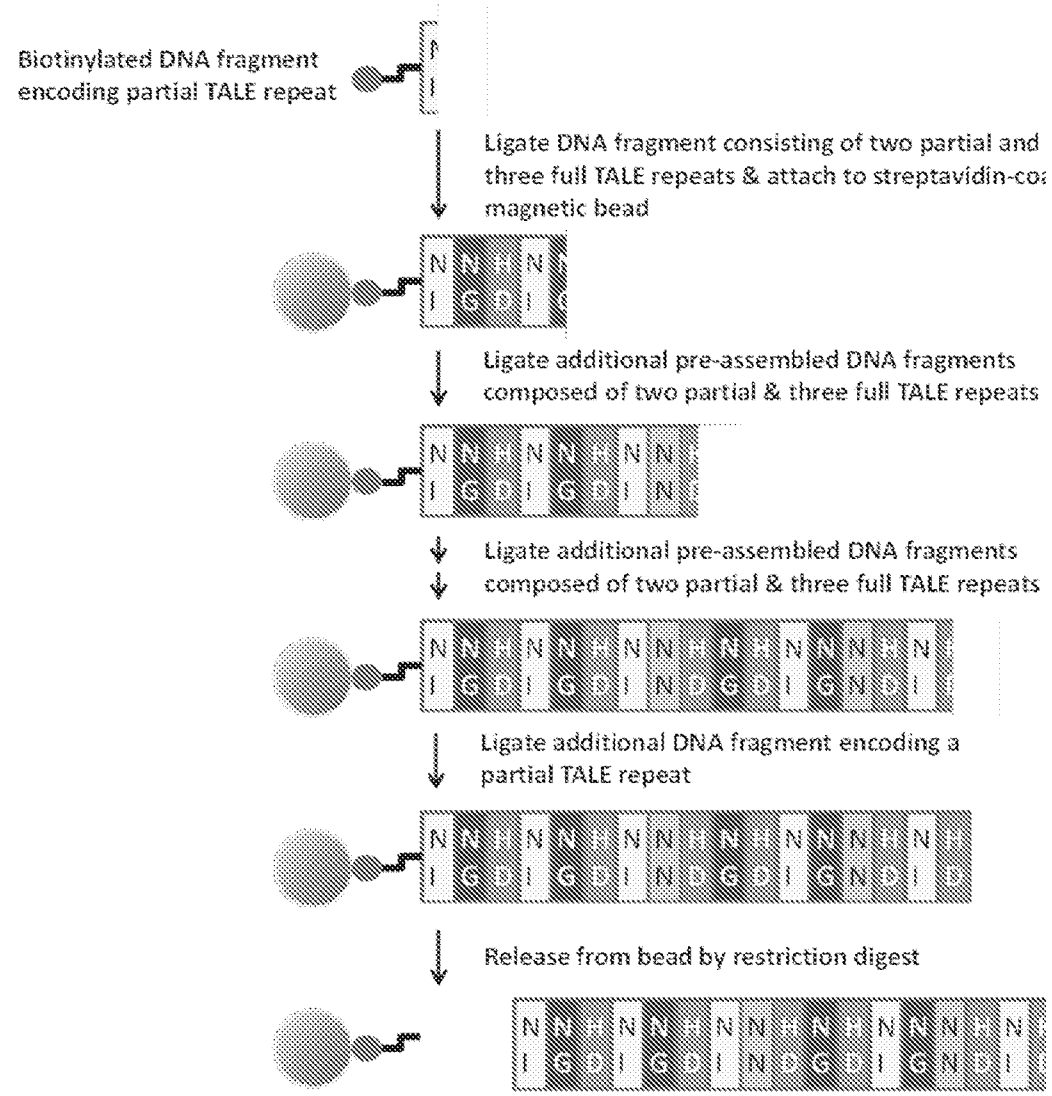
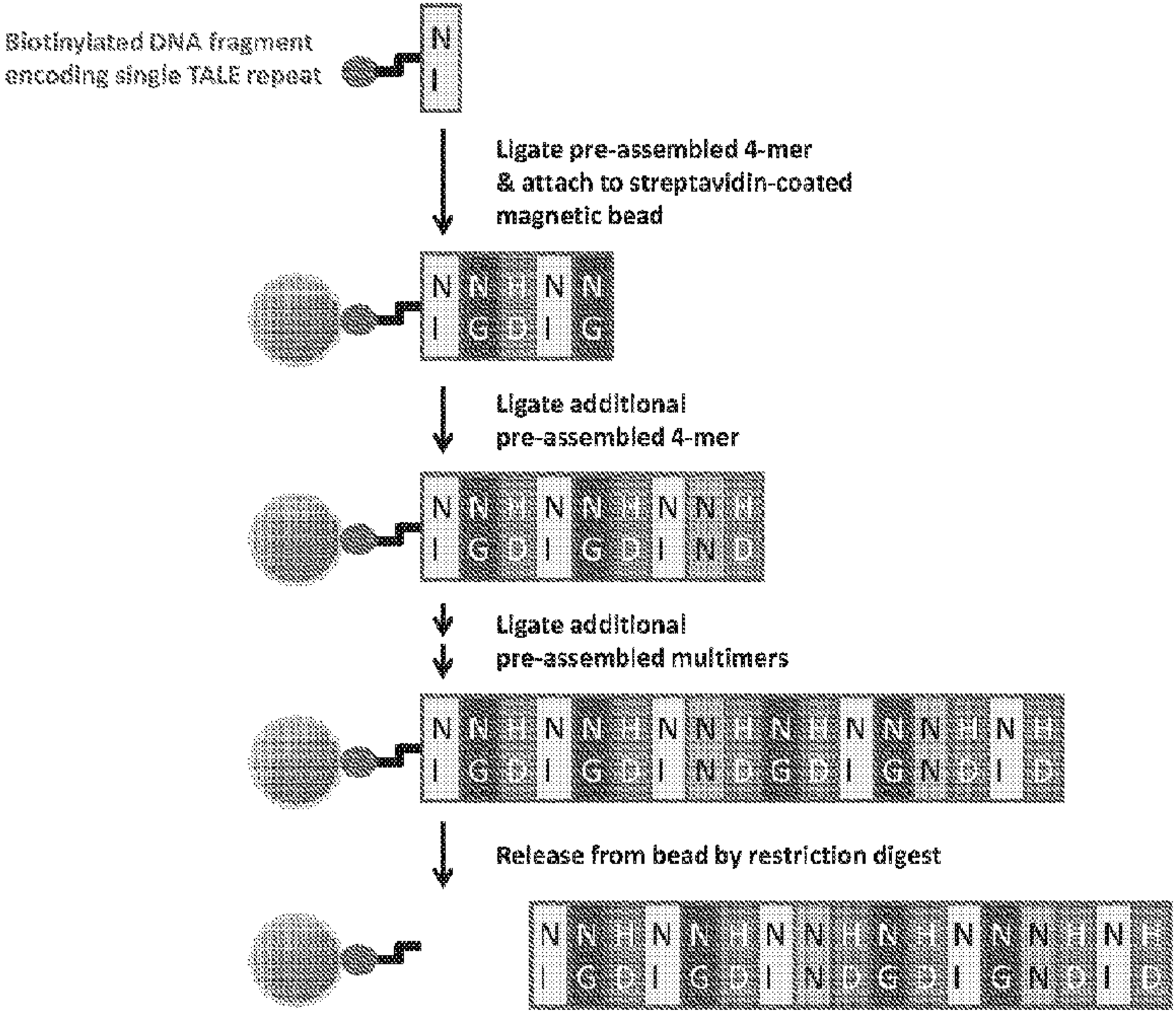

Biotinylated DNA fragment encoding partial TALE repeat
Ligate DNA fragment consisting of two partial and three full TALE repeats & attach to streptavidin-co magnetic bead
Ligate additional pre-assembled DNA fragments composed of two partial & three full TALE repeats
Ligate additional pre-assembled DNA fragments composed of two partial & three full TALE repeats
Ligate additional DNA fragment encoding a partial TALE repeat
Release from bead by restriction digest

FIG. 22


Biotinylated DNA fragment
encoding single TALE repeat
Ligate pre-assembled 4-mer
& attach to streptavidin-coated
magnetic bead
Ligate additional
pre-assembled 4-mer
Ligate additional
pre-assembled multimers
Release from bead by restriction digest