

US 20210108193A1

# (19) United States (12) Patent Application Publication (10) Pub. No.: US 2021/0108193 A1

#### Mali et al.

Apr. 15, 2021 (43) **Pub. Date:** 

#### (54) METHODS FOR SCREENING GENETIC PERTURBATIONS

- (71) Applicant: The Regents of the University of California, Oakland, CA (US)
- (72) Inventors: **Prashant Mali**, La Jolla, CA (US); Udit Parekh, La Jolla, CA (US); Yan Wu, La Jolla, CA (US); Kun Zhang, La Jolla, CA (US)
- (21) Appl. No.: 17/028,836
- (22) Filed: Sep. 22, 2020

#### **Related U.S. Application Data**

(60) Provisional application No. 62/904,614, filed on Sep. 23, 2019.

#### **Publication Classification**

(51) Int. Cl. C12N 15/10 (2006.01)C12N 5/071 (2006.01)(2006.01) C12N 15/86 A61K 35/44 (2006.01)

(52) U.S. Cl. . *C12N 15/1065* (2013.01); *C12N 5/069* (2013.01); *C12N 15/86* (2013.01); *C12N* CPC ..... 2740/15052 (2013.01); C12N 2506/45 (2013.01); C12N 2740/15043 (2013.01); A61K 35/44 (2013.01)

#### (57)ABSTRACT

Understanding the complex effects of genetic perturbations on cellular state and fitness in human pluripotent stem cells (hPSCs) has been challenging using traditional pooled screening techniques which typically rely on unidimensional phenotypic readouts. Here, Applicants use barcoded open reading frame (ORF) overexpression libraries with a coupled single-cell RNA sequencing (scRNA-seq) and fitness screening approach, a technique we call SEUSS (ScalablE fUnctional Screening by Sequencing), to establish a comprehensive assaying platform. Using this system, Applicants perturbed hPSCs with a library of developmentally critical transcription factors (TFs), and assayed the impact of TF overexpression on fitness and transcriptomic cell state across multiple media conditions. Applicants further leveraged the versatility of the ORF library approach to systematically assay mutant gene libraries and also whole gene families. From the transcriptomic responses, Applicants built genetic co-perturbation networks to identify key altered gene modules. Strikingly, we found that KLF4 and SNAI2 have opposing effects on the pluripotency gene module, highlighting the power of this method to characterize the effects of genetic perturbations. From the fitness responses, Applicants identified ETV2 as a driver of reprogramming towards an endothelial-like state.

#### Specification includes a Sequence Listing.



FIG. 17







FIG. 1F











FIG. 4











Pluripotency Network

#### METHODS FOR SCREENING GENETIC PERTURBATIONS

#### CROSS-REFERENCE TO RELATED APPLICATION

**[0001]** This application claims priority to 35 U.S.C. § 119(e) of U.S. Provisional Application Ser. No. 62/904,614, filed Sep. 23, 2019, the content of which is hereby incorporated by reference its entirety.

**[0002]** This invention was made with government support under HG009285 awarded by the National Institutes of Health. The government has certain rights in the invention.

#### SEQUENCE LISTING

**[0003]** The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Dec. 14, 2020, is named 114198-0152\_SL.txt and is 155,507 bytes in size.

#### BACKGROUND

**[0004]** Cellular reprogramming by the overexpression of transcription factors (TF), has widely impacted biological research, from the direct conversion of adult somatic cells to the induction of pluripotent stem cells, and the differentiation of hPSCs. To date, the choice of TFs that drive such reprogramming has been through a combination of the knowledge of their role in development and cellular transformation, and systematic trial-and-error. These challenges highlight the need for the development of a scalable screening method to assess the effects of TF overexpression. Such a screening method would have broad applicability in advancing a fundamental understanding of reprogramming, and as a means for the discovery of novel reprogramming factors. This disclosure addresses this need and provides related advantages as well.

#### SUMMARY

**[0005]** Described herein is a comprehensive high-throughput platform to determine an optimal method to drive the differentiation of pluripotent cells to specific somatic lineages. In some aspects, the platform utilizes a novel open reading frame (ORF) gene overexpression vector library of developmentally critical transcription factors. The platform builds genetic co-perturbation networks to identified key altered gene modules and identifies key reprogramming/ differentiation drivers from transcriptomic responses. The platform enabled identification of the key role of (previously not recognized) transcription factor ETV2 in reprogramming towards an endothelial state.

**[0006]** Thus, in one aspect, provided herein are isolated nucleic acids comprising, consisting of, or consisting essentially of (a) a nucleic acid encoding a transcription factor (TF) open reading frame (ORF); (b) a nucleic acid barcode, and (c) an optional vector comprising (a) and (b); wherein the nucleic acid barcode is located 3' to the TF ORF. In some embodiments, the TF ORF encodes a developmentally critical TF.

**[0007]** In another aspect, provided herein is a TF screening library comprising, consisting of, or consisting essentially of at least one isolated nucleic acid comprising, consisting of, or consisting essentially of (a) a nucleic acid encoding a transcription factor (TF) open reading frame (ORF); (b) a

nucleic acid barcode, and (c) an optional vector comprising (a) and (b); wherein the nucleic acid barcode is located 3' to the TF ORF. In some embodiments, the TF ORF encodes a developmentally critical TF, optionally selected from the TFs listed in Table 1.

**[0008]** In some embodiments, the TF screening library comprises, consists of, or consists essentially of at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 nucleic acids or vectors, wherein each nucleic acid or vector comprises, consists of, or consists essentially of a distinct nucleic acid encoding a TF ORF.

**[0009]** In some embodiments, the TF screening library further comprises, consists of, or consists essentially of a nucleic acid encoding a selectable marker. In some embodiments, the TF screening library further comprises, consists of, or consists essentially of a nucleic acid encoding an expression control element. In some embodiments, the expression control element is a promoter or a long terminal repeat (LTR). In some embodiments, the TF screening library further comprises, consists of, or consists essentially of a nucleic acid encoding at ranslation elongation factor, optionally wherein the translation elongation factor is Ef1a. **[0010]** In some embodiments, the vector is a retroviral vector, optionally a lentiviral vector.

**[0011]** In another aspect, provided herein is a viral packaging system comprising, consisting of, or consisting essentially of at least one isolated nucleic acid comprising, consisting of, or consisting essentially of (a) a nucleic acid encoding a transcription factor (TF) open reading frame (ORF); (b) a nucleic acid barcode, and (c) an optional vector comprising (a) and (b); wherein the nucleic acid barcode is located 3' to the TF ORF; or aTF screening library; and a packaging plasmid.

[0012] In another aspect, provided herein is a method for producing a viral particle, the method comprising, consisting of, or consisting essentially of transfecting a packaging cell line with a viral packaging system comprising, consisting of, or consisting essentially of at least one isolated nucleic acid comprising, consisting of, or consisting essentially of (a) a nucleic acid encoding a transcription factor (TF) open reading frame (ORF); (b) a nucleic acid barcode, and (c) an optional vector comprising (a) and (b); wherein the nucleic acid barcode is located 3' to the TF ORF; or aTF screening library; and a packaging plasmid under conditions suitable to package the vector or the TF screening library into a viral particle. In another aspect, also provided herein is a viral particle produced by this method, and optionally a carrier. In another aspect, also provided herein is an isolated cell comprising a nucleic acid, vector, or particle as described herein, and optionally a carrier.

**[0013]** In another aspect, provided herein is a kit comprising, consisting of, or consisting essentially of at least one of (a) a nucleic acid or vector according to any of the embodiments described herein; and/or (b) a TF screening library according to any of the embodiments described herein; and/or (c) a viral packaging system according to any of the embodiments described herein; and/or (d) a viral particle according to any of the embodiments described herein; and/or (e) an isolated cell according to any of the embodiments described herein; and/or (e) an isolated cell according to any of the embodiments described herein; and/or (e) and particle according to any of the embodiments described herein; and optionally instructions for use. **[0014]** In another aspect, provided herein is a method of performing a high throughput gene activation screen, the method comprising, consisting of, or consisting essentially

of: (a) transducing a target cell with the viral particle according to any of the embodiments described herein; and (b) performing scRNA-seq on the transduced target cell to identify the nucleic acid barcode. In some embodiments, the method further comprises or consists of determining a fitness effect in the transduced target cell. In some embodiments, the method further comprises or consists of identifying a co-perturbation network. In some embodiments, the method further comprises or consists of identifying a functional gene module. In some embodiments, the target cell is a stem cell. In some embodiments, the stem cell is an embryonic stem cell (ESC) or an induced pluripotent stem cell (iPSC). In some embodiments, the target cell is a mammalian cell, optionally wherein the mammalian cell is an equine, bovine, canine, murine, porcine, feline, or human cell. In a particular embodiment, the target cell is a human cell.

[0015] In other aspects, also provided herein is a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell. In some embodiments, ectopic expression of ETV2 is induced by transducing the stem cell with a vector comprising a nucleic acid encoding ETV2 and a nucleic acid encoding an expression control element. In some embodiments, the stem cell is an ESC or an iPSC. In some embodiments, the stem cell is a mammalian cell, optionally wherein the mammalian cell is an equine, bovine, canine, murine, porcine, feline, or human cell. In some embodiments, the stem cell is a human cell. In some embodiments, the stem cell has been genetically modified. In some embodiments, the method further comprises or consists of genetically modifying the stem cell or the endothelial cell.

**[0016]** In further aspect, also provided herein is an endothelial cell produced by a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell, and optionally a carrier. In some embodiments, the endothelial cell expresses at least one of CDH5, PECAM1, or VWF.

**[0017]** In another aspect, also provided herein is a population of endothelial cells produced by a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell, and optionally a carrier.

**[0018]** In some aspects, provided herein is a composition comprising, consisting of, or consisting essentially of an endothelial cell produced by a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell, or a population of endothelial cells produced according to a method described herein, and one or more of: a pharmaceutically acceptable carrier, a cryopreservative or a preservative. In some embodiments, the carrier is a pharmaceutically acceptable carrier. In some embodiments, the cryopreservative is suitable for long term storage

of the composition at a temperature ranging from  $-200^{\circ}$  C. to  $0^{\circ}$  C., from  $-80^{\circ}$  C. to  $0^{\circ}$  C., from  $-20^{\circ}$  C. to  $0^{\circ}$  C., or from  $0^{\circ}$  C. to  $10^{\circ}$  C.

[0019] In some aspects, provided herein is a method of treating a subject in need thereof, the method comprising, consisting of, or consisting essentially of administering an endothelial cell produced by a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell, or a population of endothelial cells produced according to a method described herein, or a composition comprising, consisting of, or consisting essentially of the endothelial cell or population and a carrier to the subject. In some embodiments of the method, an effective amount of the endothelial cell, population, or composition is administered to the subject. In some embodiments, the endothelial cell or population is allogenic or autologous to the subject being treated.

**[0020]** In some embodiments of the method, the subject has a wound, a corneal disease or condition, a myocardial infarction, or a vascular disease or condition. In some embodiments, the subject has a corneal disease or condition. In some embodiments, the administration is local or systemic. In some embodiments, the endothelial cell, population, or composition is administered to the subject's eye.

**[0021]** In some embodiments of the method, the subject is a mammal and the mammal is an equine, bovine, canine, murine, porcine, feline, or human. In some embodiments, the mammal is a human. In some embodiments, the endothelial cells are autologous or allogeneic to the subject being treated.

#### BRIEF DESCRIPTION OF THE FIGURES

[0022] FIGS. 1A-1F: SEUSS workflow and identification of significant TFs from fitness and scRNA-seq analysis. (FIG. 1A) Schematic of experimental and analytical framework for evaluation of effects of transcription factor (TF) overexpression in hPSCs: Individual TFs are cloned into the barcoded ORF overexpression vector, pooled and packaged into lentiviral libraries for transduction of hPSCs. Transduced cells are harvested at a fixed time point to be assayed as single cells using droplet based scRNA-seq to evaluate transcriptomic changes. Cells are genotyped by amplifying the overexpression transcript from scRNA-seq cDNA prior to fragmentation and library construction, and identifying the overexpressed TF barcode for each cell. The cell count for each genotype is used to estimate fitness. Gene expression matrices from scRNA-seq are used to obtain differential gene expression and clustering signatures which in turn are used for evaluation of cell state reprogramming and gene regulatory network analysis. (FIG. 1B) Fitness effect of TFs: log fold change of individual TFs, calculated as cell counts normalized against plasmid library read counts. (FIG. 1C) t-SNE projection (left panel), and cluster enrichment of significant TFs in clusters (right panel) from screens in pluripotent stem cell medium. (FIG. 1D) t-SNE projection (left panel), and cluster enrichment of significant TFs in clusters (right panel) from screens in unilineage (endothelial) growth medium. (FIG. 1E) t-SNE projection (left panel), and enrichment of significant TFs in clusters (right panel) from screens in multilineage differentiation medium. (FIG. 1F) Number of differentially expressed genes for TFs

across different growth media. The TFs in (FIG. 1C), (FIG. 1D), (FIG. 1E) and (FIG. 1F) were chosen as significant with the following criteria: cluster enrichment with a false discovery rate (FDR) of less than  $10^{-6}$  and a cluster enrichment profile different from control (mCherry) with a FDR less than  $10^{-6}$ , or if the TF drove differential expression of more than 100 genes.

[0023] FIGS. 2A-2G: Effect of TF overexpression on gene-to-gene co-perturbation network (FIG. 2A) Schematic for gene-gene co-perturbation network analysis: A SNN network is built from the linear model coefficients and the network is then segmented into gene modules. Genes have a highly weighted edge between them if they respond similarly to TF overexpression. (FIG. 2B) Gene module network: Node size indicates the number of genes in the module; Edge size indicates distance between modules. (FIG. 2C) Effect of TF overexpression on gene modules: (FIG. 2D) Schematic of functional domains of c-MYC: MYC Box I (MBI) and MYC Box II (II) which are essential for transactivation of target genes are housed in the aminoterminal domain (NTD); the basic (b) helix-loop-helix (HLH) leucine zipper (LZ) motif, which is required for heterodimerization with the MAX protein is housed in the carboxy-terminal domain (CTD); the nuclear localization signal domain (NLS) is located in the central region of the protein. (FIG. 2E) Effect of MYC mutant overexpression on gene modules. (FIG. 2F) Schematic of KLF gene family protein structure grouped by common structural and functional features (FIG. 2G) Effect of KLF family overexpression on gene modules. For heatmaps in (FIG. 2C), (FIG. 2E), (FIG. 2F), effect size was calculated as the average of the linear model coefficients for a given TF perturbation across all genes within a module.

[0024] FIGS. 3A-3H: Elucidating effects of KLF4, SNAI2 and ETV2 (FIG. 3A) Effect of KLF4 and SNAI2 on a subnetwork of the pluripotent state module, encompassing key pluripotency regulators. Node size indicates the effect size; blue nodes are downregulated, red nodes are upregulated. (FIG. 3B) PC plot of performing PCA on 200 genes from the Hallmark Epithelial Mesenchymal Transition geneset from MSigDB<sup>42</sup>. PC1 corresponds to an EMT-like signature. (FIG. 3C) Effect of KLF4 and SNAI2 on selected epithelial and mesenchymal markers, including key Cadherin genes. (FIG. 3D) Correlation between fitness estimate from scRNA-seq genotype counts and bulk fitness estimate from gDNA in hPSC medium. (FIG. 3E) Morphology change for cells transduced with either ETV2 or mCherry in EGM. (FIG. 3F) Immunofluorescence micrograph of CDH5 labelled day 6 ETV2- or mCherry-transduced cells. (FIG. 3G) qRT-PCR analysis of signature endothelial genes CDH5, PECAM1, VWF and KDR, at day 6 post-transduction. Data were normalized to GAPDH and expressed relative to control cells in pluripotent stem cell medium. (FIG. 3H) Tube formation assay for day 6 ETV2- or mCherrytransduced cells

**[0025]** FIG. **4**: Schematic of cloning strategy for synthesis of barcoded ORF vectors. The construction involved two steps: (i) insertion of a pool of barcodes into the backbone after digestion with HpaI, (ii) individually substituting mCherry with TFs after digestion with BamHI.

**[0026]** FIGS. **5**A-**5**C: Fitness analysis from genomic DNA and correlation with fitness from scRNA-seq genotyped cell counts (FIG. **5**A) Log fold-change of TF read counts amplified from genomic DNA vs plasmid library control (FIG.

**5**B) Log fold change of TF counts vs plasmid library control for genomic DNA reads vs cell counts fitness for: (FIG. **5**B) Unilineage medium (endothelial growth medium) (FIG. **5**C) Multilineage medium.

**[0027]** FIGS. **6**A-**6**D: Differential gene expression analysis of significant TFs (FIG. **6**A) Heatmap of differentially expressed genes for significant TFs in hPSC medium. (FIG. **6**B) Heatmap of differentially expressed genes for significant TFs in endothelial growth medium. (FIG. **6**C) Heatmap of differentially expressed genes for significant TFs in multilineage medium (FIG. **6**D) Heatmap showing signed log p-values of enrichment for differentially expressed homologous genes in mESCs upon overexpression of TFs<sup>25</sup>. ASCL1, CDX2, KLF4, MYOD1, and OTX2 display a high degree of overlap with overexpression of their homologs in mESCs.

**[0028]** FIGS. 7A-7F: Correlation between aggregated samples. For all plots, correlation was between the coefficients of significant hits, with a hit being defined as a gene—TF pair with the following significance criteria: (FDR<0.05, |coefl>0.025). (FIGS. 7A-7E) Correlation between significant hits in the combined hPSC dataset with hits in each individual dataset. (FIG. 7F) Correlation of hits between the two multilineage datasets.

**[0029]** FIGS. **8**A-**8**C: Correlation between fitness and transcriptomic effects. (FIG. **8**A) Correlation of the number of differentially expressed genes for each TF vs the fitness effect (log-FC) for hPSC medium (FIG. **8**B) Correlation of the number of differentially expressed genes for each TF vs the fitness effect (log-FC) for endothelial medium (FIG. **8**C) Correlation of the number of differentially expressed genes for each TF vs the fitness effect (log-FC) for multilineage medium.

[0030] FIGS. 9A-9D: Confirmatory assays for effects of KLF4 and SNAI2 on key genes in the pluripotency network and involved in EMT (FIG. 9A) qRT-PCR analysis of signature pluripotency network genes SOX2, POU5F1, NANOG, DNMT3B, DPPA4 and SALL2 at day 5 posttransduction in in pluripotent stem cell medium. (FIG. 9B) qRT-PCR analysis of signature cadherins during EMT: CDH1 and CDH2 at day 5 post-transduction in pluripotent stem cell medium. (FIG. 9C) qRT-PCR analysis of signature epithelial marker genes during EMT: EPCAM, LAMC1 and SPP1 at day 5 post-transduction in pluripotent stem cell medium. (FIG. 9D) qRT-PCR analysis of signature mesenchymal marker genes during EMT: TPM2, THY1 and VIM at day 5 post-transduction in pluripotent stem cell medium. Data for all assays were normalized to GAPDH and expressed relative to control cells.

**[0031]** FIGS. **10**A-**10**B: Correlation of KLF4 and MYC effects across samples. (FIG. **10**A) Correlation of KLF4 effects in the KLF family screen with KLF4 effects in the hPSC screen. (FIG. **10**B) Correlation of MYC effects in the MYC mutants screen with KLF4 effects in the hPSC screen.

#### DETAILED DESCRIPTION

**[0032]** Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All technical and patent publications cited herein are incorporated herein by reference in their entirety. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

[0033] The practice of the present invention will employ, unless otherwise indicated, conventional techniques of tissue culture, immunology, molecular biology, microbiology, cell biology and recombinant DNA, which are within the skill of the art. See, e.g., Sambrook and Russell eds. (2001) Molecular Cloning: A Laboratory Manual, 3rd edition; the series Ausubel et al. eds. (2007) Current Protocols in Molecular Biology; the series Methods in Enzymology (Academic Press, Inc., N.Y.); MacPherson et al. (1991) PCR 1: A Practical Approach (IRL Press at Oxford University Press); MacPherson et al. (1995) PCR 2: A Practical Approach; Harlow and Lane eds. (1999) Antibodies, A Laboratory Manual; Freshney (2005) Culture of Animal Cells: A Manual of Basic Technique, 5th edition; Gait ed. (1984) Oligonucleotide Synthesis; U.S. Pat. No. 4,683,195; Hames and Higgins eds. (1984) Nucleic Acid Hybridization; Anderson (1999) Nucleic Acid Hybridization; Hames and Higgins eds. (1984) Transcription and Translation; Immobilized Cells and Enzymes (IRL Press (1986)); Perbal (1984) A Practical Guide to Molecular Cloning; Miller and Calos eds. (1987) Gene Transfer Vectors for Mammalian Cells (Cold Spring Harbor Laboratory); Makrides ed. (2003) Gene Transfer and Expression in Mammalian Cells; Mayer and Walker eds. (1987) Immunochemical Methods in Cell and Molecular Biology (Academic Press, London); Herzenberg et al. eds (1996) Weir's Handbook of Experimental Immunology; Manipulating the Mouse Embryo: A Laboratory Manual, 3rd edition (Cold Spring Harbor Laboratory Press (2002)); Sohail (ed.) (2004) Gene Silencing by RNA Interference: Technology and Application (CRC Press).

**[0034]** All numerical designations, e.g., pH, temperature, time, concentration, and molecular weight, including ranges, are approximations which are varied (+) or (-) by increments of 0.1 or 1.0, where appropriate. It is to be understood, although not always explicitly stated that all numerical designations are preceded by the term "about." It also is to be understood, although not always explicitly stated, that the reagents described herein are merely exemplary and that equivalents of such are known in the art.

#### Definitions

**[0035]** As used in the specification and claims, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

**[0036]** As used herein, the term "comprising" or "comprises" is intended to mean that the compositions and methods include the recited elements, but not excluding others. "Consisting essentially of" when used to define compositions and methods, shall mean excluding other elements of any essential significance to the combination for the stated purpose. Thus, a composition consisting essentially of the elements as defined herein would not exclude trace contaminants from the isolation and purification method and pharmaceutically acceptable carriers, such as phosphate buffered saline, preservatives and the like. "Consisting of" shall mean excluding more than trace elements of other ingredients and substantial method steps for administering the compositions of this disclosure or process steps to

produce a composition or achieve an intended result. Embodiments defined by each of these transition terms are within the scope of this disclosure.

**[0037]** As is known to those of skill in the art, there are 6 classes of viruses. The DNA viruses constitute classes I and II. The RNA viruses and retroviruses make up the remaining classes. Class III viruses have a double-stranded RNA genome. Class IV viruses have a positive single-stranded RNA genome, the genome itself acting as mRNA Class V viruses have a negative single-stranded RNA genome used as a template for mRNA synthesis. Class VI viruses have a positive single-stranded RNA genome but with a DNA intermediate not only in replication but also in mRNA synthesis. Retroviruses carry their genetic information in the form of RNA; however, once the virus infects a cell, the RNA is reverse-transcribed into the DNA form which integrates into the genomic DNA of the infected cell. The integrated DNA form is called a provirus.

**[0038]** A "viral vector" is defined as a recombinantly produced virus or viral particle that comprises a nucleic acid to be delivered into a host cell, either in vivo, ex vivo or in vitro. Examples of viral vectors include retroviral vectors, lentiviral vectors, adenovirus vectors, adeno-associated virus vectors, alphavirus vectors and the like. Alphavirus vectors, such as Semliki Forest virus-based vectors and Sindbis virus-based vectors, have also been developed for use in gene therapy and immunotherapy. See, Schlesinger and Dubensky (1999) Curr. Opin. Biotechnol. 5:434-439 and Ying, et al. (1999) Nat. Med. 5(7):823-827.

[0039] In aspects where gene transfer is mediated by a lentiviral vector, a vector construct refers to the polynucleotide comprising the lentiviral genome or part thereof, and a therapeutic gene. As used herein, "lentiviral mediated gene transfer" or "lentiviral transduction" carries the same meaning and refers to the process by which a gene or nucleic acid sequences are stably transferred into the host cell by virtue of the virus entering the cell and integrating its genome into the host cell genome. The virus can enter the host cell via its normal mechanism of infection or be modified such that it binds to a different host cell surface receptor or ligand to enter the cell. Retroviruses carry their genetic information in the form of RNA; however, once the virus infects a cell, the RNA is reverse-transcribed into the DNA form which integrates into the genomic DNA of the infected cell. The integrated DNA form is called a provirus. As used herein, lentiviral vector refers to a viral particle capable of introducing exogenous nucleic acid into a cell through a viral or viral-like entry mechanism. A "lentiviral vector" is a type of retroviral vector well-known in the art that has certain advantages in transducing nondividing cells as compared to other retroviral vectors. See, Trono D. (2002) Lentiviral vectors, New York: Spring-Verlag Berlin Heidelberg.

**[0040]** Lentiviral vectors of this disclosure include vectors based on or derived from oncoretroviruses (the sub-group of retroviruses containing MLV), and lentiviruses (the sub-group of retroviruses containing HIV). Examples include ASLV, SNV and RSV all of which have been split into packaging and vector components for lentiviral vector particle production systems. The lentiviral vector particle according to this disclosure may be based on a genetically or otherwise (e.g. by specific choice of packaging cell system) altered version of a particular retrovirus.

**[0041]** That the vector particle according to the disclosure is "based on" a particular retrovirus means that the vector is

derived from that particular retrovirus. The genome of the vector particle comprises components from that retrovirus as a backbone. The vector particle contains essential vector components compatible with the RNA genome, including reverse transcription and integration systems. Usually these will include gag and pol proteins derived from the particular retrovirus. Thus, the majority of the structural components of the vector particle will normally be derived from that retrovirus, although they may have been altered genetically or otherwise so as to provide desired useful properties. However, certain structural components and in particular the env proteins, may originate from a different virus. The vector host range and cell types infected or transduced can be altered by using different env genes in the vector particle production system to give the vector particle a different specificity.

**[0042]** The term "an expression control element" as used herein, intends a polynucleotide that is operatively linked to a target polynucleotide to be transcribed, and facilitates the expression of the target polynucleotide. A promoter is an example of an expression control element.

[0043] The term "promoter" refers to a nucleic acid sequence (e.g., a region of genomic DNA) that initiates transcription of a particular gene. The promoter includes the core promoter, which is the minimal portion of the promoter required to properly initiate transcription and can also include regulatory elements such as transcription factor binding sites. The regulatory elements may promote transcription or inhibit transcription. Regulatory elements in the promoter can be binding sites for transcriptional activators or transcriptional repressors. A promoter can be constitutive or inducible. A constitutive promoter refers to one that is always active and/or constantly directs transcription of a gene above a basal level of transcription. An inducible promoter is one which is capable of being induced by a molecule or a factor added to the cell or expressed in the cell. An inducible promoter may still produce a basal level of transcription in the absence of induction, but induction typically leads to significantly more production of the protein. Non-tissue specific promoters include but are not limited to human cytomegalovirus (CMV), CMV enhancer/ chicken (3-actin (CBA) promoter, Rous sarcoma virus (RSV), simian virus 40 (SV40) and mammalian elongation factor  $1\alpha$  (EF1 $\alpha$ ), are non-specific promoters and are commonly used in gene therapy vectors. Promoters can also be tissue specific. A tissue specific promoter allows for the production of a protein in a certain population of cells that have the appropriate transcriptional factors to activate the promoter.

**[0044]** A "target cell" as used herein, shall intend a cell containing the genome into which polynucleotides that are operatively linked to an expression control element are to be integrated. Cells that are infected with a lentivirus or susceptible to lentiviral infection are non-limiting examples of target cells.

**[0045]** "Host cell" refers not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0046] The terms "polynucleotide," "nucleic acid," and "oligonucleotide" are used interchangeably and refer to a

polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides or analogs thereof. Polynucleotides can have any three-dimensional structure and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: a gene or gene fragment (for example, a probe, primer, EST or SAGE tag), exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A polynucleotide can comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure can be imparted before or after assembly of the polynucleotide. The sequence of nucleotides can be interrupted by non-nucleotide components. A polynucleotide can be further modified after polymerization, such as by conjugation with a labeling component. The term also refers to both double- and single-stranded molecules. Unless otherwise specified or required, any embodiment of this this disclosure that is a polynucleotide encompasses both the double-stranded form and each of two complementary single-stranded forms known or predicted to make up the double-stranded form.

**[0047]** A polynucleotide is composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); thymine (T); and uracil (U) for thymine when the polynucleotide is RNA. Thus, the term "polynucleotide sequence" is the alphabetical representation of a polynucleotide molecule. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

[0048] The term "isolated" as used herein refers to molecules or biological or cellular materials being substantially free from other materials, e.g., greater than 70%, or 80%, or 85%, or 90%, or 95%, or 98%. In one aspect, the term "isolated" refers to nucleic acid, such as DNA or RNA, or protein or polypeptide, or cell or cellular organelle, or tissue or organ, separated from other DNAs or RNAs, or proteins or polypeptides, or cells or cellular organelles, or tissues or organs, respectively, that are present in the natural source and which allow the manipulation of the material to achieve results not achievable where present in its native or natural state, e.g., recombinant replication or manipulation by mutation. The term "isolated" also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state. The term "isolated" is also used herein to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides, e.g., with a purity greater than 70%, or 80%, or 85%, or 90%, or 95%, or 98%. The term "isolated" is also used herein to refer to cells or tissues that are isolated from other cells or tissues and is meant to encompass both cultured and engineered cells or tissues.

**[0049]** As used herein, "stem cell" defines a cell with the ability to divide for indefinite periods in culture and give rise to specialized cells. At this time and for convenience, stem

cells are categorized as somatic (adult), embryonic or induced pluripotent stem cells. A somatic stem cell is an undifferentiated cell found in a differentiated tissue that can renew itself (clonal) and (with certain limitations) differentiate to yield all the specialized cell types of the tissue from which it originated. An embryonic stem cell is a primitive (undifferentiated) cell from the embryo that has the potential to become a wide variety of specialized cell types. Pluripotent embryonic stem cells can be distinguished from other types of cells by the use of markers including, but not limited to, Oct-4, alkaline phosphatase, CD30, TDGF-1, GCTM-2, Genesis, Germ cell nuclear factor, SSEA1, SSEA3, and SSEA4.

**[0050]** The term "culturing" refers to the in vitro propagation of cells or organisms on or in synthetic culture conditions such as culture media of various kinds. In some aspects, the medium is changed daily. It is understood that the descendants of a cell grown in culture may not be completely identical (i.e., morphologically, genetically, or phenotypically) to the parent cell. By "expanded" is meant any proliferation, growth, or division of cells. Disclosed herein are culture methods that support differentiation by in inclusion of nutrients and effector molecules necessary to promote or support the differentiation of stem cells into differentiated cells.

[0051] "Differentiation" describes the process whereby an unspecialized cell acquires the features of a specialized cell such as a heart, liver, pancreas, or muscle cell. "Directed differentiation" refers to the manipulation of stem cell culture conditions to induce differentiation into a particular cell type. "Dedifferentiated" defines a cell that reverts to a less committed position within the lineage of a cell. As used herein, the term "differentiates or differentiated" defines a cell that takes on a more committed ("differentiated") position within the lineage of a cell and may also include maturation or development of the cell. As used herein, "a cell that differentiates into pancreatic beta cell" defines any cell that can become a committed pancreatic cells that produces insulin. Non-limiting examples of cells that are capable of differentiating into endothelial cells include embryonic stem cells, pluripotent stem cells, induced pluripotent stem cells (iPSCs), mesenchymal stem cell, hematopoietic stem cells, and adipose stem cells.

**[0052]** As used herein, a "pluripotent cell" defines a less differentiated cell that can give rise to at least two distinct (genotypically and/or phenotypically) further differentiated progeny cells. In another aspect, a "pluripotent cell" includes an Induced Pluripotent Stem Cell (iPSC) which is an artificially derived stem cell from a non-pluripotent cell, typically an adult somatic cell, produced by inducing expression of one or more stem cell specific genes.

**[0053]** A "composition" is intended to encompass a combination of active agent and another "carrier," e.g., compound or composition, inert (for example, a detectable agent or label) or active, such as an adjuvant, diluent, binder, stabilizer, buffers, salts, lipophilic solvents, preservative, adjuvant or the like. Compositions may include stabilizers and preservatives. As used herein, the term "pharmaceutically acceptable carrier" encompasses any of the standard pharmaceutical carriers, such as a phosphate buffered saline solution, water, and emulsions, such as an oil/water or water/oil emulsion, and various types of wetting agents. For examples of carriers, stabilizers and adjuvants, see Martin (1975) Remington's Pharm. Sci., 15th Ed. (Mack Publ. Co., Easton). Carriers also include biocompatible scaffolds, pharmaceutical excipients and additives proteins, peptides, amino acids, lipids, and carbohydrates (e.g., sugars, including monosaccharides, di-, tri-, tetra-, and oligosaccharides; derivatized sugars such as alditols, aldonic acids, esterified sugars and the like; and polysaccharides or sugar polymers), which can be present singly or in combination, comprising alone or in combination 1-99.99% by weight or volume. Exemplary protein excipients include serum albumin such as human serum albumin (HSA), recombinant human albumin (rHA), gelatin, casein, and the like. Representative amino acid/antibody components, which can also function in a buffering capacity, include alanine, glycine, arginine, betaine, histidine, glutamic acid, aspartic acid, cysteine, lysine, leucine, isoleucine, valine, methionine, phenylalanine, aspartame, and the like. Carbohydrate excipients are also intended within the scope of this this disclosure, examples of which include but are not limited to monosaccharides such as fructose, maltose, galactose, glucose, D-mannose, sorbose, and the like; disaccharides, such as lactose, sucrose, trehalose, cellobiose, and the like; polysaccharides, such as raffinose, melezitose, maltodextrins, dextrans, starches, and the like; and alditols, such as mannitol, xylitol, maltitol, lactitol, xylitol sorbitol (glucitol) and myoinositol.

**[0054]** A population of cells intends a collection of more than one cell that is identical (clonal) or non-identical in phenotype and/or genotype.

**[0055]** "Substantially homogeneous" describes a population of cells in which more than about 50%, or alternatively more than about 60%, or alternatively more than 70%, or alternatively more than 75%, or alternatively more than 80%, or alternatively more than 85%, or alternatively more than 90%, or alternatively, more than 95%, of the cells are of the same or similar phenotype. Phenotype can be determined by assaying for expression of a pre-selected cell surface marker or other marker.

[0056] An "effective amount" is an amount sufficient to effect beneficial or desired results. In the context of a therapeutic cell, population, or composition, the term "effective amount" as used herein refers to the amount to alleviate at least one or more symptom of a disease, disorder, or condition (e.g., corneal condition), and relates to a sufficient amount of the cell, population, or composition to provide the desired effect (e.g., repair of the cornea). An effective amount as used herein would also include an amount sufficient to delay the development of a disease, disorder, or condition symptom, alter the course of disease, disorder, or condition symptom (for example but not limited to, slow the progression of corneal degradation), or reverse a symptom of a disease, disorder, or condition. Thus, it is not possible to specify the exact "effective amount." However, for any given case, an appropriate "effective amount" can be determined by one of ordinary skill in the art using only routine experimentation.

**[0057]** An effective amount can be administered in one or more administrations, applications or dosages. Such delivery is dependent on a number of variables including the time period for which the individual dosage unit is to be used, the bioavailability of the therapeutic agent, the route of administration, etc. It is understood, however, that specific dose levels of the therapeutic agents of the present disclosure for any particular subject depends upon a variety of factors including the activity of the specific compound employed,

the age, body weight, general health, sex, and diet of the subject, the time of administration, the rate of excretion, the drug combination, and the severity of the particular disorder being treated and form of administration. Treatment dosages generally may be titrated to optimize safety and efficacy. The dosage can be determined by a physician and adjusted, as necessary, to suit observed effects of the treatment. Typically, dosage-effect relationships from in vitro and/or in vivo tests initially can provide useful guidance on the proper doses for patient administration. In general, one will desire to administer an amount of the compound that is effective to achieve a serum level commensurate with the concentrations found to be effective in vitro. Determination of these parameters is well within the skill of the art. These considerations, as well as effective formulations and administration procedures are well known in the art and are described in standard textbooks. Consistent with this definition, as used herein, the term "therapeutically effective amount" is an amount sufficient to inhibit RNA virus replication ex vivo, in vitro or in vivo. Consistent with this definition, as used herein, the term "therapeutically effective amount" is an amount sufficient to achieve the result of the method.

**[0058]** The term "administration" shall include without limitation, administration by oral, parenteral (e.g., intramuscular, intraperitoneal, intravenous, ICV, intracisternal injection or infusion, subcutaneous injection, or implant), by inhalation spray nasal, vaginal, rectal, sublingual, urethral (e.g., urethral suppository) or topical routes of administration (e.g., gel, ointment, cream, aerosol, etc.) and can be formulated, alone or together, in suitable dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvants, excipients, and vehicles appropriate for each route of administration. The invention is not limited by the route of administration, the formulation or dosing schedule.

[0059] An "enriched population" of cells intends a substantially homogenous population of cells having certain defined characteristics. The cells are greater than 60%, or alternatively greater than 65%, or alternatively greater than 70%, or alternatively greater than 75%, or alternatively greater than 80%, or alternatively greater than 85%, or alternatively greater than 90%, or alternatively greater than 95%, or alternatively greater than 98% identical in the defined characteristics. In one aspect, the substantially homogenous population of cells express markers that correlate with pluripotent cell identity such as expression of stem-cell specific genes like OCT4 and NANOG. In another aspect, the substantially homogenous population of cells express markers that are correlated with definitive endoderm cell identity such SOX17, CXCR4, FOXA2, and GATA4. In another aspect, the substantially homogenous population of cells express markers that are correlated with posterior foregut cell identity such as HNF1β, HNF4A while suppressing expression of HHEX, HOXA3, CDX2, OCT4, and NANOG. In another aspect, the substantially homogenous population of cells express markers that are correlated with pancreatic progenitor cell identity such as PDX1 (pancreatic duodenal homeobox gene 1). In another aspect, the substantially homogenous population of cells express markers that are correlated with endocrine pancreas cell identity such as NKX6.1, NEURO-D1, and NGN3. In yet another aspect, the substantially homogenous population of cells express markers that are correlated with islet precursor cell identity such as INS. This population may further be identified by its ability to secrete C-peptide.

**[0060]** A "gene" refers to a polynucleotide containing at least one open reading frame that is capable of encoding a particular RNA, polypeptide, or protein after being transcribed and/or translated. The term "express" refers to the production of a gene product. As used herein, "expression" refers to the process by which polynucleotides are transcribed into RNA and/or the process by which the transcribed RNA such as mRNA is subsequently being translated into peptides, polypeptides, or proteins. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. A "gene product" or alternatively a "gene expression product" refers to the amino acid (e.g., peptide or polypeptide) or functional RNA (e.g. a tRNA, miRNA, rRNA, or shRNA) generated when a gene is transcribed and translated.

[0061] The term "treating" (or "treatment") of a pancreatic or immune disorder or condition refers to ameliorating the effects of, or delaying, halting or reversing the progress of, or delaying or preventing the onset of, a pancreatic or immune condition such as diabetes, pre-diabetes, juvenile onset (Type I) diabetes mellitus, including pediatric insulindependent diabetes mellitus (IDDM), and adult onset diabetes mellitus (Type II diabetes). Treatment includes preventing the disease or condition (i.e., causing the clinical symptoms of the disease not to develop in a patient that may be predisposed to the disease but does not yet experience or display symptoms of the disease), inhibiting the disease or condition (i.e., arresting or reducing the development of the disease or its clinical symptoms), or relieving the disease or condition (i.e., causing regression of the disease or its clinical symptoms).

**[0062]** A mammalian stem cell, as used herein, intends a stem cell having an origin from a mammal. Non-limiting examples include, e.g., a murine, a canine, an equine, a simian and a human. An animal stem cell intends a stem cell having an origin from an animal, e.g., a mammalian stem cell.

**[0063]** A "subject," "individual" or "patient" is used interchangeably herein, and refers to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, rats, rabbit, simians, bovines, ovine, porcine, canines, feline, farm animals, sport animals, pets, equine, and primate, particularly human. Besides being useful for human treatment, the methods and compositions disclosed herein are also useful for veterinary treatment of companion mammals, exotic animals and domesticated animals, including mammals, rodents, and the like which is susceptible to diabetes or other immune or pancreatic diseases or conditions. In one embodiment, the mammals include horses, dogs, and cats. In another embodiment of the present disclosure, the human is an adolescent or infant under the age of eighteen years.

**[0064]** An immature stem cell, as compared to a mature stem cell, intends a phenotype wherein the cell expresses or fails to express one or more markers of a mature phenotype. Examples of such are known in the art, e.g., telomerase length or the expression of actin for mature cardiomyocytes derived or differentiated from a less mature phenotype such as an embryonic stem cell. An immature beta cell intends a pancreatic cell that has insulin secretory granules but lacks

GSIS. In contrast, mature beta cells typically are positive for GSIS and have low lactate dehydrogenase (LDH).

#### Descriptive Embodiments

[0065] Understanding the complex effects of genetic perturbations on cellular state and fitness in human pluripotent stem cells (hPSCs) has been challenging using traditional pooled screening techniques which typically rely on unidimensional phenotypic readouts. Here, Applicants use barcoded open reading frame (ORF) overexpression libraries with a coupled single-cell RNA sequencing (scRNA-seq) and fitness screening approach, a technique Applicants call SEUSS (ScalablE fUnctional Screening by Sequencing), to establish a comprehensive assaying platform. Using this system, Applicants perturbed hPSCs with a library of developmentally critical transcription factors (TFs), and assayed the impact of TF overexpression on fitness and transcriptomic cell state across multiple media conditions. Applicants further leveraged the versatility of the ORF library approach to systematically assay mutant gene libraries and also whole gene families. From the transcriptomic responses, Applicants built genetic co-perturbation networks to identify key altered gene modules. Strikingly, Applicants found that KLF4 and SNAI2 have opposing effects on the pluripotency gene module, highlighting the power of Applicants' method to characterize the effects of genetic perturbations. From the fitness responses, Applicants identified ETV2 as a driver of reprogramming towards an endothelial-like state.

Isolated Nucleic Acids and Transcription Factor Screening Libraries

**[0066]** This disclosure provides isolated polynucleotides or nucleic acids comprising, consisting of, or consisting essentially of (a) a polynucleotide or nucleic acid encoding a transcription factor (TF) open reading frame (ORF); (b) a nucleic acid barcode, and (c) an optional vector comprising (a) and (b); wherein the nucleic acid barcode is located 3' to the TF ORF.

**[0067]** Transcription factors are proteins that bind (directly or indirectly through recruitment factors) to enhancer or promoter regions of DNA (e.g. a genome) and interact to activate, repress, or maintain the current level of transcription of a particular gene or genetic locus. Many transcription factors can bind to specific DNA sequences. Non-limiting examples of TFs can be found at TFCat (Genome Biol. 2009; 10(3): R29).

**[0068]** An ORF refers to the part of a gene or polynucleotide that has the potential to be transcribed and/or translated. ORFs span intron/exon regions, which in some embodiments can be spliced together after transcription of the ORF to yield a final mRNA for protein translation. Thus, ORFs include both introns and exons, when applicable. In some embodiments, an ORF is a continuous stretch of codons that contain a start codon and a stop codon. In some embodiments, the transcription termination site is located after the ORF, beyond the translation stop codon.

**[0069]** In some embodiments, the TF ORF encodes a developmentally critical TF. As used herein, "developmentally critical" refers to a transcription factor that regulates development and/or differentiation by modulating transcription. Regulation may include, for example, suppression of one or more specific developmental or differentiation gene expression programs, activation of one or more specific

developmental or differentiation gene expression programs, and/or maintenance of a specific level of activation or suppression of a specific developmental or differentiation program. For example, a developmentally critical transcription factor may function upstream of a lineage-specific gene network and direct a stem or progenitor cell to differentiate into that specific cell lineage. Examples of developmentally critical TFs include but are not limited to ASCL1, ASCL3, ASCL4, ASCL5, ATF7, CDX2, CRX, ERG, ESRRG, ETV2, FLI1, FOXA1, FOXA2, FOXA3, FOXP1, GATA1, GATA2, GATA4, GATA6, GLI1, HAND2, HNF1A, HNF1B, HNF4A, HOXA1, HOXA10, HOXA11, HOXB6, KLF4, LHX3, LMX1A, MEF2C, MESP1, MITF, MYC, MYCL, MYCN, MYOD1, MYOG, NEUROD1, NEU-ROG1, NEUROG3, NRL, ONECUT1, OTX2, PAX7, POU1F1, POU5F1, RUNX, SIX1, SIX2, SNAI2, SOX10, SOX2, SOX3, SPI1, SPIB, SPIC, SRY, TBX5, and TFAP2C. [0070] In some embodiments, the vector is a retroviral vector, optionally a lentiviral vector.

**[0071]** This disclosure provides a vector comprising, or alternatively consisting essentially of, or yet further consisting of a viral backbone. In one aspect, the viral backbone contains essential nucleic acids or sequences for integration into a target cell's genome. In one aspect, the essential nucleic acids necessary for integration of the genome of the target cell include at the 5' and 3' ends the minimal LTR regions required for integration of the vector.

[0072] In one aspect, the term "vector" intends a recombinant vector that retains the ability to infect and transduce non-dividing and/or slowly-dividing cells and integrate into the target cell's genome. In several aspects, the vector is derived from or based on a wild-type virus. In further aspects, the vector is derived from or based on a wild-type lentivirus. Examples of such, include without limitation, equine infectious anaemia virus (EIAV), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), and human immunodeficiency virus (HIV). Alternatively, it is contemplated that other retrovirus can be used as a basis for a vector backbone such murine leukemia virus (MLV). It will be evident that a viral vector need not be confined to the components of a particular virus. The viral vector may comprise components derived from two or more different viruses, and may also comprise synthetic components. Vector components can be manipulated to obtain desired characteristics, such as target cell specificity.

**[0073]** The recombinant vectors of this disclosure are derived from primates and non-primates. Examples of primate lentiviruses include the human immunodeficiency virus (HIV), the causative agent of human acquired immunodeficiency syndrome (AIDS), and the simian immunodeficiency virus (SIV). The non-primate lentiviral group includes the prototype "slow virus" visna/maedi virus (VMV), as well as the related caprine arthritis-encephalitis virus (CAEV), equine infectious anaemia virus (EIAV) and the more recently described feline immunodeficiency virus (FIV) and bovine immunodeficiency virus (BIV). Prior art recombinant lentiviral vectors are known in the art, e.g., see U.S. Pat. Nos. 6,924,123; 7,056,699; 7,07,993; 7,419,829 and 7,442,551, incorporated herein by reference.

**[0074]** U.S. Pat. No. 6,924,123 discloses that certain retroviral sequence facilitate integration into the target cell genome. This patent teaches that each retroviral genome comprises genes called gag, pol and env which code for virion proteins and enzymes. These genes are flanked at both

ends by regions called long terminal repeats (LTRs). The LTRs are responsible for proviral integration, and transcription. They also serve as enhancer-promoter sequences. In other words, the LTRs can control the expression of the viral genes. Encapsidation of the retroviral RNAs occurs by virtue of a psi sequence located at the 5' end of the viral genome. The LTRs themselves are identical sequences that can be divided into three elements, which are called U3, R and U5. U3 is derived from the sequence unique to the 3' end of the RNA. R is derived from a sequence repeated at both ends of the RNA, and U5 is derived from the sequence unique to the 5'end of the RNA. The sizes of the three elements can vary considerably among different retroviruses. For the viral genome and the site of poly (A) addition (termination) is at the boundary between R and U5 in the right hand side LTR. U3 contains most of the transcriptional control elements of the provirus, which include the promoter and multiple enhancer sequences responsive to cellular and in some cases, viral transcriptional activator proteins.

**[0075]** With regard to the structural genes gag, pol and env themselves, gag encodes the internal structural protein of the virus. Gag protein is proteolytically processed into the mature proteins MA (matrix), CA (capsid) and NC (nucleocapsid). The pol gene encodes the reverse transcriptase (RT), which contains DNA polymerase, associated RNase H and integrase (IN), which mediate replication of the genome.

**[0076]** In another aspect, provided herein is a TF screening library comprising, consisting of, or consisting essentially of at least one isolated nucleic acid comprising, consisting of, or consisting essentially of (a) a nucleic acid encoding a transcription factor (TF) open reading frame (ORF); (b) a nucleic acid barcode, and (c) an optional vector comprising (a) and (b); wherein the nucleic acid barcode is located 3' to the TF ORF. In some embodiments, the TF ORF encodes a developmentally critical TF, optionally selected from the TFs listed in Table 1.

**[0077]** In some embodiments, the TF screening library comprises, consists of, or consists essentially of at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 nucleic acids or vectors, wherein each nucleic acid or vector comprises, consists of, or consists essentially of a distinct nucleic acid encoding a TF ORF.

**[0078]** In some embodiments, the TF screening library further comprises, consists of, or consists essentially of a nucleic acid encoding a selectable marker (e.g., hygromycin). In some embodiments, the TF screening library further comprises, consists of, or consists essentially of a nucleic acid encoding an expression control element. In some embodiments, the expression control element is a promoter or a long terminal repeat (LTR). In some embodiments, the TF screening library further comprises, consists of, or consists essentially of a nucleic acid encoding a nucleic acid encoding a translation elongation factor, optionally wherein the translation elongation factor is Ef1a.

**[0079]** For the production of viral vector particles, the vector RNA genome is expressed from a DNA construct encoding it, in a host cell. The components of the particles not encoded by the vector genome are provided in trans by additional nucleic acid sequences (the "packaging system", which usually includes either or both of the gag/pol and env genes) expressed in the host cell. The set of sequences required for the production of the viral vector particles may be introduced into the host cell by transient transfection, or

they may be integrated into the host cell genome, or they may be provided in a mixture of ways. The techniques involved are known to those skilled in the art.

**[0080]** In another aspect, provided herein is a viral packaging system comprising, consisting of, or consisting essentially of at least one isolated nucleic acid comprising, consisting of, or consisting essentially of (a) a nucleic acid encoding a transcription factor (TF) open reading frame (ORF); (b) a nucleic acid barcode, and (c) an optional vector comprising (a) and (b); wherein the nucleic acid barcode is located 3' to the TF ORF; or aTF screening library; and a packaging plasmid.

[0081] In another aspect, provided herein is a method for producing a viral particle, the method comprising, consisting of, or consisting essentially of transfecting a packaging cell line with a viral packaging system comprising, consisting of, or consisting essentially of at least one isolated nucleic acid comprising, consisting of, or consisting essentially of (a) a nucleic acid encoding a transcription factor (TF) open reading frame (ORF); (b) a nucleic acid barcode, and (c) an optional vector comprising (a) and (b); wherein the nucleic acid barcode is located 3' to the TF ORF; or aTF screening library; and a packaging plasmid under conditions suitable to package the vector or the TF screening library into a viral particle. In another aspect, also provided herein is a viral particle produced by this method, and optionally a carrier. In another aspect, also provided herein is an isolated cell comprising a nucleic acid, vector, or particle as described herein, and optionally a carrier.

**[0082]** Retroviral vectors for use in the methods and compositions described herein include, but are not limited to Invitrogen's pLenti series versions 4, 6, and 6.2 "ViraPower" system. Manufactured by Lentigen Corp.; pHIV-7-GFP, lab generated and used by the City of Hope Research Institute; "Lenti-X" lentiviral vector, pLVX, manufactured by Clontech; pLKO.1-puro, manufactured by Sigma-Aldrich; pLemi®, manufactured by Open Biosystems; and pLV, lab generated and used by Charité Medical School, Institute of Virology (CBF), Berlin, Germany.

**[0083]** This invention also provides the suitable packaging cell line. In one aspect, the packaging cell line is the HEK-293 cell line. Other suitable cell lines are known in the art, for example, described in the patent literature within U.S. Pat. Nos. 7,070,994; 6,995,919; 6,475,786; 6,372,502; 6,365,150 and 5,591,624, each incorporated herein by reference.

**[0084]** Yet further provided is an isolated cell or population of cells, comprising, or alternatively consisting essentially of, or yet further consisting of, a retroviral particle of this invention, which in one aspect, is a viral particle. In one aspect, the isolated host cell is a packaging cell line.

#### Kits

**[0085]** In another aspect, provided herein is a kit comprising, consisting of, or consisting essentially of at least one of (a) a nucleic acid or vector according to any of the embodiments described herein; and/or (b) a TF screening library according to any of the embodiments described herein; and/or (c) a viral packaging system according to any of the embodiments described herein; and/or (d) a viral particle according to any of the embodiments described herein; and/or (e) an isolated cell according to any of the embodiments described herein; and/or (e) an isolated cell according to any of the embodiments described herein; and/or (b) a transfer according to any of the embodiments described herein; and/or (e) an isolated cell according to any of the embodiments described herein; and optionally instructions for use.

#### High Throughput Gene Activation Screens

[0086] In another aspect, provided herein is a method of performing a high throughput gene activation screen, the method comprising, consisting of, or consisting essentially of: (a) transducing a target cell with the viral particle according to any of the embodiments described herein; and (b) performing single cell RNA sequencing (scRNA-seq) on the transduced target cell to identify the nucleic acid barcode.

[0087] In some embodiments, scRNA-seq methods comprise the following steps: isolation of single cell and RNA, reverse transcription (RT), optional amplification, library generation, and sequencing. Several scRNA-seq protocols appropriate for use with the disclosed methods have been published: Tang et al. (Nat Methods. 6 (5): 377-82) STRT (Islam, S. et al. (2011). Genome Res. 21 (7): 1160-7), SMART-seq (Ramskold, D. et al. (2012). Nat. Biotechnol. 30 (8): 777-82) CEL-seq (Hashimshony, T. et al. (2012) Cell Rep. 2 (3): 666-73), and Quartz-seq (Sasagawa, Y. et al. (2013) Genome Biol. 14 (4): R31).

[0088] In some embodiments, the method further comprises or consists of determining a fitness effect in the transduced target cell. Fitness effects include but are not limited to effects on cell proliferation, effects on cell viability, effects on rate of senescence, effects on apoptosis, effects on DNA repair mechanisms, effects on genome stability, effects on gene transcription, and effects on stress response. In some embodiments, fitness effects are calculated from genomic DNA or mRNA reads,

[0089] In some embodiments, the method further comprises or consists of identifying a co-perturbation network. In some embodiments, the method further comprises or consists of identifying a functional gene module. In some embodiments, the target cell is a stem cell. In some embodiments, the stem cell is an embryonic stem cell (ESC) or an induced pluripotent stem cell (iPSC). In some embodiments, the target cell is a mammalian cell, optionally wherein the mammalian cell is an equine, bovine, canine, murine, porcine, feline, or human cell. In a particular embodiment, the target cell is a human cell.

Endothelial Differentiation Methods and Compositions

[0090] Also provided herein is a method driving or directing differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 (Ets variant 2, Entrez gene: 2116) in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell.

[0091] In some embodiments, ectopic expression of ETV2 is induced by transducing the stem cell with a vector (e.g., AAV) comprising a nucleic acid encoding ETV2 and a nucleic acid encoding an expression control element. In other embodiments, the vector encodes an open reading frame of ETV2. In other embodiments, the vector encodes a cDNA of ETV2 (RefSeq: NM 001300974; NM 001304549; NM 014209). A non-limiting example of the sequence of an ETV2 cDNA is provided:

					(S	EO ID NO: 1
1	tteetgttge	agataagccc	agettageee	agctgacccc	agaccetete	ccctcactcc
61	ccccatgtcg	caggatcgag	accctgaggc	agacagcccg	ttcaccaagc	cccccgcccc
121	gcccccatca	ccccgtaaac	ttctcccagc	ctccgccctg	ccctcaccca	gcccgctgtt
181	ccccaagcct	cgctccaagc	ccacgccacc	cctgcagcag	ggcagcccca	gaggccagca
241	cctatccccg	aggctggggt	cgaggctcgg	ccccgcccct	gcctctgcaa	cttgagcctg
301	gctgcgaccc	ctgctctgac	gtctcggaaa	attccccctt	gcccaggccc	ttgggggagg
361	gggtgcatgg	tatgaaatgg	ggctgagacc	cccggctggg	ggcagaggaa	cccgccagag
421	aaggagccaa	attaggcttc	tgtttccctg	atctggcact	ccaagggggac	acgccgacag
481	cgacagcaga	gacatgctgg	aaaggtacaa	gctcatccct	ggcaagcttc	ccacagctgg
541	actgggggctc	cgcgttactg	cacccagaag	ttccatgggg	ggcggagccc	gactctcagg
601	ctcttccgtg	gtccgggggac	tggacagaca	tggcgtgcac	agcctgggac	tcttggagcg
661	gegeetegea	gaccctgggc	cccgcccctc	tcggcccggg	ccccatcccc	gccgccggct
721	ccgaaggcgc	cgcgggccag	aactgcgtcc	ccgtggcggg	agaggccacc	tcgtggtcgc
781	gcgcccaggc	cgccgggagc	aacaccagct	gggactgttc	tgtggggccc	gacggcgata
841	cctactgggg	cagtggcctg	ggcgggggagc	cgcgcacgga	ctgtaccatt	tcgtggggcg
901	ggcccgcggg	cccggactgt	accacctcct	ggaacccggg	gctgcatgcg	ggtggcacca
961	cctctttgaa	gcggtaccag	agctcagctc	tcaccgtttg	ctccgaaccg	agcccgcagt
1021	cggaccgtgc	cagtttggct	cgatgcccca	aaactaacca	ccgaggtccc	attcagctgt
1081	ggcagttcct	cctggagctg	ctccacgacg	gggcgcgtag	cagctgcatc	cgttggactg
1141	gcaacagccg	cgagttccag	ctgtgcgacc	ccaaagaggt	ggctcggctg	tggggcgagc

#### -continued

- 1201 gcaagagaaa gccgggcatg aattacgaga agctgagccg gggccttcgc tactactatc
- 1261 geogegacat cgtgegeaag agegggggge gaaagtacae gtacegette gggggeegeg
- 1321 tgcccagcct agcctatccg gactgtgcgg gaggcggacg gggagcagag acacaataaa

**[0092]** In some embodiments, the stem cell is an ESC or an iPSC. In some embodiments, the stem cell is a mammalian cell, optionally wherein the mammalian cell is an equine, bovine, canine, murine, porcine, feline, or human cell. In some embodiments, the stem cell is a human cell. In some embodiments, the stem cell has been genetically modified. In some embodiments, the method further comprises or consists of genetically modifying the stem cell or the endothelial cell.

**[0093]** In further aspect, also provided herein is an endothelial cell produced by a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell, and optionally a carrier. In some embodiments, the endothelial cell expresses at least one of CDH5 (VE-Cadherin, Entrez gene: 1003; RefSeq: NM 001114117, NM 00179, PECAM1 (Platelet endothelial cell adhesion molecule, Entrez gene: 5175; RefSeq: NM 000442), or VWF (Von Willebrand Factor, Entrez gene: 7450, RefSeq: NM 000552).

**[0094]** In another aspect, also provided herein is a population of endothelial cells produced by a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell, and optionally a carrier.

[0095] In some aspects, provided herein is a composition comprising, consisting of, or consisting essentially of an endothelial cell produced by a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell, or a population of endothelial cells produced according to a method described herein, and one or more of: a pharmaceutically acceptable carrier, a cryopreservative or a preservative. In some embodiments, the carrier is a pharmaceutically acceptable carrier. In some embodiments, the cryopreservative is suitable for long term storage of the composition at a temperature ranging from -200° C. to 0° C., from -80° C. to 0° C., from -20° C. to 0° C., or from  $0^{\circ}$  C. to  $10^{\circ}$  C.

#### Methods of Treatment

**[0096]** In some aspects, provided herein is a method of treating a subject in need thereof, the method comprising, consisting of, or consisting essentially of administering an endothelial cell produced by a method driving differentiation of a stem cell into an endothelial cell, the method comprising, consisting of, or consisting essentially of inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an

endothelial cell, or a population of endothelial cells produced according to a method described herein, or a composition comprising, consisting of, or consisting essentially of the endothelial cell or population and a carrier to the subject. In some embodiments of the method, an effective amount of the endothelial cell, population, or composition is administered to the subject. In some embodiments, the endothelial cell or population is allogenic or autologous to the subject being treated. In one aspect, the treatment excludes prevention.

**[0097]** In some embodiments of the method, the subject has a wound, a corneal disease or condition, a myocardial infarction, or a vascular disease or condition. In some embodiments, the subject has a corneal disease or condition. In some embodiments, the administration is local or systemic. In some embodiments, the endothelial cell, population, or composition is administered to the subject's eye.

[0098] An effective amount can be administered in one or more administrations, applications or dosages. Such delivery is dependent on a number of variables including the time period for which the individual dosage unit is to be used, the bioavailability of the therapeutic agent, the route of administration, etc. It is understood, however, that specific dose levels of the therapeutic agents of the present disclosure for any particular subject depends upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, sex, and diet of the subject, the time of administration, the rate of excretion, the drug combination, and the severity of the particular disorder being treated and form of administration. Treatment dosages generally may be titrated to optimize safety and efficacy. The dosage can be determined by a physician and adjusted, as necessary, to suit observed effects of the treatment. Typically, dosage-effect relationships from in vitro and/or in vivo tests initially can provide useful guidance on the proper doses for patient administration. In general, one will desire to administer an amount of the compound that is effective to achieve a serum level commensurate with the concentrations found to be effective in vitro. Determination of these parameters is well within the skill of the art. These considerations, as well as effective formulations and administration procedures are well known in the art and are described in standard textbooks. Consistent with this definition, as used herein, the term "therapeutically effective amount" is an amount sufficient to achieve the result of the method.

**[0099]** The term "administration" shall include without limitation, administration by oral, parenteral (e.g., intramuscular, intraperitoneal, intravenous, ICV, intracisternal injection or infusion, subcutaneous injection, or implant), by inhalation spray nasal, vaginal, rectal, sublingual, urethral (e.g., urethral suppository) or topical routes of administration (e.g., gel, ointment, cream, aerosol, etc.) and can be formulated, alone or together, in suitable dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvants, excipients, and vehicles

<sup>1381</sup> aatteeeggt caaaceteaa aaaaaaaaaa aaa

appropriate for each route of administration. The invention is not limited by the route of administration, the formulation or dosing schedule.

**[0100]** In some embodiments of the method, the subject is a mammal and the mammal is an equine, bovine, canine, murine, porcine, feline, or human. In some embodiments, the mammal is a human. In some embodiments, the endothelial cells are autologous or allogeneic to the subject being treated.

**[0101]** Having been generally described herein, the follow examples are provided to further illustrate this invention.

#### Example 1

**[0102]** Recently, screens combining genetic perturbations with scRNA-seq readouts have emerged as promising alternatives to traditional screens, enabling high-throughput, high-content screening by profiling the transcriptomes of tens of thousands of individual cells simultaneously. Unlike array-based methods scRNA-seq screens are scalable, while unlike traditional pooled screening techniques, they enable direct readout of cell state changes. In addition, they also enable the evaluation of heterogeneous cellular response to perturbations. While several groups have demonstrated CRISPR-Cas9 based knock-out and knock-down scRNA-seq screens, to Applicants' knowledge, gene activation screens have yet to be demonstrated.

[0103] Here, Applicants use barcoded ORF overexpression libraries with a coupled scRNA-seq and fitness screen, a technique Applicants call SEUSS, to systematically overexpress TFs and assay both, the transcriptomic and fitness effects on hPSCs. Applicants chose open-reading frame (ORF) constructs for several reasons, namely that ORF constructs yield strong, stable expression of the gene of interest, enable the ability to express a targeted isoform of the gene, and allow for the ability to express engineered or mutant forms of the gene, aspects otherwise not accessible through endogenous gene activation. Applicants screened a pooled library of TFs that are either developmentally critical, specific to key lineages, or are pioneer factors capable of binding closed chromatin (Table 1). From the transcriptomic readouts, Applicants built a gene-gene co-perturbation network, segmented the network genes into functional gene modules, and used these gene modules to also elucidate the impact of TF overexpression on the pluripotent cell state. Notably, Applicants also leveraged the versatility of the ORF library approach and SEUSS to systematically assay mutant gene libraries (MYC) and whole gene families (KLF). Finally, Applicants also leveraged the complementary fitness information via SEUSS to ascertain that ETV2 is a novel reprogramming factor for hPSCs, whose overexpression yields rapid differentiation towards the endothelial lineage. [0104] Applicants designed Applicants' ORF overexpression vector such that each TF was paired with a unique 20 bp barcode sequence located downstream of the 3' end of a hygromycin resistance transgene (FIG. 1A, FIG. 4), and 200 bp upstream of the lentiviral 3'-long terminal repeat (LTR) region. This yields a polyadenylated transcript bearing the barcode proximal to the 3' end, thereby facilitating efficient capture and detection in scRNA-seq. To construct the ORF library, transcription factors were amplified out of a multitissue human cDNA pool or directly synthesized as doublestranded DNA fragments, and individually cloned into the backbone vector (FIG. 4). The final library consisted of 61 developmentally critical or pioneer TFs (Table 1). Applicants chose this library size to ensure that within a single scRNA-seq run of up to 10,000 cells, each perturbation was represented by at least 50-100 cells. However, SEUSS can be scaled up to include all known TFs.

**[0105]** Applicants conducted the overexpression screens by transducing lentiviral ORF libraries into human embryonic stem cells (hESCs), maintaining them under antibiotic selection for 5 days after transduction, for screens in hPSC medium, and 6 days after transduction, for screens in unlineage (endothelial) and multilineage (high serum) medium, and then performing scRNA-seq on the transduced and selected cells. TF barcodes were recovered and associated with scRNA-seq cell barcodes by targeted amplification from the unfragmented cDNA, allowing genotyping of each cell for downstream analysis (FIG. 1A). Genotyped cell counts, although an under-sampling of the bulk population, also allowed Applicants to obtain an estimate of fitness, which was strongly correlated with bulk fitness obtained from genomic DNA (FIG. 1A, FIG. 3D, FIGS. 5A-5C).

**[0106]** To analyze the effect of the TF perturbations, Applicants used the Seurat computational pipeline to cluster the cells from the scRNA-seq expression matrix (FIG. 1C, FIG. 1D, FIG. 1E). In parallel, a linear model was used to identify genes whose expression levels are appreciably changed by the perturbation. To select TFs for downstream analysis, Applicants calculated over-enrichment of TFs in clusters using Fisher's exact test (FIG. 1C, FIG. 1D, FIG. 1E). Subsequently, Applicants focused Applicants' analysis on TFs that were either significantly enriched for at least one cluster (FDR<10<sup>-6</sup>), or had at least 100 significant differentially expressed genes. For TFs that had significant over-enrichment in a cluster, Applicants repeated the linear regression analysis, only including cells that fell into enriched clusters (FIG. 1F).

[0107] This framework was used to conduct screens in hPSC medium, aggregating 12,873 cells across five samples. Applicants found that these independent experiments were well correlated with the combined dataset (Pearson R>0.84), implying overall reproducibility and the absence of strong batch effects (FIGS. 7A-7E). To study the interplay of ORF overexpression with growth media conditions, Applicants also conducted screens in a unilineage medium, specifically endothelial growth medium, on 5,646 cells and in a multilineage (ML) differentiation medium, specifically a high serum growth medium, on 3476 cells (Table 3). Two samples were aggregated for analysis in the ML medium, again showing good correlation (FIG. 7F; Pearson R=0.68). [0108] From Applicants' screen in hPSC medium, Applicants found that transcriptomic changes do not necessarily correlate with changes in fitness (FIG. 5), thus Applicants' coupled screening method enables a more comprehensive profiling of impacts on both fitness and cell state. Among the most significantly depleted TFs, was the haemato-endothelial master regulator ETV2, (FIG. 3D, FIG. 5), which guided Applicants' choice of EGM for a unilineage medium screen. [0109] Applicants find that certain TFs show consistent effects across all media conditions (CDX2, KLF4), while some TFs have medium-specific effects. For instance, SNAI2 effects were specific to hPSC medium, MITF to ML medium, and GATA4 to EGM (FIG. 1F). To benchmark Applicants' results, Applicants compared expression profiles for significant TFs in hPSC medium with a previously

reported bulk RNA-seq screen of TF perturbations in

mESCs. For TFs present in both datasets, Applicants found

a strong overlap, suggesting the effectiveness of Applicants' screen for studying perturbations (FIG. 6D).

[0110] To interpret the effects of the significant TFs, Applicants used the regression coefficients of the linear model to build a weighted gene-to-gene co-perturbation network, where genes with a highly weighted edge between them respond to TF perturbations in a similar manner (FIG. 2A). Using this network, Applicants identified 11 altered gene modules via a modularity optimization graph clustering algorithm. Many of these gene modules showed a strong enrichment for Gene Ontology (GO) terms, and gene module identity was assigned using GO enrichment paired with manual inspection of genes in each module. In this network, Applicants found that the pluripotency gene module and the chromatin accessibility module are highly interconnected, reflecting the relationship between those two biological processes (FIG. 2B), and suggesting that this network may serve as a resource to understand the cascading effects of genetic perturbations (FIG. 2B, Table 5).

**[0111]** Applicants next calculated the effect of each significant TF on the gene modules (FIG. **2**C). Applicants found that the annotated neural specifiers NEUROD1, NEUROG1, and NEUROG3, which show similar cluster enrichment and differential expression patterns, upregulate the neuron differentiation module, consistent with their known effects. ASCL1 and MYOD1, which also show similarity in clustering and expression patterns, upregulate the Notch pathway module (FIG. **2**C). This similarity between ASCL1 and

MYOD1 may be due to a myogenic program initiated by ASCL1. Notably, for the TFs with consistent effects across medium conditions, Applicants find that both CDX2 and KLF4 strongly downregulate the pluripotency gene module, while CDX2 also upregulates the embryonic development gene module, potentially reflecting its role in trophectoderm development, and KLF4 tends to upregulate the cytoskeleton and motility gene modules.

[0112] Next, since in Applicants' screens MYC was found to drive significant transcriptomic changes in hPSC medium in its wild type form (FIG. 1F), Applicants chose to focus on it in demonstrating the ability of Applicants' platform to also systematically screen mutant forms of proteins. Specifically, Applicants constructed a library of mutant MYC proteins, where functional domains were systematically deleted (FIG. 2D), or mutations at known hotspots were incorporated (Glu-39, Thr-58 and Ser-62). Screening this library in pluripotent stem cell medium, Applicants found that while some variants, such as known hotspot mutations, as well as deletion of the nuclear localization signal (NLS) sequence maintain an effect similar to the wild type MYC, a majority of the other mutant forms show a greater overlap with the control mCherry-transduced cells, suggesting the essential requirement of the mapped domains for function of MYC in hPSCs (FIG. 2E).

MYC Mutants Library:

#### [0113]

GENE	SEQUENCE	SEQ ID NO:	MUTATION
MYC	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC	2	Deletion of MYC
AMBI	TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA		Box I
	GGAGGAGAACTTCTACCAGCAGCAGCAGCAGAGCGAGCT		
	GCAGCCCCCGGCGGGATCAGGTAGCGGTAGCCGCCGCTC		
	CGGGCTCTGCTCGCCCTCCTACGTTGCGGTCACACCCTTCT		
	CCCTTCGGGGAGACAACGACGGCGGTGGCGGGAGCTTCT		
	CCACGGCCGACCAGCTGGAGATGGTGACCGAGCTGCTGG		
	GAGGAGACATGGTGAACCAGAGTTTCATCTGCGACCCGG		
	ACGACGAGACCTTCATCAAAAACATCATCATCCAGGACTG		
	TATGTGGAGCGGCTTCTCGGCCGCCGCCAAGCTCGTCTCA		
	GAGAAGCTGGCCTCCTACCAGGCTGCGCGCAAAGACAGC		
	GGCAGCCCGAACCCCGCCGCGGCCACAGCGTCTGCTCCA		
	CCTCCAGCTTGTACCTGCAGGATCTGAGCGCCGCCGCCTC		
	AGAGTGCATCGACCCCTCGGTGGTCTTCCCCTACCCTCC		
	AACGACAGCAGCTCGCCCAAGTCCTGCGCCTCGCAAGACT		
	CCAGCGCCTTCTCCCGTCCTCGGATTCTCTGCTCTCCTCG		
	ACGGAGTCCTCCCCGCAGGGCAGCCCCGAGCCCCTGGTGC		
	TCCATGAGGAGACACCGCCCACCACCAGCAGCGACTCTG		
	AGGAGGAACAAGAAGATGAGGAAGAAATCGATGTTGTTT		
	CTGTGGAAAAGAGGCAGGCTCCTGGCAAAAGGTCAGAGT		
	CTGGATCACCTTCTGCTGGAGGCCACAGCAAACCTCCTCA		
	CAGCCCACTGGTCCTCAAGAGGTGCCACGTCTCCACACAT		
	CAGCACAACTACGCAGCGCCTCCCTCCACTCGGAAGGACT		
	ATCCTGCTGCCAAGAGGGTCAAGTTGGACAGTGTCAGAGT		
	CCTGAGACAGATCAGCAACAACCGAAAATGCACCAGCCC		
	CAGGTCCTCGGACACCGAGGAGAATGTCAAGAGGCGAAC		
	ACACAACGTCTTGGAGCGCCAGAGGAGGAACGAGCTAAA		
	ACGGAGCTTTTTTGCCCTGCGTGACCAGATCCCGGAGTTG		
	GAAAACAATGAAAAGGCCCCCCAAGGTAGTTATCCTTAAA		
	AAAGCCACAGCATACATCCTGTCCGTCCAAGCAGAGGAG		
	CAAAAGCTCATTTCTGAAGAGGACTTGTTGCGGAAACGAC		
	GAGAACAGTTGAAACACAAACTTGAACAGCTACGGAACT		
	CTTGTGCG		
c-MYC	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC	3	Deletion of MYC
AMBII	TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA		Box II
	GGAGGAGAACTTCTACCAGCAGCAGCAGCAGAGCGAGCT		
	GCAGCCCCCGGCGCCCAGCGAGGATATCTGGAAGAAATT		

14

-continued

GENE	SEQUENCE	SEQ ID NO:	MUTATION
	TCCGGGCTCTGCTCGCCCTCCTACGTTGCGGTCACACCCTT CTCCTCGGGGAGACAACGACGCGGGGGGGGGG		
MYC ANLS	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCAGCAGCAGCAGCAGCAGCAGCA GCAGCCCCGGCCCCAGCCAG	4	Deletion of nuclear localization signal sequence
мүс Дь	ATGCCCCTCAACGTTAGCTTCACCACAGGAACTATGACC TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCAGCAGCAGCAGCAGAGCGAGCT GCAGCCCCCGGCGCCCAGCGAGGATATCTGGAAGAAATT CGAGCTGCTGCCCCCCCCCC	5	Deletion of basic motif

-continued

15

GENE	SEQUENCE	SEQ NO:	ID	MUTATION
	GTATGTGGAGCGGCTTCTCGGCCGCCGCCAAGCTCGTCTC AGAGAAGCTGGCCTCCTACCAGGCTGCGCGCAAAGACAG CGGCAGCCGGAACCCGCCGCCGGGCCACAGGTCTGCTCC ACCTCCAGCTTGTACCTGCAGGATCTGAGCGCCGCCGCT CAGAGTCGACGAGCCCTCGGGTGTCTCCCCTACCCTTC AACGACAGCAGCTCGCCCCAGGTCTCCCGGCAGAGACT CCAGGGCCTTCTCCCGCCCGGGGCAGCCCCGGGGCCTCCTG ACGGAGTCCTCCCCGCAGGCACCCCGAGCCCCTGGTGC TCCATGAGGAGAACAGAGCCCCCAGCACCAGCAACGCCCTGG AGGAGAACAAGAAGAAGCACGCCACAGCAAAGGTCAGAG CTGGATCACCTTCTGCGCGCGCCGCAGCAAAGGTCAGAG CTGGATCACCTTCTGCGCGCGCCCGAGCACCCCGAGCACCCCC CAGCCCACTGGTCCTCAAGAGGCCCCCAGCAACCTCCTC CAGCCCACTGGTCCTCAAGAGGTCCCGCCACACACGCAACCCCCCC CAGCCCACTGGTCCTCAAGAGGTCCCGCCCCCCCCACCACCACACACA			
ИYC NHLH	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA GAGGAGAAACTTCTACCACAGCAGCAGCGAGCT CCGACCCCCGGCCCCAGCGAGCATATCTGGAAGAAATT CGAGCTGCTGCCCACCCGCCCCTGCCCCTAGCCGCGCCC TCCGGGCTCTGCTCGCCCCCTCCTACGTTGCGGTCACACCCTT CTCCCTCGGGGAGACAACGACGGCGGTGGCGGGAGCTT CTCCCTCGGGGAGCACAGCGGGGGGGGGG	6		Deletion of helix- loop-helix motif
4YC \LZ	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCAGCAGCAGCAGCAGAGCGAGCT GCAGCCCCCGGCCCCAGCGAGATATCTGGAAGAAATT CGAGCTGCTGCCCACCCGCCCCTGTCCCCTAGCGCGCCGC TCCGGGGCTCTGCTCGCCCTCCTACGTTGCGGTCACACCCTT CTCCCTCGGGGAGACAACGACGACGGTGGACCGAGGCGGGACCTT CTCCACGGCCGACCCGCTGGAGATGGTGGACCGAGGCGGCGCG GGAGGACAATGGTGAACCAGAGTTTCATCTGCGACCCG GACGACGAGACCTTCACAAAACATCATCATCCACGACCT GTATGTGGAGCGGCCTCCTACCAGGCCGCCCAAGCTCGTCC AGGACAACGCGGCCCCCGCCGCGCCAAGGCTCTCCC ACCTCCAGGCCGACCCCGCCGCGCCACAGCGCCTCCC ACCTCCAGGCCGCCCCCCCCGCGGCCACAGCGCCTCCC CAGAGCGCCGAACCCCGCCGCGGCCACAGCGCCTCCC CAGACGCCGACCCCCCCCCGCGGCCCCCCCCCC	7		Deletion of leucine zipper motif

CCAGCGCCTTCTCCCGTCCTCGGATTCTCTGCTCTCCTCG

16

-continued

		SEQ ID	MITTER BIT ON
GENE	SEQUENCE	NO:	MUTATION
	ACGGAGTCCTCCCCGCAGGGCAGCCCCGAGCCCCTGGTGC TCCATGAGAAGACACCGCCACCACCACCAGCAGCGACTCTG AGGAGGACAAGAAGAAGAGGAGAGAAAACGACGACGATGTT CTGTGGAAAAGAGGCAGGTCCTGGCAAAAGGTCAGAGT CTGGATCACCTTCTGCTGGAGGGCCCAGGCAAAACCTCCTCA CAGCCACTGGTCCTCAAGAGGTGCCACGTCTCCACACAT CAGCACAACTACGCAGGGCCACGTCCCCCCCCGGAAGGACT ATCCTGGCTGCCAAGAGGGTCCAGGTGGACAGTGCAGAGT CCTGAGACAGATCAGCAACACGAAAATGCACCAGCC CAGGTCCTCGGACACCGAGGAGAATGTCAAGGGCGAAC ACACAACGTCTTGGAGCGCCCGAGGAGGAAGACGAGCTAAA ACGGAGCTTTTTTGCCCTGCGTGACCAGATCCCGAGATTG GAAACAATGAAAAGGCCCCCCAAGGTAGTTATCCTTAAA AAAGCCACAGCATACATCCTGTCCGTCCAAGCAGAGGAGA		
MYC ANTD	ATGGGATCAGGTAGCGGTCTCGTCTCAGAGAAGCTGGCCT CCTACCAGGCTGCGCGCAAAGACACGGGCAGCCCGAACC CCGCCQGGGCCACAGCGTCTGCTCCACCTCCAGCTTGTA CCTGCAGGATCTGAGCGCCGCCGCCTCCAGAGTGCATCGAC CCCTCGGTGGTCTCCCCTACCCTCTCAACGGCAGCGCTCCC CGCCAAGTCCTGGGCCTCGCAGAGACTCCAGCGCGCTCTCC TCCGTCCTCGGATTCTCTGCTCTCCTCGACGGAGTCCTCCC CGCAGGCCACCCCAGCAGCGCCTCGGTGCTCCATGAGGAGA CACCGCCCACCACCAGCAGCGACTCTGAGGAGAAGA AAGATGAGGAAGAAATCGATGTTGTGGAAAAGA GGCAGGCTCCTGGCAAAAGGTCAGAGTCTGGGCACACTTC TGCTGGAGGCCCAGGCAACTCCTCACAGCCCACTGGTC CTCAAGAGGTGCCACGTCTCCCACACACCACGCCACTGGTC CTCAAGAGGTGCCACGTCTCCCACACTCAGGACAACTACG CACCGCCCCCCCCCC	8	Deletion of amino- terminal domain: Housing MYC Box I and II
MYC ACTD	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC TCGACTACGACTCGTGCAGCCGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCAGCAGCAGCAGCAGCAGCAGCAGCAG GGAGGAGAAACTTCTACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT GCAGCCCCCGGCCCCCGCCCCTGTCCCCTAGCGCGCCG TCGGGGCTGCCCCCCCCTCCCCT	9	Deletion of carboxy- terminal domain: Housing basic helix- loop-helix leucine zipper motif, governing heterodimerization with MAX protein
MYC Glu39Ala	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCAGCAGCAGCAGCAGAGCGCGCT GCAGCCCCCGGCGCCCAGCGAGGATATCTGGAAGAAATT CGAGCTGCCGCCCCCCCCCC	10	Point mutation changing Glutamic Acid to Alanine at amino acid 39

17

-continued

		SEO	TD	
GENE	SEQUENCE	NO:	<u></u>	MUTATION
	CTCCCTTCGGGGAGACAACGACGGCGGTGGCGGGAGCTT CTCCACGGCCGACCAGCTGGAGATGGTGACCGAGCTGCTG GGAGGAGAACCTTCATCAAAAACATCATCATCGGCACCG GACGACGAGACCTTCATCAAAAACATCATCATCAGGACT GTATGTGGAGCGGCTTCTCGCCGCGCGCAAAGCCGC CGGCAGCCCGAACCCCGCGCGCG			
MYC Thr58Ala	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA GCAGGAGAGAACTTCTACCAGCAGCAGCAGCAGAGCGAGC	11		Point mutation changing Threonine to Alanine at amino acid 58
MYC Ser62Ala	ATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACC TCGACTACGACTCGGTGCAGCCGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCAGCAGCAGCAGCGAGCT GCAGCCCCCGGCCCCAGCCAGCAGCAGACAATT CGAGCTGCTGCCCACCCCGCCCCTGGCCCCTAGCCGCCGC TCCGGGCTCTGCTCCCCCCCCCC	12		Point mutation changing Serine to Alanine at amino acid 58

-continued

GENE	SEQUENCE	SEQ NO:	ID	MUTATION	
	GACGACGAGACCTTCATCAAAAACATCATCATCCAGGACT				
	GTATGTGGAGCGGCTTCTCGGCCGCCGCCAAGCTCGTCTC				
	AGAGAAGCTGGCCTCCTACCAGGCTGCGCGCAAAGACAG				
	CGGCAGCCCGAACCCCGCCGCGGCCACAGCGTCTGCTCC				
	ACCTCCAGCTTGTACCTGCAGGATCTGAGCGCCGCCGCCT				
	CAGAGTGCATCGACCCCTCGGTGGTCTTCCCCTACCCTCTC				
	AACGACAGCAGCTCGCCCAAGTCCTGCGCCTCGCAAGACT				
	CCAGCGCCTTCTCCCGTCCTCGGATTCTCTGCTCTCCTCG				
	ACGGAGTCCTCCCCGCAGGGCAGCCCCGAGCCCCTGGTGC				
	TCCATGAGGAGACACCGCCCACCACCAGCAGCGACTCTG				
	AGGAGGAACAAGAAGATGAGGAAGAAATCGATGTTGTTT				
	CTGTGGAAAAGAGGCAGGCTCCTGGCAAAAGGTCAGAGT				
	CTGGATCACCTTCTGCTGGAGGCCACAGCAAACCTCCTCA				
	CAGCCCACTGGTCCTCAAGAGGTGCCACGTCTCCACACAT				
	CAGCACAACTACGCAGCGCCTCCCTCCACTCGGAAGGACT				
	ATCCTGCTGCCAAGAGGGTCAAGTTGGACAGTGTCAGAGT				
	CCTGAGACAGATCAGCAACAACCGAAAATGCACCAGCCC				
	CAGGTCCTCGGACACCGAGGAGAATGTCAAGAGGCGAAC				
	ACACAACGTCTTGGAGCGCCAGAGGAGGAACGAGCTAAA				
	ACGGAGCTTTTTTGCCCTGCGTGACCAGATCCCGGAGTTG				
	GAAAACAATGAAAAGGCCCCCAAGGTAGTTATCCTTAAA				
	AAAGCCACAGCATACATCCTGTCCGTCCAAGCAGAGGAG				
	CAAAAGCTCATTTCTGAAGAGGACTTGTTGCGGAAACGAC				
	GAGAACAGTTGAAACACAAACTTGAACAGCTACGGAACT				
	CTTGTGCG				

**[0114]** Additionally, the consistent and strong effects of KLF4 overexpression motivated the investigation of the full KLF zinc finger transcription factor family (FIG. **2**F) as a demonstration of the utility of Applicants' technique in studying patterns of perturbation effects across gene families. A screen including all 17 members of the KLF family was conducted in pluripotent stem cell medium. Gene module analysis showed that KLF5 and KLF17 also have similar effects as KLF4 (FIG. **2**G), which may reflect their similar

role in promoting or maintaining epithelial cell states. On the other hand, unlike most of the KLF family, KLF13 and KLF16 fail to activate the cytoskeleton and motility module (FIG. **2**G).

KLF Family Library

#### [0115]

GENE	SEQUENCE	SEQ ID NO:
KLF1	ATGGCGACTGCGGAGACAGCACTTCCATCAATCTCAACACTCACT	13
	GGGCCATTTCCAGATACCCAGGACGATTTCCTTAAGTGGTGGCGGTCCGAA	
	GAGGCTCAAGACATGGGACCTGGTCCGCCGGATCCCACCGAACCTCCTCTG	
	CATGTCAAAAGTGAAGATCAGCCTGGCGAGGAAGAGGATGACGAAAGGG	
	GTGCCGACGCCACTTGGGACTTGGATCTTCTCCTTACCAATTTCTCTGGTCC	
	GGAACCTGGCGGGGCACCACAGACGTGCGCTCTCGCTCCCTCAGAAGCGA	
	GCGGGGCTCAGTACCCACCCCCTCCCGAAACTCTGGGAGCCTATGCTGGGG	
	GTCCTGGACTGGTGGCTGGGTTGCTTGGTAGTGAGGACCATTCTGGCTGG	
	TACGCCCCGCTTTGAGGGCCCGCGCTCCGGACGCCTTTGTGGGACCGGCGC	
	TCGCTCCTGCACCGGCTCCGGAACCAAAAGCCCTCGCGCTGCAGCCCGTGT	
	ACCCCGGACCCGGAGCCGGATCCTCAGGGGGGATACTTCCCACGGACCGGA	
	CTCAGCGTTCCAGCGGCTTCCCGGGGCGCCATACGGATTGTTGAGCGGCTAC	
	CCGGCTATGTATCCCGCTCCCCAGTACCAAGGACACTTCCAATTGTTCCGG	
	GGTCTTCAAGGGCCTGCGCCCGGGCCTGCTACCAGTCCCAGTTTCCTCAGT	
	TGTCTGGGACCGGGAACTGTTGGCACTGGACTTGGCGGGACTGCAGAGGA	
	CCCAGGCGTTATAGCAGAGACAGCGCCAAGTAAAAGGGGCCGACGAAGCT	
	GGGCCAGGAAACGCCAAGCTGCGCACACTTGTGCCCATCCAGGTTGCGGT	
	AAATCCTACACGAAGAGCAGTCATCTTAAAGCACATCTTCGCACACACA	
	GGGCGAGAAGCCCTACGCCTGTACTTGGGAAGGTTGCGGCTGGAGATTCG	
	CTAGATCTGACGAGCTCACCCGGCATTATCGAAAACACACTGGCCAGCGA	
	CCGTTCCGGTGCCAACTCTGCCCAAGGGCGTTCAGTCGCTCAGATCATCTG	
	GCTTTGCATATGAAGCGACACCTT	
KLF2	ATGGCCCTTAGTGAACCCATTCTTCCCAGCTTTTCCACGTTCGCGTCTCCTT	14
	GCCGAGAGAGAGGCCTTCAGGAAAGGTGGCCGAGGGCTGAACCCGAGTCT	
	GGAGGTACGGATGATGATCTTAACAGTGTGCTCGATTTCATACTCTCAATG	
	GGACTGGACGGGCTGGGAGCGGAGGCAGCTCCTGAACCACCACCACCCC	
	TCCGCCCCCAGCGTTTTACTACCCGGAGCCAGGTGCGCCGCCGCCATATTC	
	AGCCCCGGCGGGTGGCTTGGTGTCCGAGCTCCTCCGGCCTGAATTGGATGC	

#### -continued

GENE	SEQUENCE	SEQ NO :	ID
	CCCGCTCGGCCCGGCGCTGCATGGTAGATTTCTGCTCGCGCCTCCGGGTCG ACTCGTTAAGGCTGAACCTCCTGAGGCTGATGGTGGAGGTGGCTACGGAT GTGCCCCCGGCTTACCCGAGGAGCCGAGAGGGCCCGGTGGGAGGCCACCGG CCTGGCCCGGCTGCAAGCTGTATGCGGGGGCCCGGTGGGAGGCCTCCCCC GCCCCCAGATACACCCCCCCTTAGTCCAGATGGACGACCAGCTCGACTTCCGC ACCTGGCCCCAGGCGAGTTTCACTACGCCCCCCCTGGAGCTCCTGG CGCCCCAGGTCCTGGACTTCACTACGCCCCCCCTGGAGCCCCTGGGCCTTTGG CGCCCAGGTCCTGGACTTCACTACGCCCCCCCTGGAGCTCCTGA AGCCAAGCCGAAGCGCGCGCAGCCCCTGGAGCCCCTGGACCTCTGA AGCCAAGCCGAAGCGCAGGACGCAGATCATGGCCGCGCCAAGGCGACAGCT ACGCATACCTGCTCATTGCGCGCCACCGGCAAGCGGACAGCT ACGCATACCTGCTCATTGCGCGCCACCGGCAAAGCCCACAAAGAGTTC ACACCTTAAAGCGCACCTCGGCCACCAGCGGAGAAACCATATCATT GTAACTGGGACGGATGGAAGGACTCGGCCACCGGCCAAGGCCACATATCCGA GACATTATCGAAAGCATACCGGACATCGGCCTTCGACACAGAGAGGCTACCG ACAGCATTTTCCCCGGTCTGACCACCTCGCCTTGCACATGAAGAGGCACA TG		
KLF3	ATGCTCATGTTTGACCCAGTTCCTGTCAAGCAAGAGGGCCATGGACCCTGTC TCAGTGTCATACCCATCTAATTACATGGAATCCATGAAGCCTAACAAGTAT GGGGTCATCTACTCCACACCATTGCCTGAGAAGTTCTTTCAGACCCCAGAA GGTCTGTCGCACGGAATACAGATGGAGCCAGTGGACCTCACGGTGAACAA GCGGAGTTCACCCCCTTCGGGTGGGGACTCGCCTCCTCTCAGAGTTCCC GTCCTCACACCGGAGAGCCCCGCTGGGTTGAGCATGCCTTCTTCCAGCCC ACCGATAAAAAAATACTCACCCCCTTCTCCAGGCGTGCAGCCCTTCGGCGT GCCGCTGTCCATGCCACCAGTGATGGCAGCTGCCCTCTCGCGGCATGGCAG CCCCCTTTATGTCACACAGTGATGGCAGCTGCCCTCTCGGGGCAGCCCG TCCCCCTTTATGTACACACAGTCACCCGCCGTGCGGGTGGCAGCCCG GCGCTGTCCATGCCACCACGAGTAGCCAGCCGCGTGGTGGCAGCCCG TCCCCCTTTATGTACACACAGTCACCCCGCCGCGGGGGGGG	15	
KLF5	ATGGCTACAAGGGTGCTGAGCATGAGCGCCGCCCGCCTGGGACCCGTGCCCA GCCGCCGGCGCCGCAGGACGAGCCGGCGTGTTCGCGCAGCTCAAGCCGGTGC TGGGCGCCGCGAGCCCGCCGCGCGCGCGCCGCGGCCGCGGCG	16	
KLF6	ATGGACGTGCTCCCCATGTGCAGCATCTTCCAGGAGCTCCAGATCGTGCAC GAGACCGGCTACTTCTCGGCGCTGCCGTCTCTGGAGGAGTACTGGCAACAG	17	

GAGACCGGCTACTTCTCGGCGCTGCCGTCTCTGGAGGAGTACTGCAACAG ACCTGCCTAGAGCTGGAACGTTACCTCCAGAGCGAGCCCTGCTATGTTTCA GCCTCAGAAATCAAATTTGACAGCCAGGAAGATCTGTGGACCAAAATCAT TCTGGCTCGGGAGAAAAAGGAGGAATCCGAACTGAAGATATCTTCCAGTC

## Apr. 15, 2021

#### -continued

GENE	SEQUENCE	SEQ ID NO:
	CTCCAGAGGACACTCTCATCAGCCGAGCTTTTGTTACAACTTAGAGACCA ACAGCCTGAACTCAGATGTCAGCAGCGAATCCTCTGACAGCTCCGAGGAA CTTTCTCCCACGGCCAAGTTTACCTCCGACCCCATTGGCGAAGTTTTGGTCA GCTCGGGAAAATTGAGCTCCTCTGTCACCTCCACGCCTCCATCTCTCTC	
KLF7	ATGGACGTGTTGGCTAGTTATAGTATATTCCAGGAGCTACAACTTGTCCAC GACACCGGCTACTTCTCAGGCTTTACCATCCCTGGAGGAGACCTGGCAGCAG ACATGCCTTGGAATGGAA	18
KLF8	ATGGT CGATATGGATAAACTCATAAACAACTTGGAGGTCCAACTTAATTCA GAAGGTGGCTCAATGCAGGTATTCAAGCAGGTCACTGCTTCTGTTCGGAAC AGAGATCCCCCTGAGATAGAATACAGAAGTAATATGACTTCTCCAACACTC CTGGATGCCAACCCCATGGAGAACCCAGCACTGTTTAATGACATCAAGATT GAGCCCCCAGAAGAACTTTTGGCTAGTGATTTCAGCCTGCCCAAGTGGAA CCAGTTGACCTCTCTTTCACAAGCCCCAAGGCTCCTCTCCAGCCTGCTAGC ATGCTACAAGCTCCAATACGTCCCCCCAAGCCACAGTCTTCTCCCCAGACC CTTGTGGTGTCCACGTCAACATCTGACATGAGCACTTCAGCAAACATTCCT ACTGTTCTGACCCCAGGCTCTGTCCTGACCTCCTCCAGAGCACTGGTAGC CAGCAGATCTTACATGTCATCCCACATCCCCCTCAGCAACCATTCCT ACTGTTCTGACCCCAGGCTCTGTCCTGACCTCCTCAGAGCACTGGTAGC CAGCAGATCTTACATGTCATTCACCACTATCCCCTCAGTGCAGCCATGGCGCAAAT AAGATGGGTGGCCTGAAGACCATCCCAGTGGTAGTCAGCCCATG GTGTATACTACTTTGCCTGCAGAGTGGGGGCCCTGCAGCCATTACAGTCCCA CTCCATTGCAGGAGATGGTAAAAATGCTGGATCAGGGAGAGAGTCACCCCAC CTCCATGTCTCCACTGGAAATTCCAAGTGACAGTGAAGGTGACCCCAC CAGCAATGGCCCAAAGGCAGGAGAGAGTGCCTTGAAGAGAACA CGGATTCACCAATGGCAGGAGAGAGGCGCTTGCAGGAGAGAGA	19
KLF9	ATGTCCGCGGCCGCCTACATGGACTTCGTGGCTGCCCAGTGTCTGGTTTCC ATTTCGAACCGCGCTGCGGTGCCGGAGCATGGGGTCCCGGACGCCGA GCGGCTGCGACTACCTGAGCGCGAGGTGACCAACGAGACGCGGACGCCGG GGGACACCTGGAAGATTACTGCACACTGGTCACCATCGCCAAGAGCTTG TTGGACCTGGAACAAGTACCGACCATCCAGACCACCGCCGACGAGCGAC AGTCTGGAAAGTCCAGATGAGGATATGGGATCCGACAGCGACGACCAC CGAATCTGGGTCGAGTCCTTCCCACACCCGACGAGGAGAAGACAGAATCCTG GCAGCGCCCCAGCCCGCTCTCCCACACCCCGAGGAGAAGACAGAATCCTG GCAGCGCCCCGCCC	20

#### -continued

GENE	SEQUENCE	SEQ NO :	ID
KLF10	ATGCTCAACTTCGGTGCCTCTCTCCAGCAGACTGCGGAGGAAAGAATGGA AATGATTTCTGAAAGGCCAAAAGAGAGTATGTATTCCTGGAACAAAACTG CAGAGAAAAGTGATTTTGAAGCTGTAGAAGAGCTTATGTCAATGAGCTGC AGTTGGAAGTCTGATTTTAAGAAATACGTTGAAAACAGACCTGTTACACCA GTATCTGATTTGTCAGAGGAAGAAAACGTTCACAGACCTGTTACACCA GTATCTGATTTGTCAGAGGAAGAAATCTGCTTCCGGGAACACCTGATTTT CATACAATCCCAGCATTTTGTTTGACTCCACGTCCACGGCCATCTCTGGACTTTG AACCCTCTCAAGTGTCAAATCTGCTGACGACCAGCGCCATCTACTGACACT TCAAGTCACTCTCAGATACTGCCAAACCTCACATTGCCGCACCTTTCAAAG AGGAAGAAAAGAGCCCAGTATCTGCCCCCAAACTCCCCAAAGCTCACGTGCC GCCCAATGAAACCAGCCAGCATCCTCAACTATCGCAGCACCTTTCAAAG AGAAGAACCCACCTAAATGTGGAGCTGCAAGAAAGAACATACCATGTGC CGCTGTCTCCCAAATGTGGAGGCTGCAAGAAAGAACATACCATGTGC CGCTGTGTCACCAAACAGTCCAAATGTGGAGAGAAACACAGTGGCAGATG TTGATGAGAAAGCAAGTGCTGCACTTTATGACTTTTCTGTGCCTTCCTCAG AGACGTCATCTGCAGGTGCCCCTTCCTGCCAACAACCACTGGCAGAATG CAGTGTTCGCCAAATGTCCCCACCTGCCACCAACAGCGCCACCAAACGAAGT CAGTGTTCCCCACCTCCCCCACCAGCCGCCGCTGTGTCCCCCCACCAAAGAAGAA CAGTCGTCCCCCAGCACCACGCCCCTGTGTCCCCCCACCAACAGAAGT CAGTGTTCCCCAGCACCCCCCCCGCCCCCGGCGCTGTCGTGGCCACCAAGC CCCAGCCCTTGGCCACCAAGTCCCCAAAGGCCCTCTGTTGTGGCACA AGACGTCCTCCCAGCACCAAGTCCCCAAAGGCGCTGTCATGTTTGTGGCAC CCCAGCCCTTGCCAGAGTTCAAAGCCTTCCGGGGGGTGTCCCCCTCGAGCACAA AAGTCACTCCCCAGAGTGACACAAGCCTTCCAGGCGGTGTGCCCCACCAACAAGCC ACCCAGGATGGCCACAAGACCATCTTTAAAAGTTCCCATCTGTAGGCC ACCCAGGACGCACCACGAGCATACTTTAAAAGTTCCCATCTGTAGGCC ACCCAGGACGCACCACGAGCATACTTTAAAAGTTCCCATCTGTAGGCC ACCCAGGACGCACCACGAGCATACTTTAAAAGTTCCCATCTGTAGGCCCA ACGAGGACGCACCACAGGCATACTTTAAAAGTTCCCATCTGTAGGCCCA ACGAGGACGCACCACAGGAAAAGCCTTTCCAGCCAACAGCGGCGATTG TGAAAGGAAGTTTGCCCGTTCTGATGAACGACCACACGGCGGCCCCC ACCCAGGACGCACACAGGACATACTTTAAAAGTTCCCATCTGTAGACCGAACCC ACCCAGGACGCACCACGAGACAAACCCTTCCAGCCACACAGGCGAACCC ACCCAGGACGCACACAGGAAAAGCCTTTCCAGCCACACAGGCGAACCC ACCCAGGACGCACACAGGACAAACGCTTCCAGCCACACAGGCGAACCC ACCCAGGGTGAAAAAAATTGCCGTCCTGATGAACGACCACACGGCGGTTCATG AGGAGTGACCATTTGACCAAGCATGCCCGGCCCACCACACAGGCGAACCC ACCCAGGCTGAAAAAGAATTTGCCCGCCCCGC	21	
KLF11	ATGCATACTCCTGATTTCGCTGGACCTGACGACGCCCGAGCCGTGGACATT ATGGACATTTGTGAATCTATACTCGAAAGAAAGAGACATGATTCAGAGCG AAGTACATGCTCTATCCTCGAGCAACAGACATGGAGGCGATGATACAACTAT GAATCCGACGCTTACTCCAGTTCCGATAGCGGCGACGTACAACTACTG TTCATATGGACGCTACCCGGTTCCCGATAGCGGCGACGTACACACTACTG TTCATATGGACGCAGCCACGCCTGAGCTGCCCAAAGACTTCACAGCCTCT CAACTCTTTGCATCACTCCACCACCAGTCTCCAGACGACCATCAA CCCGGACCCCTGTTAGCCCGCAAGTTCAAGAGCTTTGCGAACCATCAA CCCGGACCCCTGTTAGCCCGCAAGTTCAAGAGCGTCTCCTTGTAGA CGCAAGGTGTTGTGGGTCTTGACCCGACCC	22	
KLF12	ATGAACATTCACATGAAGCGCAAGACGATAAAGAACATCAATACATTCGA GAACCGAATGTTGATGTTGGATGGCATGCCCGCTGTACGGGTAAAAACCG AGCTCCTGGAGTCTGAACAAGGATCCCCAAACGTCCACAACTACCCGGAT ATGGAGGCAGTGCCGGCTCTTGCTCAACAATGTGAAGGGAGAGCCGCCCTGA GGACTCTCTCCGTAGATCATTTCCAACAACGGCAGACTGAGCCCGGTAGATCT TTCAATTAACAAAGCCAGAACATCTCCTACTGCGGTAAGTTCTTCTCCCGT AAGTATGACAACGAAGTGCATCTAGTCCAAGTTCTACCAGACACTAGCACTTC	23	

TTCATCTAGTAGACTTGCTAGTTCACCAACGGTGATCACAAGTGTTTCTAG CGCCAGCAGCAGCACGGTCAACGGTACTGACTCCCGGTCCACTCGTGGCAAGCG CTAGTGGCGTGGGTGGCCAACAATTTCTCCCATATTATTCACCCCGTGCCTC

### -continued

GENE	SEQUENCE	SEQ ID NO:
	CGTCTAGTCCGATGAATCTCCAGAGCAACAAGCTTAGTCACGTACATAGGA TCCCCGTCGTCGTCCAGTCAGTTCCCGTCGTCTACACAGCTGTGCGATCCCC TGGGAATGTCAATAATACTATAGTTGTTCCTTTGCTTGAGGATGGTAGGGG CCATGGGAAAGCACGATGATGGACCCCGCGGCGTTGTCACCGAGACAGTCTA AATCCGATAGTGACACGATGATTTGCCTAAGGTACAGGACAGTCTGA ACGAGACCGGGAGTACCGCTCTGTCAATCGCTAGGGCCGTACAGGAGGTC CACCCAAGCCCTGTGTCACGAGTCCGAGGTAACAGGATGAATAATCAGAA ATTTCCCTGTAGCATCAGCCCATTTTCTATAGAGTCCACTCGGAGACAGCG ACGAAGTGAATCACCCGACTCCAGAAAAAGGAGATACATCGCTGTGACT TTGAGGGCTGTAACAGGTCTACCGTACAGAGAGATCACACCTCAGGGCAT CGACGGAGCATACTGGGCAAAAACCGTACAAAAGTCACACCTCAGGGGATG CACGTGGAAATTGCCGGCTCTGACGAGATGAATGCACTCGGGAGGGA	
KLF13	ATGGCTGCGGCTGCATATGTGGATCATTTTGCGGCTGAGTGCCTGGTGTCA ATGTCTAGTAGAGCGGTGGTACACCGGTCCCAGAGAGGCCCAGAATCACG CCCAGAGGGCGCCGCCGTCGCTGCAACACCGACGCTGCCTCGGGTCGAG AGCGCCGCGACGGAACGGCCGCGCGCACCCGCGCGACGCCG GGAGGGCCGCTGCCAGAACGGCTCCAGCACCATGCCGCTGCGCCGC CTGCGCCAGAACCCACAGGTCCAGGACCATGCCGCTGCGCCCC CTGCGCCAGAACCCACAGTCCAGGTGCCGAAGGTGCGGCGGCGGCTGCCCT CCTTCACCGGCCTGCTGCAGACCAGAC	24
KLF14	ATGTCAGCCGCAGTCGCATGCCTTGATTACTTCGCGGCCGAGTGTCTTGTTT CCATGTCAGCGGGGGCTGTCGTTCACAGAAGACCACCAGACCCCGGAGGGA GCGGGAGGGGCAGCTGGATCTGAAGTCGGCGCGGCTCCACCTGAATCAGC GCTTCCCGGCCCTGGTCCTCGAGGTCCGCGGAGCGGCCCCCAACTCCCTGC AGCATCAGTGTGGGCCGATCTTCCGGGAAGCTCCGGGGAGGGGCCCCTGGG AAAACAGCGGAGAGGCCCCCGCGAGCTTCAAGCGGCTTTTCCGATCCAATC CCTTGCAGTGTTCAAACCCCATGCTCGAGCTCGCGCGCGC	25
KLF15	ATGGTGGACCACTTACTTCCAGTGGACGAGAACTTCTCGTCGCCAAAATGC CCAGTTGGGTATCTGGGTGATAGGCTGGTTGGCCGGCGGGCATATCACATG CTGCCCTCACCCGTCTCTGAAGACAGCGATGCCTCCAGCCCTGGTCC TGTTCCAGTCCCGACTCTCAAGCCTCTGGTCCTATGGTGGAGGCCTG GGCACCGAGAGCCAGGACAGCAGCAGCAGCAGTGGCCAGCAGTGGCCC CTGGGCAGTGGCGGGGGCAGCGGCAGTAGCATTGGGGCCAGCAGTGGCC CTGGCCTGGGGCCCTGGCGAAGGGCAGCAGCAGCCCCTGTGAAAGGGGGAC ATTCTGCTTGCCCGAGTTTCCTTTGGGTGATCCTGATGACGTCCCACGGC CTTCCAGCCTACCCTGGAGAGGACAGCAGCAGCAGCAGCAGCAGCAGCAC AGCCTGGAGTCCAAGGAGGCACCAGGAACAGCAAGGAACATGG AGCCTGGAGTCCAAGGAGGCCCCCTGCAAGGACCTTGGATGCC TGCAGCCAGCTCTCAGCTGGGCCACCACAAGAACTTGGATGCC AGCGGGAGAGAGGCCTGTCCCCCCCACCAGGTGGTGCCAGTGCAGGAGG TGCCCAGGGCCAGGCGGGGCCCCCACGCCTGATGGCCCATCCCAGGTGT CCTGCCAGGCCCAGGCGGGGCCCCCACGCCTGATGGCCCACCGCGTG CCCCCGGGCCAGGCCGTGCCCGGCCCG	26

#### -continued

GENE	SEQUENCE	SEQ NO :	ID
	GGGCCAGAAGTTCCCCAAGAACCCAGCCGCAGAACTCATCAAAATGCACA AATGTACTTTCCCTGGCTGCAGCAAGATGTACACCAAAAGCAGCCACCTCA AGGCCACCTGCGCCGGCACAGGGTGAGAAGCCCTTCGCCTGCCCCTGG CCAGGCTGGGGCTGGAGGTTCTCGCGCTCTGAGAGCTGTCGCGGGCACAG GCGCTCGCACTCAGGTGTGAAGCCGTACCAGTGTCCTGTGTGCGAGAAAA AGTTCGCGCGGAGCGACCACCTCTCCCAAGCACATCAAGGTGCACCGCTTCC CGCGGAGCAGCCGCTCCGTGCGCTCCGTGAAC		
KLF16	ATGTCAGCCGCGGTCGCGTGGGTGGATTATTTTGCAGCAGATGTGCTGATG GCAATTTCATCCGGTGCAGTAGTTCATCGCGGGAAGACCAGGTCCTGAGGGT GCGGGGCCTGCGCGCGGGTTGGATGTTCGCGCCGCGC	27	
KLF17	ATGTACGGCCGACCGCAGGCTGAGATGGAACAGGAGGCTGGGGAGCTGAG CCGGTGGCAGGCGGCGCCCCAGGCTGCCCAGGATAACGAGAACTCAGCGC CCATCTTGAACATGTCTTCATCTTCTGGAAGCTCTGGAGTGCACACCTCTTG GAACCAAGGCCTACCAAGCATTCAGCACTTTCCTCACAGCGCAGAGATGCT GGGGTCCCCTTTGGTGTGAGCAGCCCTGGGGGGGGGG	28	

**[0116]** To further demonstrate the applicability of the network analysis to uncover novel phenomena, Applicants focused on two TFs, SNAI2 and KLF4, which seemed to have opposite effects on the pluripotency module. Since KLF4 and SNAI2 are known to play critical and opposing roles in epithelial-mesenchymal transition (EMT) Applicants assessed whether they cause changes along an EMT-like axis in hPSCs as well. A PCA analysis using 200 genes from a consensus EMT geneset from MSigDB demonstrated a distinct stratification of KLF4-transduced cells towards an epithelial-like state and SNAI2-transduced cells towards a mesenchymal-like state. The scRNA-seq data also demonstrates expression level changes in signature genes consistent with EMT (FIG. **3**C), which Applicants confirmed with qRT-PCR (FIG. **9**).

**[0117]** Finally, Applicants chose to focus on ETV2, which has the greatest average fitness loss across all medium conditions (FIG. 1B), as an exemplary case for investigation of a TF showing markedly reduced fitness in all medium

conditions. Applicants hypothesized that the reduced fitness could be due to a proliferation disadvantage if ETV2transduced cells are undergoing massive reprogramming without division. Focused experiments revealed that while ETV2-transduced cells undergo extensive cell death in pluripotent medium, there is a morphology change, indicative of an endothelial phenotype, in endothelial medium (FIG. 3E). Confirmatory qRT-PCR assays demonstrated a strong upregulation of the key endothelial markers CDH5, PECAM1 and VWF (FIG. 3F). Immunofluorescence revealed a distinct distribution of CDH5, with greater localization at cell-cell junctions (FIG. 3G), consistent with known results. In addition, functional testing confirmed tube formation (FIG. 3H), suggesting that a single TF, ETV2, may be able to drive reprogramming from a pluripotent to an endothelial-like state.

**[0118]** To Applicants' knowledge, this is the first demonstration of a high-throughput gene over-expression screening approach that can simultaneously assay both fitness and
transcriptome-wide effects. Applicants' use of ORF overexpression drove strong phenotypic effects, allowing Applicants to capture subtle transcriptomic signals. Additionally, Applicants demonstrated the versatility of the SEUSS screening platform, by assaying mutant forms of a single TF, and assaying all the TFs in a gene family to uncover patterns and differences. Applicants note that the effects of gene overexpression are context dependent. In Applicants' assays, since hPSCs were transduced with pooled libraries, transcriptomic changes driven by cell-cell interactions could increase variability, even supporting the survival of certain cells or disrupting the pluripotent state of control cells. Applicants also assume, in aggregating multiple batches from independent experiments, that each batch is relatively similar. Additionally, while Applicants believe the gene co-perturbation network is a valuable resource, it is dependent on the set of perturbations and conditions used in the experiment.

**[0119]** Taken together, SEUSS has broad applicability to study the effects of overexpression in diverse cell types and contexts; it may be extended to novel applications such as high-throughput screening of large-scale protein mutagenesis, and is amenable to scale-up. In combination with other methods of genetic and epigenetic perturbation it may allow Applicants to generate a comprehensive understanding of the pluripotent and differentiation landscape.

## Example 1 Methods

## Cell Culture

**[0120]** H1 hESC cell line was maintained under feederfree conditions in mTeSR1 medium (Stem Cell Technologies). Prior to passaging, tissue-culture plates were coated with growth factor-reduced Matrigel (Corning) diluted in DMEM/F-12 medium (Thermo Fisher Scientific) and incubated for 30 minutes at 37° C., 5% CO<sub>2</sub>. Cells were dissociated and passaged using the dissociation reagent Versene (Thermo Fisher Scientific).

#### Library Preparation

[0121] A lentiviral backbone plasmid was constructed containing the EF1 $\alpha$  promoter, mCherry transgene flanked by BamHI restriction sites, followed by a P2A peptide and hygromycin resistance enzyme gene immediately downstream. Each transcription factor in the library was individually inserted in place of the mCherry transgene. Since the ectopically expressed transcription factor would lack a polyadenylation tail due to the presence of the 2A peptide immediately downstream of it, the transcript will not be captured during single-cell transcriptome sequencing which relies on binding the poly-adenylation tail of mRNA. Thus, a barcode sequence was introduced to allow for identification of the ectopically expressed transcription factor. The backbone was digested with HpaI, and a pool of 20 bp long barcodes with flanking sequences compatible with the HpaI site, was inserted immediately downstream of the hygromycin resistance gene by Gibson assembly. The vector was constructed such that the barcodes were located only 200 bp upstream of the 3'-LTR region. This design enabled the barcodes to be transcribed near the poly-adenylation tail of the transcripts and a high fraction of barcodes to be captured during sample processing for scRNA-seq.

[0122] To create the transcription factor library, individual transcription factors were PCR amplified out of a human cDNA pool (Promega Corporation) or obtained as synthesized double-stranded DNA fragments (gBlocks, IDT Inc) with flanking sequences compatible with the BamHI restriction sites. MYC mutants were obtained as gBlocks with a 6-amino acid GSGSGS linker (SEQ ID NO: 29) substituted in place of deleted domains (Table 1). The lentiviral backbone was digested with BamHI HF (New England Biolabs) at 37° C. for 3 hours in a reaction consisting of: lentiviral backbone, 4 µg, CutSmart buffer, 5 µl, BamHI, 0.625 µl, H<sub>2</sub>0 up to 50 µl. After digestion, the vector was purified using a QIAquick PCR Purification Kit (Qiagen). Each transcription factor vector was then individually assembled via Gibson assembly. The Gibson assembly reactions were set up as follows: 100 ng digested lentiviral backbone, 3:10 molar ratio of transcription factor insert, 2× Gibson assembly master mix (New England Biolabs), H<sub>2</sub>0 up to 20 µl. After incubation at 50° C. for 1 h, the product was transformed into One Shot Stb13 chemically competent Escherichia coli (Invitrogen). A fraction (150 µL) of cultures was spread on carbenicillin (50 µg/ml) LB plates and incubated overnight at 37° C. Individual colonies were picked, introduced into 5 ml of carbenicillin (50 µg/ml) LB medium and incubated overnight in a shaker at 37° C. The plasmid DNA was then extracted with a QIAprep Spin Miniprep Kit (Qiagen), and Sanger sequenced to verify correct assembly of the vector and to extract barcode sequences.

**[0123]** To assemble the library, individual transcription factor vectors were pooled together in an equal mass ratio along with a control vector containing the mCherry transgene which constituted 10% of the final pool.

#### Viral Production

[0124] HEK 293T cells were maintained in high glucose DMEM supplemented with 10% fetal bovine serum (FBS). In order to produce lentivirus particles, cells were seeded in a 15 cm dish 1 day prior to transfection, such that they were 60-70% confluent at the time of transfection. For each 15 cm dish 36 µl of Lipofectamine 2000 (Life Technologies) was added to 1.5 ml of Opti-MEM (Life Technologies). Separately 3 µg of pMD2.G (Addgene no. 12259), 12 µg of pCMV delta R8.2 (Addgene no. 12263) and 9  $\mu g$  of an individual vector or pooled vector library was added to 1.5 ml of Opti-MEM. After 5 minutes of incubation at room temperature, the Lipofectamine 2000 and DNA solutions were mixed and incubated at room temperature for 30 minutes. During the incubation period, medium in each 15 cm dish was replaced with 25 ml of fresh, pre-warmed medium. After the incubation period, the mixture was added dropwise to each dish of HEK 293T cells. Supernatant containing the viral particles was harvested after 48 and 72 hours, filtered with 0.45 µm filters (Steriflip, Millipore), and further concentrated using Amicon Ultra-15 centrifugal ultrafilters with a 100,000 NMWL cutoff (Millipore) to a final volume of 600-800 µl, divided into aliquots and frozen at -80° C.

## Viral Transduction

**[0125]** For viral transduction, on day -1, H1 cells were dissociated to a single cell suspension using Accutase (Innovative Cell Technologies) and seeded into Matrigel-coated plates in mTeSR containing ROCK inhibitor, Y-27632 (10

 $\mu$ M, Sigma-Aldrich). For transduction with the TF library, cells were seeded into 10 cm dishes at a density of  $6 \times 10^6$  cells for screens conducted in mTeSR or  $4.5 \times 10^6$  cells for screens conducted in endothelial growth medium (EGM) or multilineage (ML) medium (DMEM+20% FBS.) For transduction with individual transcription factors cells were seeded at a density of  $4 \times 10^5$  cells per well of a 12 well plate for experiments conducted in mTeSR or  $3 \times 10^5$  cells per well for experiments conducted in the alternate media.

**[0126]** On day 0, medium was replaced with fresh mTeSR to allow cells to recover for 6-8 hours. Recovered cells were then transduced with lentivirus added to fresh mTeSR containing polybrene (5  $\mu$ g/ml, Millipore). On day 1, medium was replaced with the appropriate fresh medium: mTeSR, endothelial growth medium or high glucose DMEM+20% FBS. Hygromycin (Thermo Fisher Scientific) selection was started from day 2 onward at a selection dose of 50  $\mu$ g/ml, medium containing hygromycin was replaced daily.

#### Single Cell Library Preparation

**[0127]** For screens conducted in mTeSR cells were harvested 5 days after transduction while for alternate media, EGM or ML, cells were harvested 6 days after transduction with the TF library. Cells were dissociated to single cell suspensions using Accutase (Innovative Cell Technologies). For samples sorted with magnetically assisted cell sorting (MACS), cells were labelled with anti-TRA-1-60 antibodies or with dead cell removal microbeads and sorted as per manufacturer's instructions (Miltenyi Biotec). Samples were then resuspended in 1×PBS with 0.04% BSA at a concentration between 600-2000 per  $\mu$ l. Samples were loaded on the 10× Chromium system and processed as per manufacturer's instructions (10× Genomics). Unused cells were centrifuged at 300 rcf for 5 minutes and stored as pellets at -80° C. until extraction of genomic DNA.

**[0128]** Single cell libraries were prepared as per the manufacturer's instructions using the Single Cell 3' Reagent Kit v2 (10× Genomics). Prior to fragmentation, a fraction of the sample post-cDNA amplification was used to amplify the transcripts containing both the TF barcode and cell barcode.

# Barcode Amplification

**[0129]** Barcodes were amplified from cDNA generated by the single cell system as well as from genomic DNA from cells not used for single cell sequencing. Barcodes were amplified from both types of samples and prepared for deep sequencing through a two-step PCR process.

**[0130]** For amplification of barcodes from cDNA, the first step was performed as three separate 50  $\mu$ l reactions for each sample. 2  $\mu$ l of the cDNA was input per reaction with Kapa Hifi Hotstart ReadyMix (Kapa Biosystems). The PCR primers used were, Nexterai7\_TF\_Barcode\_F: GTCTCGTGGGGCTCGGAGATGTGTATAAGA-

GACAGAGAACTATTTCCTGGCTGTTACG CG (SEQ ID NO: 30) and NEBNext Universal PCR Primer for Illumina (New England Biolabs). The thermocycling parameters were 95° C. for 3 min; 26-28 cycles of 98° C. for 20 s; 65° C. for 15 s; and 72° C. for 30 s; and a final extension of 72° C. for 5 min. The numbers of cycles were tested to ensure that they fell within the linear phase of amplification. Amplicons (~500 bp) of 3 reactions for each sample were pooled, size-selected and purified with Agencourt AMPure XP beads at a 0.8 ratio. The second step of PCR was performed with two separate 50  $\mu$ l reactions with 50 ng of first step purified PCR product per reaction. Nextera XT Index primers were used to attach Illumina adapters and indices to the samples. The thermocycling parameters were: 95° C. for 3 min; 6-8 cycles of (98° C. for 20 s; 65° C. for 15 s; 72° C. for 30 s); and 72° C. for 5 min. The amplicons from these two reactions for each sample were pooled, size-selected and purified with Agencourt AMPure XP beads at a 0.8 ratio. The purified second-step PCR library was quantified by Qubit dsDNA HS assay (Thermo Fisher Scientific) and used for downstream sequencing on an Illumina HiSeq platform.

**[0131]** For amplification of barcodes from genomic DNA, genomic DNA was extracted from stored cell pellets with a DNeasy Blood and Tissue Kit (Qiagen). The first step PCR was performed as three separate 50 µl reactions for each sample. 2 µg of genomic DNA was input per reaction with Kapa Hifi Hotstart ReadyMix. The PCR primers used were, NGS\_TF-Barcode\_F: ACACTCTTTCCCTA-CACGACGCTCTTCCGATCTAGAACTAT-

TTCCTGGCTGTTACGCG (SEQ ID NO: 31) and NGS\_TF-Barcode R:

GACTGGAGTTCAGACGTGTGCTCTTCC-

GATCTTGTCTTCGTTGGGAGTGAATTAGC (SEQ ID NO: 32). The thermocycling parameters were: 95° C. for 3 min; 26-28 cycles of 98° C. for 20 s; 55° C. for 15 s; and 72° C. for 30 s; and a final extension of 72° C. for 5 min. The numbers of cycles were tested to ensure that they fell within the linear phase of amplification. Amplicons (200 bp) of 3 reactions for each sample were pooled, size-selected with Agencourt AMPure XP beads (Beckman Coulter, Inc.) at a ratio of 0.8, and the supernatant from this was further size-selected and purified at a ratio of 1.6. The second step of PCR was performed as two separate 50 µl reactions with 50 ng of first step purified PCR product per reaction. Next Multiplex Oligos for Illumina (New England Biolabs) Index primers were used to attach Illumina adapters and indices to the samples. The thermocycling parameters were: 95° C. for 3 min; 6 cycles of (98° C. for 20 s; 65° C. for 20 s; 72° C. for 30 s); and 72° C. for 2 min. The amplicons from these two reactions for each sample were pooled, size-selected with Agencourt AMPure XP beads at a ratio of 0.8, and the supernatant from this was further size-selected and purified at a ratio of 1.6. The purified second-step PCR library was quantified by Qubit dsDNA HS assay (Thermo Fisher Scientific) and used for downstream sequencing on an Illumina MiSeq platform.

Single Cell RNA-Seq Processing and Genotype Deconvolution

**[0132]** Using the  $10 \times$  genomics CellRanger pipeline [citation], Applicants aligned Fastq files to hg38, counted UMIs to generate counts matrices, and aggregated samples across  $10 \times$  runs with cellranger aggr. All cellranger commands were run using default settings.

**[0133]** To assign one or more transcription factor genotypes to each cell, Applicants aligned the plasmid barcode reads to hg38 using BWA, and then labeled each read with its corresponding cell and UMI tags. To remove potential chimeric reads, Applicants used a two-step filtering process. First, Applicants only kept UMIs that made up at least 0.5% of the total amount of reads for each cell. Applicants then counted the number of UMIs and reads for each plasmid barcode within each cell, and only assigned that cell any barcode that contained at least 10% of the cell's read and UMI counts. Barcodes were mapped to transcription factors within one edit distance of the expected barcode. The code for assigning genotypes to each cell can be found on github at: github.com/yanwu2014/genotyping-matrices

#### Clustering and Cluster Enrichment

[0134] Clustering was performed on the aggregated counts matrices using the Seurat pipeline. Applicants first filtered the counts matrix for genes that are expressed in at least 2% of cells, and cells that express at least 500 genes. Applicants then normalized the counts matrix, found overdispersed genes, and used a negative binomial linear model to regress away library depth, batch effects, and mitochondrial gene fraction. Applicants performed PCA on the overdispersed genes, keeping the first 20 principal components. Applicants then used the PCs to generate a K Nearest Neighbors graph, with K=30, used the KNN graph to calculate a shared nearest neighbors graph, and used a modularity optimization algorithm on the SNN graph to find clusters. Clusters were recursively merged until all clusters could be distinguished from every other cluster with an out of the box error (oobe) of less than 5% using a random forest classifier trained on the top 15 genes by loading magnitude for the first 20 PCs. Applicants used tSNE on the first 20 PCs to visualize the results.

**[0135]** Cluster enrichment was performed using Fisher's exact test, testing each genotype for over-enrichment in each cluster. The p-value from the Fisher test for each genotype and cluster combination was corrected using the Benjamini-Hochberg method.

Differential Expression, Identification of Significant Genotypes, and Genotype Trimming

**[0136]** Applicants used a modified version of the MIMOSCA linear model to analyze the differentially expressed genes for each genotype. In this model, Applicants used the R glmnet package with the multigaussian family, with alpha (the lasso vs ridge parameter) set to 0.5. Lambda (the coefficient magnitude regularization parameter) was set using 5-fold cross validation.

**[0137]** In order to account for unperturbed cells, Applicants "trimmed" the cells in each transcription factor genotype to only include cells that belonged to a cluster that the genotype was enriched for. Specifically, Applicants first obtained a set of transcription factor genotypes with strong cluster enrichment, such that each significantly enriched genotype was enriched for a cluster with an FDR>1e-6, and whose cluster enrichment profile was different from the control mCherry profile with an adjusted chi-squared p-value of less than 1e-6. For each significantly enriched genotype, Applicants only kept cells that were part of a cluster that the genotype can be enriched for at FDR<0.01 level. Each genotype can be enriched for more than one cluster. After trimming the significantly enriched genotypes, Applicants repeated the differential expression.

**[0138]** TFs were chosen as significant for downstream analysis if they were enriched for one or more clusters as described, or if the TF drove statistically significant differential expression of greater than 100 genes.

Gene Co Perturbation Network and Module Detection

**[0139]** Applicants took the genes by genotypes coefficients matrix from the regression analysis with trimmed

genotypes and used it to calculate the Euclidean distance between genes, using the significant genotypes as features. Applicants then built a k-nearest neighbors graph from the Euclidean distances between genes, with k=30. From this kNN graph, Applicants calculated the fraction of shared nearest neighbors (SNN) for each pair of genes to build and SNN graph. For example, if two genes share 23/30 neighbors, Applicants create an edge between them in the SNN graph with a weight of 23/30=0.767.

**[0140]** To identify gene modules, Applicants used the Louvain modularity optimization algorithm. For each gene module, Applicants identified enriched Gene Ontology terms using Fisher's exact test (Table 5). Applicants also ranked genes in each gene module by the number of enriched Gene Ontology terms the gene is part of, to identify the most biologically significant genes in each module (Table 5). Gene module identities were assigned based on manual inspection of enriched GO terms and the genes within each module. The effect of each genotype on a gene module was calculated by taking the average of the regression coefficients for the genotype and the genes within the module.

# Dataset Correlation

**[0141]** To compare how the combined hPSC medium dataset correlated with the five individual datasets, Applicants correlated the regression coefficients of the combined dataset with the coefficients for each individual dataset, subsetting for coefficients that were statistically significant in either the individual dataset, or the combined dataset. Each coefficient represents the effect of a single TF on a single gene. The two datasets for the multilineage lineage screens were correlated in the same manner.

## Fitness Effect Analysis

**[0142]** To calculate fitness effects from genomic DNA reads, Applicants first used MagECK to align reads to genotype barcodes and count the number of reads for each genotype in each sample, resulting in a genotypes by samples read counts matrix. Applicants normalized the read counts matrix by dividing each column by the sum of that column, and then calculated log fold-change by dividing each sample by the normalized plasmid library counts, and then taking a log 2 transform. For the stem cell media, Applicants averaged the log fold change across the non MACS sorted samples.

**[0143]** To calculate fitness effects from genotype counts identified from single cell RNA-seq, Applicants used a cell counts matrix instead of a read counts matrix, and repeated the above protocol.

## Epithelial Mesenchymal Transition Analysis

**[0144]** Applicants took 200 genes from the Hallmark Epithelial Mesenchymal Transition geneset from MSigDB and ran PCA on those genes with the stem cell medium dataset, visualizing the first two principal components. The first principal component was an EMT-like signature and Applicants used the gene loadings, along with literature research to identify a relevant panel of EMT related genes to display. All analysis code can be found at github.com/ yanwu2014/SEUS S-Analysis.

# RNA Extraction, and qRT-PCR

[0145] RNA was extracted from cells using the RNeasy Mini Kit (Qiagen) as per the manufacturer's instructions. The quality and concentration of the RNA samples was measured using a spectrophotometer (Nanodrop 2000, Thermo Fisher Scientific). cDNA was prepared using the Protoscript II First Strand cDNA synthesis kit (New England Biolabs) in a 20 µl reaction and diluted up to 1:5 with nuclease-free water. qRT-PCR reactions were setup as: 2 µl cDNA, 400 nM of each primer, 2× Kapa SYBR Fast Master Mix (Kapa Biosystems), H<sub>2</sub>O up to 20 µl. qRT-PCR was performed using a CFX Connect Real Time PCR Detection System (Bio-Rad) with the thermocycling parameters: 95° C. for 3 min; 95° C. for 3 s; 60° C. for 20 s, for 40 cycles. All experiments were performed in triplicate and results were normalized against a housekeeping gene, GAPDH. Relative mRNA expression levels, compared with GAPDH, were determined by the comparative cycle threshold  $(\Delta\Delta C_T)$ method. Primers used for qRT-PCR are listed in Table 6.

#### Immunofluorescence

**[0146]** Cells were fixed with 4% (wt/vol) paraformaldehyde in PBS at room temperature for 30 minutes. Cells were then incubated with a blocking buffer: 5% donkey serum, 0.2% Triton X-100 in PBS for 1 hour at room temperature followed by incubation with primary antibodies diluted in the blocking buffer at 4° C. overnight. Primary antibodies used were: VE-Cadherin (D87F2, Cell Signaling Technology; 1:400). Secondary antibodies used were: DyLight 488 labelled donkey anti-rabbit IgG (ab96891, Abcam; 1:250).

**[0147]** After overnight incubation with primary antibodies, cells were labelled with secondary antibodies diluted in 1% BSA in PBS for 1 hour at 37° C. Nuclear staining was done by incubating cells with DAPI for 5 minutes at room temperature. All imaging was conducted on a Leica DMi8 inverted microscope equipped with an Andor Zyla sCMOS camera and a Lumencor Spectra X multi-wavelength fluorescence light source.

## Endothelial Tube Formation Assay

**[0148]** A mCherry expressing H1 cell line was created by transducing H1 cells with a lentivirus containing the EF1 $\alpha$  promoter driving expression of the mCherry transgene, internal ribosome entry site (IRES) and a puromycin resistance gene. Cells were then maintained under constant puromycin selection at a dose of 0.75 µg/ml. mCherry labelled H1 cells were transduced with either ETV2 lentivirus or control mCherry lentivirus, hygromycin selection was started on day 2 and cells were used for tube formation assay on day 6.

**[0149]** Growth-factor reduced Matrigel (Corning) was thawed on ice and 250  $\mu$ l was deposited cold per well of a 24-well plate. The deposited Matrigel was incubated for 60 minutes at 37° C., 5% CO<sub>2</sub>, to allow for complete gelation and the ETV2-transduced or control cells were then seeded on it at a density of  $3.2 \times 10^5$  cells per well in a volume of 500  $\mu$ l EGM. Imaging was conducted 24 hours after deposition of the cells.

#### Example 2

#### Corneal Endothelial Stem Cell Transplant

[0150] Skin fibroblasts are isolated from a patient with a corneal eye disease. iPSCs are generated from the fibroblasts using techniques known in the art. Briefly, the isolated fibroblasts are reprogrammed by forced expression of one or more pluripotency genes selected from: OCT3/4, SOX1, SOX2, SOX15, SOX18, KLF1, KLF2, KLF4, KLF5, n-MYC, c-MYC, L-MYC, NANOG, LIN28, and GLIS1. [0151] Next, the iPSCs are directed to differentiate into endothelial cells by introducing expression of ETV2. Expression is introduced by infecting the cells with an AAV virus encoding ETV2. After the cells differentiate into endothelial cells, they are expanded ex vivo and harvested. [0152] The cells are administered to the patient by transplant to the cornea following removal of the diseased corneal tissue. After corneal transplant with the endothelial cells, repair of the cornea is identified by achieving full or partial restoration of corneal function in the patient.

TABLE 1

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
mCherry	ATGGTGAGCAAGGGCGAGGAGGAT	33	Non-functional	
Control	AACATGGCCATCATCAAGGAGTTC		control vector	
	ATGCGCTTCAAGGTGCACATGGAG			
	GGCTCCGTGAACGGCCACGAGTTC			
	GAGATCGAGGGCGAGGGCGAGGGC			
	CGCCCCTACGAGGGCACCCAGACC			
	GCCAAGCTGAAGGTGACCAAGGGT			
	GGCCCCCTGCCCTTCGCCTGGGACA			
	TCCTGTCCCCTCAGTTCATGTACGG			
	CTCCAAGGCCTACGTGAAGCACCC			
	CGCCGACATCCCCGACTACTTGAAG			
	CTGTCCTTCCCCGAGGGCTTCAAGT			
	GGGAGCGCGTGATGAACTTCGAGG			
	ACGGCGGCGTGGTGACCGTGACCC			
	AGGACTCCTCCCTGCAGGACGGCG			
	AGTTCATCTACAAGGTGAAGCTGC			
	GCGGCACCAACTTCCCCTCCGACGG			
	CCCCGTAATGCAGAAGAAGACCAT			
	GGGCTGGGAGGCCTCCTCCGAGCG			
	GATGTACCCCGAGGACGGCGCCCT			
	GAAGGGCGAGATCAAGCAGAGGCT			
	GAAGCTGAAGGACGGCGGCCACTA			
	CGACGCTGAGGTCAAGACCACCTA			

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CAAGGCCAAGAAGCCCGTGCAGCT GCCCGGCGCCTACAACGTCAACAT CAAGTTGGACATCACCTCCCACAAC GAGGACTACACCATCGTGGAACAG TACGAACGCCGCCGAGGGCCGCCAC TCCACCGGCGGCATGGACGAGCTG TACAAG			
ASCL1	ATGGAGTCTTCTGCTAAAATGGAGT CCGGAGGCGCGGGACAACAACCAC AACCGCAACCACAACAACCACTCCT GCCGCCGCCGCGCACATCTTTTTCCCG GCGGCTGCTGCTGCTGCAGCGGCGG GCGGCTGCTGCCGCGCGCGAATCC GCCAACAGCAACAACAACAACAG CAGCAGCAGCAACAACAACAACAG CAGCAGCAGCAACAACAACAACAG CCTCCGAAGCAGCACAAGAGCG GCTCCGAAGCAGGGTAAAAGGCAG GCTCCGAAGCAGGTTAAAAGGCAG AGGAGCAGTAGTCCCGAACTGATG CGATGTAAGAGGCGCCTCAATTTTA GCGGTTTTGGTTACTCTTTGCCCCA GCAGCAGCCGGCTGCCGTAGCTCG CCGAAATGAGCGGGAAAGGAACCG CGTTAAACTTGTGAATCTCGGTTTC GCGACATTGAGAGCACGTACCA AATGGGGCAGCTAACAAGAAATG AGTAAAGTTGAGACACTGCGGTCT GCAGTGGGAGTATATTAGAGCTCTTC AACAATTGCTTGACGAGCACGATG CCGTATCAGCCGCATTTCAAGCGGG GGTGCTGTCCCCAACAATATCTCCG AACTACAGCAATGATCTCGGTTCC CCGAATGAGCACGATTCCAAGCAGATG CCGTATCAGCCGCATTTCAAGCCGG GGTGCTGTCCCCCAACAATATCCCCG AACTACAGCAATGATCTTATAAGC ATGGCGGGAAGTCCCGTTTCCTCCT ACTCCTCTGATGAGGGCAGCTACG ACCTCTCAGTCCCGAGGAGCAAG AGCTTCTTGACTTCACTAACTGGTT	34	Involved in neuronal specification and differentiation. Demonstrated to drive neuronal differentiation from hPSCs	<pre>Wilkinson, G. et al. Proneural genes in neocortical development. Neuroscience 253, 256-273 (2013). Chanda, S. et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. Stem cell reports 3, 282-96 (2014).</pre>
ASCL3	ATGATGGACAACAGAGGCAACTCT AGTCTACCTGACAAACTTCCTATCT TCCCTGATTCTGCCCGCTTGCCACT TACCAGGTCCTTCTATCTGGAGCCC ATGGTCACTTTCCACGTGCACCCAG AGGCCCCGGTGCCACGCTGCCTTAT CCCAGCGACTCTCTTATCCTGGGAA ATTACAGTGAACCCTGCCCTTCT TTTCCCGATGCCTTATCCAAATTAC AGAGGGTGCGAGTACTCCTACGGG CCAGCCTTCACCGGAAAAGGAAT GAGCGGAAAGGCAGCGGGTGAAA TGTGTCAATGAAGGCACCGGGTGAAA TGTGTCAATGAAGGCACCGAGCAGT ATTTGGAAAGCCACCGAGCACCA GGAAACCTCACGAGCTCCAGCATCA	35	Involved in salivary gland cell development	Bullard, T. et al. Ascl3 expression marks a progenitor population of both acinar and ductal cells in mouse salivary glands. Dev. Biol. 320, 72- 78(2008)

TCTGTACCCTGATAAAGCTGAGACA AAGAATAACCCTGGAAAAGTTTCC TCCATGATAGCAACCACCAGCCAC CATGCTGACCCTATGTTCAGAATTG TTTGCCCAACTTTCTTGTACAAAGT

TGTCCCC

TABLE 1-continued

TABLE 1-continued

		SEQ ID		
GENE	SEQUENCE	NO :	ROLE	REFERENCES
ASCL4	ATGGAGACGCGTAAACCGGCGGAA	36	Involved in	Jonsson, M. et
	CGGCTGGCCTTGCCATACTCGCTGC		development of	al. Hash4, a
	GCACCGCGCCCCTGGGCGTTCCGG		skin	novel human
	GGACCCTGCCCGGACTCCCGCGGA			achaete-scute
	GGGACCCCCTCAGGGTCGCCCTGC			homologue
	GTCTGGACGCCGCGTGCTGGGAGT			found in fetal
	GGGCGCGCAGCGGCTGCGCACGGG			skin.
	GATGGCAGTACTTGCCCGTGCCGCT			Genomics 84,
	GGACAGCGCCTTCGAGCCCGCCTTC			859-866
	CTCCGCAAGCGCAACGAGCGCGAG			(2004)
	CGGCAGCGGGTGCGCTGCGTGAAC			
	GAGGGCTATGCGCGCCTCCGAGAC			
	CACCTGCCCCGGGAGCTGGCAGAC			
	AAGCGCCTCAGCAAAGTGGAGACG			
	CTCCGCGCTGCCATCGACTACATCA			
	AGCACCTGCAGGAGCTGCTGGAGC			
	GCCAGGCCTGGGGGCTCGAGGGCG			
	CGGCCGGCGCCGTCCCCCAGCGCA			
	GGGCGGAATGCAACAGCGACGGGG			
	AGTCCAAGGCCTCTTCGGCGCCTTC			
	GCCCAGCAGCGAGCCCGAGGAGGG			
	GGGCAGC			
ASCL5	ATGCCGATGGGGGCAGCAGAAAGA	37	Paralog of	Wang, C. et
	GGTGCTGGGCCCCAATCATCTGCAG		ASCL4	al. Systematic
	CACCATGGGCTGGTTCAGAAAAGG			analysis of the
	CGGCAAAGAGAGGGCCATCAAAAA			achaete-scute
	GCTGGTACCCAAGAGCTGCTGCATC			complex-like
	TGATGTCACGTGCCCGACTGGTGGT			gene signature
	GATGGAGCTGACCCAAAACCTGGA			in clinical
	CCTTTTGGAGGTGGTTTAGCTTTAG			cancer
	GGCCTGCGCCCAGAGGAACAATGA			patients.
	ATAATAATTTCTGCAGGGCCCTTGT			Molecular and
	TGACAGAAGGCCTTTAGGACCCCCT			Clinical
	TCATGTATGCAATTAGGTGTAATGC			Oncology 6,
	CACCGCCAAGACAAGCGCCCCTCC			(Spandidos
	CGCCGGCTGAACCCCTTGGAAATGT			Publications,
	ACCTTTCCTCCTATACCCTGGCCCA			2017).
	GCTGAACCACCATATTATGATGCAT			
	ATGCTGGTGTTTTCCCATATGTGCC			
	TTTCCCTGGTGCTTTTGGTGTATAT			
	GAATACCCTTTTGAGCCGGCTTTTA			
	TCCAAAAGAGGAATGAAAGAGAGA			
	GACAGAGAGTGAAGTGTGTGAATG			
	AAGGATACGCCAGATTGAGAGGCC			
	ATTTGCCTGGTGCCCTGGCAGAAAA			
	GAGATTATCAAAAGTTGAAACCCT			
	GAGGGCGGCAATCAGATATATAAA			
	ATACCTCCAAGAACTCCTTTCATCA			
	GCACCTGATGGATCGACACCACCG			
	GCTTCAAGAGGTTTACCTGGAACTG			
	GACCATGCCCTGCACCGCCTGCTAC			
	ACCAAGGCCAGACAGACCTGGAGA			
	TGGAGAAGCAAGAGCACCTTCTTC			
	CCTTCTCCCTCAATCTCTCTCTCAATCA			
	AGAAAGIGAAGAATUUTGGUA			

		SEQ ID		
SENE	SEQUENCE	NO :	ROLE	REFERENCES
አጥሮን	ATCCCACACACCACACACCCTTTCTC	20	Involved in	Peters C S
AIF/	TIGGAGACGACAGACCGIIIGIG	50		receib, C. S.
			earry cerr	et al. AIF-7,
	AGATITACAAACGAGGACCACCIG		signaling, binds	a novel billp
	GCAGTTCATAAACACAAGCATGAG		CAMP response	protein,
	ATGACATTGAAATTTGGCCCAGCCC		element	interacts with
	GAACTGACTCAGTCATCATTGCAGA			the PRL-1
	TCAAACGCCTACTCCAACTAGATTC			protein-
	CTGAAGAACTGTGAGGAGGTGGGA			tyrosine
	CTCTTCAATGAACTAGCTAGCTCCT			phosphatase.
	TTGAACATGAATTCAAGAAAGCTG			J. Biol. Chem.
	CAGATGAGGATGAGAAAAAGGCAA			276, 13718-
	GAAGCAGGACTGTTGCCAAAAAAC			26 (2001).
	TGGTGGCTGCTGCTGGGCCCCTTGA			Hamard, PJ.
	CATGTCTCTGCCTTCCACACCAGAC			et al. A
	ATCAAAATCAAAGAAGAAGAGCCA			functional
	GTGGAGGTAGACTCATCCCCACCTG			interaction
	ATAGCCCTGCCTCTAGTCCCTGTTC			between
	CCCACCACTGAAGGAGAAGGAGGAGGT			ATF7 and
	TACCCCAAAGCCTGTTCTGATCTCT			TAF12 that is
	ACCCCCACACCACCATTGTACGTC			modulated by
	CTGGCTCCCTGCCTCTCCACTTGGG			TAF4.
	CTATGATCCACTTCATCCAACCCTT			Oncogene 24,
	CCCTCCCCAACCTCTGTCATCACAC			3472-3483
	AGGCTCCACCATCCAACAGGCAAA			(2005).
	TGGGGTCTCCCACTGGCTCCCTCCC			
	TCTTGTCATGCATCTTGCTAATGGA			
	CAGACCATGCCTGTGTTGCCAGGGC			
	CTCCAGTACAGATGCCGTCTGTTAT			
	ATCGCTGGCCAGACCTGTGTCCATG			
	GTGCCCAACATTCCTGGTATCCCTG			
	GCCCACCAGTTAACAGTAGTGGCTC			
	CATTTCTCCCTCTGGCCACCCTATA			
	CCATCAGAAGCCAAGATGAGACTG			
	AAAGCCACCCTAACTCACCAAGTCT			
	CCTCAATCAATGGTGGTTGTGGAAT			
	GGTGGTGGGTACTGCCAGCACCAT			
	GGTGACAGCCCGCCCAGAGCAGAG			
	CCAGATTCTCATCCAGCACCCTGAT			
	GCCCCATCCCCTGCCCAGCCACAG			
	GTCTCACCAGCTCAGCCCACCCCTA			
	GTACTGGGGGGCGACGGCGCGCA			
	CAGTAGATGAAGATCCAGATGAGC			
	GACGGCAGCGCTTTCTGGAGCGCA			
	ACCGGGCTGCAGCCTCCCGCTGCCG			
	CCAAAAGCGAAAGCTGTGGGTGTC			
	CTCCCTAGAGAAGAAGGCCGAAGA			
	ACTCACTTCTCAGAACATTCAGCTG			
	AGTAATGAAGTCACATTACTACGC			
	AATGAGGTGGCCCAGTTGAAACAG			
	CTACTGTTAGCTCATAAAGACTGCC			
	CAGTCACTGCACTACAGAAAAAGA			
	CTCAAGGCTATTTAGAAAGCCCCA			
	AGGAAAGCTCAGAGCCAACGGGTT			
	CTCCAGCCCCTGTGATTCAGCACAG			
	CTCAGCAACAGCCCCTAGCAATGG			
	CCTCAGTGTTCGCTCTGCAGCTGAA			
	GCTGTGGCCACCTCGGTCCTCACTC			
	AGATGGCCAGCCAAAGGACAGAAC			
	TGAGCATGCCGATACAATCGCATGT			

AATCATGACCCCACAGTCCCAGTCT

GCGGGCAGA

TABLE 1-continued

GENE	SEQUENCE	SEQ II NO:	ROLE	REFERENCES
CDX2	ATGTACGTGAGCTACCTCCTGGACA AGGACGTGAGCATGTACCCTAGCT CCGTGCGCCCCCTCTGGCGGCCTCAA CCTGGCGCCGCAGAACTTCGTCAGC GCTACCACGTGGCGCGGCGC	39	Involved in trophectoderm specification and differentiation	<pre>Strumpf, D. et al. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development 132, 2093- 102 (2005).</pre>
CRX	ATGATGGCGTATATGAACCCGGGG CCCCACATTCTGTCAACGCCTTGG CCCTAAGTGGCCCCAGTGTGGATCT GATGCACCAGGCTGTGCCCTACCCA AGCGCCCCCAGGAGCAGCGCGGGGAGCCCCACCTTCACCCGGAGC CAACTGGAGGAGCTGGAGGCACTG TTTGCCAAGACCCAGTACCCAGAC GTCTATGCCGTGAGGAGGTGGCTC TGAAGATCAATCTGCCTGAGTCCAG GGTTCAGGTTGGTTCAAGAACCGG AGGGCTAAATGCAGGCAGCAGCAGC CCCCCAGGGGCCAGGCAGCAGCAG CCCCCAGGGGCCAGGCAGCAGCAG CCCCCCAGGGGCCAGGAGCGCC CGGCCTGCCAAGAGGAGGCGCAGCC CGGCCTGCCAAGAGCAGCAGCGC CGGCCTGCCAGAGCCCTCTGGGCATCT CAGATTCCTACAGTCCCCACAGAT GTGTGTCCAGACCCTCTGGGCATCT CAGATCCCTCAGGCCCCCCTCTGC CGGCCCCCCAGGCCGCCCCTTGC CTGAGGCCCCCCAGGCCTCTGGCCCCTCTGC CCGCCCCCAGGCCGCCTCTGGCCCCCCCCCC	40	Involved in photoreceptor differentiation	Furukawa, T., Morrow, E. M. & Cepko, C. L. Crx, a novel otx-like homeobox gene, shows photoreceptor- specific expression and regulates photoreceptor differentiation. Cell 91, 531-541 (1997).

CCGTGGATAGCTTGGAATTCAAGG

TABLE 1-continued

TARLE	1-continued
TADUE	I-COILCIIIded

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	ACCCCACGGGCACCTGGAAATTCA			
	CCTACAATCCCATGGACCCTCTGGA			
	CTACAAGGATCAGAGTGCCTGGAA			
	GTTTCAGATCTTG			
ERG	ATGGCCAGCACTATTAAGGAAGCC	41	Involved in	Mclaughlin,
	TTATCAGTTGTGAGTGAGGACCAGT		endothelial cell	F. et al.
	CGTTGTTTGAGTGTGCCTACGGAAC		specification	Combined
	GCCACACCTGGCTAAGACAGAGAT		and	genomic and
	GACCGCGTCCTCCTCCAGCGACTAT		differentiation	antisense
	GGACAGACTTCCAAGATGAGCCCA			analysis
	CGCGTCCCTCAGCAGGATTGGCTGT			reveals that
	CTCAACCCCCAGCCAGGGTCACCAT			the
	CAAAATGGAATGTAACCCTAGCCA			transcription
	GGTGAATGGCTCAAGGAACTCTCCT			factor Erg is
	GATGAATGCAGTGTGGCCAAAGGC			implicated in
	GGGAAGATGGTGGGCAGCCCAGAC			endothelial
	AUUGIIGGGATGAAUTACGGCAGC			differentiation
	IACAIGGAGGAGAGGAGACAIGUCA			Plood 99
	CCCCCARACAIGACCACGACGAG			3332-3330
	CCTACGCTATGGAGTACAGACCAT			(2001)
	GTGCGGCAGTGGCTGGAGTGGGCG			(2001).
	GTGAAAGAATATGGCCTTCCAGAC			
	GTCAACATCTTGTTATTCCAGAACA			
	TCGATGGGAAGGAACTGTGCAAGA			
	TGACCAAGGACGACTTCCAGAGGC			
	TCACCCCCAGCTACAATGCCGACAT			
	CCTTCTCTCACATCTCCACTACCTC			
	AGAGAGACTCCTCTTCCACATTTGA			
	CTTCAGATGATGTTGATAAAGCCTT			
	ACAAAACTCTCCACGGTTAATGCAT			
	GCTAGAAACACAGGGGGTGCAGCT			
	TTTATTTTCCCAAATACTTCAGTAT			
	ATCCTGAAGCTACGCAAAGAATTA			
	CAACTAGGCCAGATTTACCATATGA			
	GCCCCCCAGGAGATCAGCCTGGAC			
	GTCGAAAGCTGCTCAACCATCTCCT			
	GEAGUICUIGICGGACAGUTUCAA			
	CIUTGTACAAGTACCCCTCAGACCT			
	GEGEECCACCETCCAGECCTCCCCG			
	TGACATCTTCCAGTTTTTTTGCTGCC			
	CCAAACCCATACTGGAATTCACCA			
	ACTGGGGGTATATACCCCAACACT			
	AGGCTCCCCACCAGCCATATGCCTT			
	CTCATCTGGGCACTTACTAC			

	CROURNER	SEQ ID	DOLD	DEEDDINGEG
JENE	SEQUENCE	NO :	ROLE	REFERENCES
RCDDC		40	Truelued de	Alermial V
UNACE	AIGICAAACAAAGAICGACACATT GATTCCAGCTGTTCGTCCTTCATCA	42	cardiac	Alaynick, W. A of al FDDv
	AGACGGAACCTTCCAGCCCAGCCT		development	Directs and
	CCCTGACGGACAGCGTCAACCACC		dovolopmono	Maintains the
	ACAGCCCTGGTGGCTCTTCAGACGC			Transition
	CAGTGGGAGCTACAGTTCAACCAT			to Oxidative
	GAATGGCCATCAGAACGGACTTGA			Metabolism in
	CTCGCCACCTCTCTACCCTTCTGCT			the Postnatal
	CCTATCCTGGGAGGTAGTGGGCCTG			Heart. Cell
	TCAGGAAACTGTATGATGACTGCTC			Metab. 6, 13-
	CAGCACCATTGTTGAAGATCCCCAG			24 (2007).
	CACTATCCCCTACCATCCCTAC			
	GCCTGCAAGGCATCCTTCAAGAGG			
	ACAATTCAAGGCAATATAGAATAC			
	AGCTGCCCTGCCACGAATGAATGT			
	GAAATCACAAAGCGCAGACGTAAA			
	TCCTGCCAGGCTTGCCGCTTCATGA			
	AGTGTTTAAAAGTGGGCATGCTGA			
	AAGAAGGGGTGCGTCTTGACAGAG			
	TACGTGGAGGTCGGCAGAAGTACA			
	AGCGCAGGATAGATGCGGAGAACA			
	GCCCATACCTGAACCCTCAGCTGGT			
	TCAGCCAGCCAAAAAGCCATTGCT			
	CTGGTCTGATCCTGCAGATAACAAG			
	ATTGTCTCACATTTGTTGGTGGCTG			
	AACCGGAGAAGATCTATGCCATGC			
	CATCAAAGCCCTCACTACACTGTGT GACTTGGCCCACCCACAGTGGTG			
	GTTATCATTGGATGGGCGAAGCAT			
	ATTCCAGGCTTCTCCACGCTGTCCC			
	TGGCGGACCAGATGAGCCTTCTGC			
	AGAGTGCTTGGATGGAAATTTTGAT			
	CCTTGGTGTCGTATACCGGTCTCTT			
	TCGTTTGAGGATGAACTTGTCTATG			
	CAGACGATTATATAATGGACGAAG			
	ACCAGTCCAAATTAGCAGGCCTTCT			
	TGATCTAAATAATGCTATCCTGCAG			
	CTGGTAAAGAAATACAAGAGCATG			
	AAGCTGGAAAAAGAAGAATTTGTC			
	ACCCTCAAAGCTATAGCTCTTGCTA			
	ATTCAGACTCCATGCACATAGAAG			
	ATGTTGAAGCCGTTCAGAAGCTTCA			
	GGATGTUTTACATGAAGCGCTGCA			
	GGATTATGAAGUTGGUCAGCACAT			
	GARGREGATGACACTCCACTCCAC			
	AGGCAGACCTCTACCALGCCALGCCTG			
	CAGCATTTCTACAACGCCGIG			
	AAGGCAAAGTCCCAATGCACAAAC			
	TTTTTTTGGAAATGTTGGAGGCCAA			
	GGTC			
TV2	ATGGATCTTTGGAACTGGGATGAA	43	Involved in	Lee, D. et al.
	GCTTCCCCTCAAGAAGTTCCCCCCG		haemato-	ER71 acts
	GAAATAAACTCGCGGGGCTTGGAA		endothelial	downstream
	GACTCCCTCGCCTTCCGCAACGCGT		specification	of BMP,
	CTGGGGCGGATGCCCTGGTGGAGC		and	Notch, and
	CTCAGCGGACCCAAACCCTTTGTCT		differentiation,	Wnt signaling
			and in	in blood and
	TTCTGCTTCCCGGATCTTGCTTTGC AACCCCATTACTCCAAACCCCAACCC		vasculogenesis	vessel
	AAGGUGATAUTUCAAUGGUGAUGG			progenitor
	CAGAGACCIGIIGGAAAAGGCACCA GTAGCTCCCTGGCCAGCTTTCCCCCA			Cell Stem
	GIAGCICCCIGGCCAGCIIICCGCA			Cell 2 40-
	CATCCCGAAGTTCCCTGGGGGGGCC			507 (2008)
	GAACCCGACTCCCAAGCCCTTCCCT			207 (2000).
	GGAGTGGTGATTGGACAGATATGG			
	CATGCACAGCCTGGGACAGTTGGT			

CCGGGGCGTCACAGACATTGGGAC

TABLE 1-continued

		050 55			
GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES	
	CAGCCCCACTTGGACCGGGGCCTAT				
	CCCCGCAGCAGGAAGCGAAGGAGC				
	TGCTGGTCAGAACTGTGTGCCCGTG				
	GCTGGTGAGGCTACCAGTTGGTCCA				
	GGGCCCAGGCAGCAGGCAGTAACA				
	CCAGCTGGGATTGCTCAGTGGGGC				
	CTGACGGGGATACTTATTGGGGCTC				
	TGGTCTTGGTGGAGAACCGAGAAC				
	GGACTGTACGATAAGTTGGGGCGG				
	TCCAGCTGGGCCTGATTGTACTACG				
	TCATGGAATCCTGGCTTGCACGCCG				
	ATCA A AGTTCA GCCCTTA CAGTTTC				
	CTCAGAACCTTCCCCCCCAAAGTGAC				
	CGAGCGTCACTGGCGCGATGTCCTA				
	AAACTAATCATCGAGGGCCGATCC				
	AGTTGTGGCAGTTTTTGCTTGAACT				
	CCTTCACGATGGCGCGAGGAGCAG				
	TTGCATCAGATGGACCGGTAACAG				
	CAGGGAGTTCCAATTGTGTGACCCC				
	AAGGAAGTGGCTCGACTGTGGGGT				
	GAGCGCAAACGGAAGCCTGGTATG				
	AATTACGAAAAGTTGAGTAGGGGT				
	TIGUGATATIAUTATAGGUGUGACA				
	AGTACACATACACATTCCCCCCCCC				
	GCGTACCATCTCTTGCATACCCTGA				
	TTGCGCAGGCGGGGGTAGGGGTGC				
	GGAAACACAA				
7LI1	ATGGACGGGACTATTAAGGAGGCT	44	Involved in	Liu, F. et al.	
	CTGTCGGTGGTGAGCGACGACCAG		haemato-	Fli1 Acts at	
	TCCCTCTTTGACTCAGCGTACGGAG		endothelial	the Top of the	
	CGGCAGCCCATCTCCCCAAGGCCG		specification	Transcriptional	
	ACATGACTGCCTCGGGGAGTCCTG		and	Network	
	ACTACGGGCAGCCCCACAAGATCA		differentiation	Driving Blood	
	ACCECCTCCCACCACAGCAGGAGT			and	
	GGATCAATCAGCCAGTGAGGGTCA			Endotnellal	
	TGAATCGATCCACCACCACCACCACCACCACCACCACCACCACCACC			Curr Biol 18	
	TGGACTGCAGCGTTAGCAAATGCA			1234-1240	
	GCAAGCTGGTGGGCGGAGGCGAGT			(2008).	
	CCAACCCCATGAACTACAACAGCT			(/	
	ATATGGACGAGAAGAATGGCCCCC				
	CTCCTCCCAACATGACCACCAACGA				
	GAGGAGAGTCATCGTCCCCGCAGA				
	CCCCACACTGTGGACACAGGAGCA				
	TGTGAGGCAATGGCTGGAGTGGGC				
	CATAAAGGAGTACAGCTTGATGGA				
	GATCGACACATCCTTTTTCCAGAAC				
	AIGGATGGCAAGGAACTGTGTAAA				
	AIGAACAAGGAGGAUTTUUTUUGU GCCDCCDCCCTCTDCDACCCCCD				
	GTGCTGTTGTCACACCTCAGTTACC				
	TCAGGGAAAGTTCACTGCTGGCCTA				
	TAATACAACCTCCCACACCGACCA				
	ATCCTCACGATTGAGTGTCAAAGA				
	AGACCCTTCTTATGACTCAGTCAGA				
	AGAGGAGCTTGGGGCAATAACATG				
	AATTCTGGCCTCAACAAAAGTCCTC				
	CCCTTGGAGGGGCACAAACGATCA				
	GTAAGAATACAGAGCAACGGCCCC				
	AGCCAGATCCGTATCAGATCCTGG				
	GCCCGACCAGCAGTCGCCTAGCCA				
	ACCCIGGAAGCGGGCAGATCCAGC TGTGGGADATTCCTCCTCCTCCAGC				
	CTCCGACAGCGCCAACGCCAGCTGCT				
	TATCACCTGGGAGGGGACCAACGG				
	GGAGTTCAAAATGACGGACCCCGA				
	TGAGGTGGCCAGGCGCTGGGGCGA				
	GCGGAAAAGCAAGCCCAACATGAA				
	TTACGACAAGCTGAGCCGGGCCCT				
	CCGTTATTACTATGATAAAAACATT				
	ATGACCAAAGTGCACGGCAAAAGA				

TABLE 1-continued

GENE	SEQUENCE	SEQ NO:	ID	ROLE	REFERENCES
	TATGCTTACAAATTTGACTTCCACG GCATTGCCCAGGCTCTGCAGCCACA TCCGACCGAGTCGTCCATGTACAAG TACCCTTCTGACATCTCCTACATGC CTTCCTACCATGCCCACCAGCAGAA GGTGAACTTGTCCCTCCCACCCA CCTCCATGCCTGTCACTTCCTCCA GCTTCTTTGGAGCCGCCATCACAATA CTGGACCTCCCCCACGGGGGGAAAT CTACCCCAACCCCACGTGCCTTCAC ACTTAGGCAGCTACTAC				
FOXA1	ACTTAGGCACGTGAGATGGAG GGGCATGAGACAAGCGACTGGAAT TCCTACTACGCGGATACCCAAGAA GCGTATTCTTCAGTTCCCGTAAGCA ATATGAACTCCGGATTGGGGAGCA TGAATAGTATGGACACCAGCGG GCACATGAACACGGCCTCCTTTAA TATGTCATATGCGAACCCCGGCG GCCGCGGGCGTGACGGCCTCCTTTAA TATGTCATATGCGAACCCCGGGGGGAGC GCCGGGGCGTGACGCCCGGGGGGAGC GCCGGGGCGTGACGCCCGGGGGGGG GCCGGAGCGATGAACTCCATGACC GCTGCGGGCCTGCACCACGGGGGAGC GCCGGAGCGATGAACTCCATGACC GCCGGGGCGTGCACGCCAGGGGAGC GCCGGGGCGTGCACGCCGGGGGGGC GCCGGGGCGTGACGCCCGGGGGGGC GCCGGGGCGTGCACGCCGGGGGGGC GCCGGAGCGATGAACTCCATGACC GCTCCGGGCCGTGCACGCCGGGGGGGC GCCGGGGCGTGACGCCCAGGGAATG GGCCCTGCATGAATGGGGCCCAGCAGAC GCCCCGGCGGGCGGAGAGC GCCCCAATGAATGGGATCCCTGCAT GTCCCCTATGGCTCACACGGGGGC GGTGGTGGCGATGCAAACCCTTC AAGCCAAGGTATCCCTCAAGGCC GCTCCATTGCTCACCAGGGCCGC GGTGGTGGCGATGCCAAAACCTTC AAGCGAAGTTATCCTCATGCGAGG CCTCCTTATCATATATATCCTTGAT TACCGATGGCGATCCAGAGCCCC GTCTAAGATGCTGACCAAGCCCC GTCTAAGATGCTGACTCGGAGCCCC GTCTAAGATGCTGACTCTGAAGG ATATACCAGTGGATCATGGACCTTT TTCCTTACTACCGGCAAAACCTAC GAGATGGCAAACCCAATACGCCA TAGCCCGGTAAAGGTCCTGAGTGAG ATATACCAGTGGATCATGGACCTTT TTCCTTACTACCGGCAAAACCTAC GAGATGGCCAAACCTAATGGAC AAGCCGGTAAAGGTCCTATTGG ACCCTTCATCCAGATAGGCCCA AAGCCGGAAAACCTAATGGAC AAGCCGGAAACGATCCAATACGCCA TAGCCAGGAACGATCCAATACGCCA CAGGCAGAAACGATCCAATACGCCA CAGGCCAGGACGCCCGGGGGGGGGG	45		Involved in branching morphogenesis, development of lung, liver, prostate, and pancreas	Friedman, J. R. et al. The Foxa family of transcription factors in development and metabolism. Cell. Mol. Life Sci. 63, 2317-2328 (2006).

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CCTGCAGTATTCTCCATATGGCTCT ACACTTCCTGCTTCTCTTCCATTGG GGTCTGCAAGTGTGACAACGCGCT CCCCAATCGAGCCAAGTGCCCTCG AGCCTGCTTATTATCAAGGAGTATA TTCCCGACCAGTTTTGAATACAAGT			
FOXA2	ATGCTGGGAGCGTGAAGATGGAA GGGCACGAGCCGTCCGACTGGAGG AGCTACTATGCAGAGCCCGACGGGC TACTCCTCCGTGAGCACATGAACG CCGGCCTGGGGATGAACGGCATGA ACACGTACATGAGCAGCGCCGGGCA ACATGAGCGCGGGCACGGGCGCGGG CCGCCATGGGCGGGGCTCCGGGGG CCGCCGTCCTGGGGGGGCGCGG CCCCGGCCGG	46	Involved in branching morphogenesis, development of notochord, lung, liver, prostate, and pancreas.	<pre>Friedman, J. R. et al. The Foxa family of transcription factors in development and metabolism. Cell. Mol. Life Sci. 63, 2317-2328 (2006).</pre>

CAGGGGGTGTACTCCCGGCCCATTA

TGAACTCCTCTTTG

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
FOXA3	ATGCTGGGCTCAGTGAAGATGGAG GCCCATGACCTGGCCGAGTGGAGC TACTACCCGGAGGCGGCGAGTC TACTCCCCGGTGACCCCAGTGCCA CCATGGCCCCCCTCAACTCCTACAT GACCCTGAATCCTCAACTCCTACAT GACCCTGGAGCGCCCCTGCGGC CCCCCCCCCC	47	Involved in cell glucose homeostasis	<pre>Friedman, J. R. et al. The Foxa family of transcription factors in development and metabolism. Cell. Mol. Life Sci. 63, 2317-2328 (2006).</pre>
FOXP1	Argatgcaagaatctroggactaag Acaaaagtaacggtrcagccatc Cagaatgggtcggcggcgcgtcttc Gggaggggcggtctac Ggcggcggtgacatcggaggag Ctgacctcgcccacgccagcag Agcagcaacagtggcatctcataa Accatcagccctctaggagtcccag Cagttggcttaggagtcccag Cagttggcttaggagttggagtccc Agcccttgggagttggagtcctgc Agcccttgggagttggagtcctgt Gggacgaagatgagtggagtcccattt	48	Involved in development of haematopoetic cells, lung and oesophagus, and neuronal development	Hu, H. et al. Foxp1 is an essential transcriptional regulator of B cell development. Nat. Immunol. 7, 819-826 (2006). Shu, W. et al. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. Development 134, 1991- 2000 (2007).

TABLE 1-continued

2000 (2007). Bacon, C. et al. Brain-specific Foxp1 deletion impairs neuronal

GENF	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
GENE	SEQUENCE		NULE	development and causes autistic-like behaviour. Mol. Psychiatry 20, 632-639 (2015).
GATA1	ATGGAGTTCCCTGGCCTGGGGTCCC TGGGGACCTCAGAGCCCTCCCCCA GTTTGTGGATCTGGTTCGTCT TCCACACAGAATCAGGGGTTTCT TCCCCTCTGGGCCTGAGGGGTTGGA TGCAGAGACACTCCCCACTGCGCG GCACTGGCCTACTACAGGAGAGCT GAGGCCTACAGACACTCCCCAGTCT TTCAGGTGTACCCATGCTCAACTG TATGGAGGGGATCCCAGGGGCGTACGG CAAGACGGGGCTCACCAGGGGCCTACGG CAAGACGGGGCTCTACCCAGGGGCCTACGG CAAGACGGGGCTCTACCCCGCGAGGACT CTCCTCCCCAGGCGTGGCCTACGG GCTGAGCCCAGACGCCGCGGAGACT CCTCTCCCCCAGGCGTGGACAGCT GGATGGAAAAGCCACACCCTG GGCTGAGCCCAGCCTCCTGACCCTG GGCCGACACTGCCTCAGAGGG CCTGACCCCAGCGCCTCAGGGG CCTGACCCCCATCCTGACCCTG GGCCGCACTGCCTCCTGACCCTG GGCCGAGCCCAGACCTCCTGACCCTG GGCGGAGGCCCGGAGCCCCCCCAATTC AGCAGCTTTCCAGTACCTCCT GTGGAACTTTCCCCCCCCCC	49	Involved in erythroid development	<pre>Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. &amp; Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA- 1. PNAS 93, 12355-12358 (1996).</pre>
GATA2	ATGGAGGTGGCGCCCGAGCAGCCG CGCTGGATGGCGCCCGGCCGTG CTGAATGCGCAGCACCCCGGCCGTG CACACCCGGGCCTGGCGCACAAC TACATGGAACCCCGGCAGCTGCTG CCTCCAGACGAGGTGGACGTCTTCT TCAATGACCTCGACTCGCAGGGCA ACCCCTACTATGCCAACCCCGCTGA CGCCGGGCGCGCGCCTCCTCCTACAG CCCCGCGCACGCCGCCTGACCGG AGGCCAGATGTGCCGCCCACACTT GTTGCACAGCCCGGGTTTGCCCTGG CTGGACGGGGGCCAAGCCCCCCTC TCTGCCCGCGCCGCG	50	Involved in haematopoetic development	<pre>Pimanda, J. E. et al. Gata2, Fli1, and Scl form a recursively wired gene- regulatory circuit during early hematopoietic development. Proc. Natl. Acad. Sci. U. S. A. 104, 17692-7 (2007). Lugus, J. J. et al. GATA2 functions at multiple steps in hemangioblast development</pre>

TABLE 1-continued

TABLE 1-continued

		SEQ ID			
GENE	SEQUENCE	NO:	ROLE	REFERENCES	
	ACGGGGGCTGCGTCCAGCCTCAT CTTCCGCGGGGGGGTAGTGCAGCCC GAGGAGAGGGCAAGGACGGCGTCA AGTACCAGGTGTCACTGACGGGGA GCATGAAGATGGAAAGTGGCAGTC CCCTGCGCCCAGGCTAGCTACTAT GGGCACCCAGCCTGCTACCACCA CCCCATCCCCAGCCTGCTACCACCA CCCCATCCCCAGCTTCCACCCCCGA GGCTTCCTGGGGGGGACCGGCCTCC AGCTTCCTGGGGGGGACCGGCCACC AGCTCCTGGGGGGGACCGGCCACC GGGAGTGGTCAACTGTGGGGCCA CAGCCACCCCTCTCTGGCGGGGGG ACGGCACCGGCCTCTCTGGCGGCGGG ACGCCCCGGCCACTACCTGTGCA ATGCCTGTGGCCCTCACCACAGAC CAGCCCCAGCACACCGCCACT CCCCAGAAGAGCCGCCCCCACG CGCCAGAAGAGCCGCCCCCACTCAT GAATGGGCAGAACCGACCACCACTCAT CAAGCCCCAGCGCAGCACCACCACTCAT GAAATGGCCAGCCGCCGCACCCCTGT GGCCTCTACTACAAGCTGCACACCAC CACCTTATGGCGCCGAAACGCCCAC CGCCAGAAGAGCCGCCACTGCTG TGCAAATTGTCAGACGACAACCAC CACCTTACTACAAGCTGCACAACG CGGCAGCAGCCCCCCCCCC			and differentiation. Development 134,393-405 (2007).	
GATA4	ATGTACCAGAGCCTGGCTATGGCTG CTAATCATGGACCTGGCTGCCCCTGGAGC CTATGAAGCCGGAGGAGCCTGGCGCCGCT TTTTATGCATGGAGCGCGCGCT TCTTCTCCCGTGTATGTGGCGCACC CTAGAGTGCCCAGCAGCGGGGGGGG GCCTTTCTTATCTTCAGGGGAGGGG	51	Involved in cardiovascular development	Xin, M. et al. A threshold of GATA4 and GATA6 expression is required for cardiovascular development. Proc. Natl. Acad. Sci. U. S. A. 103, 11189-94 (2006). Rivera- Feliciano, J. et al. Development of heart valves requires Gata4 expression in endothelial- derived cells. Development 133, 3607-18 (2006).	

TABLE 1-continued

	CRO TR		
SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
CTGTGGACTGTACATGAAGCTGCAC GGCGTGCCCAGACCTCTGGCCATG AGAAAGGAGGGCATCCAGACCAGA			
ACCGCC ATGGCCCTGACCGACGCGGATGG TGTCTCCCTAAAAGATTCGGCGCCG CTGGCGCTGATGCTTCTACAGCAG AGCCTTCCCCGCTAGGAAACCCAG CTCAAGCTCTAGCTGTAGCAGAG CTCAAGCTCTAGCTGTAGCAGAG CTCAAGCTCTAGCTGTAGCAGAG CTCAAGCTCTAGCTGTAGCAGAG CGGAGAGAGAGGCGCCCCCCCGGCGGCCA CCAGCCAG	52	Involved in cardiac, lung, endoderm and extraembryonic development	<pre>Xin, M. et al. A threshold of GATA4 and GATA6 expression is required for cardiovascular development. Proc. Natl. Acad. Sci. U. S. A. 103, 11189-94 (2006). Morrisey, E. E. et al. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes Dev. 12, 3579- 3590 (1998). Koutsourakis, M.; Langeveld, A.; Patient, R.; Beddington, R.; Grosveld, F. The transcription factor GATA6 is essential for early extraembryonic development. Development 126, 723-732 (1999). Zhang, Y. et al. A Gata6- Wnt pathway required for epithelial stem cell development</pre>
	SEQUENCE CTGTGGACTGTACATGAAGCTGCAC GGCGTGCCCAGACCTCAGACCAGA AAGAGGAAGGCCCAGAACCTCAGAC CAGAGCAGAG	SEQ ID SEQUENCE NO: SEQUENCE NO: SEQUENCE NO: CTGTGGACCTGACACCTGACCGAC GGCGTGCCCAGACCTCTGGCCTCAG CGCTGGGCCGAGACCTCAGACCAGA AAGAGAAAGGCGCCAGCAGCACTGAC AAGAGCAGACGCCCCCCGCTGCTCCAG CCTCTGGAGCGCAGCACCACTGCTGC TGAAGGCCCAGCACACTTCTTCTGA GGAAGTGAGGCCCACACACTCTGCT CTAAGGCCACCACACTTCTTCTGA GGCAGCCTGAGCAGCACTACGG CCACGCCTGAGCAGCACTACGG CCCAGGCTTGAGCAGCACTACAGG CCCAGGCCTGAGAGTGGCGCCAC AGGCCTCAGACTCCAGCACGACGA GGCCACACCCCGGAGATTCCACC ACGCC ATGGCCCTGAAGCTGCAGCGCGATGG TGTCTCCCTGAAAGCTCGGCGCGG TGTCTCCCTGAAGGCGCGAGAG CCCACACCACCAGCCCCAGCAGAG CCCACCCCCCCCCC	SEQUENCE NO: ROLE CTOTOGACTOTACATGAAGCTOCAC GOOTGCCCAGACCTOTGCCACG AGAAGGAGGCCACCAGCAATAGCT CTGGAAGGCAGGCAATCACT CTGGAAGGCAGCACCTOCTOCTOC CTGAAGGCACACCACTOTTCTTOTGG GCCAGGCCTGAGCAGCCATAGCG CCCAGCCTGAGCAGCCATCACG GCCAGGCCTGGACGTGGGCCAC CCACAGCTTGAGCAGCCATCACG GCCAGGCCTGGAGTGCGGCGGC CACGGCCTGGAGTGCGCGGCGG CGCGGCGCGGACGCGGCGGGCG TGGGCGGCGGCGGCGGGCGGCG TGTCTACCGGCGGCGGCGGCG CGCCGCGCCGCCGCGGCGCG CGCGGCGCGCGGCGGCGGCG CGCGGCGCGCGGCGGCGGCG GGGGCGGACGCCGCGGCGGCG GGGGCGGACGCCGCGGGCGG GGGGCGGACGCCGCGGGGGGGG GGGGCGGACGCCGGCGGGCG GGGGCGGACGCGGCGGGGGGG GGGGCGGACGCGGCGGGCGG GGGGCGGACGCGGCGGGGGGG GGGGCGGACGCGGCGGGGGGG GGGGCGGACGCGGCGGGGGGGG GGGGCGGACGCGGCGGGGGGG GGGGCGGACGCGGCGGGCGG GGGGCGGACGTGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGACCTGG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGACCTGGGGG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGAACCTG GGGGCGGACCTGGGGGAACCTG GGGGCGGACCTGGGGGGAACCTG GGGGCGGACCTGGGGGAACCTG GGGGCGGACCTGGGGGAACCTG GGGGCGGACCGCGGGGGGAACCTG GGGGCGGACGCGGCGGGGGGA AGGGGCGGACGCGGGGGGAGGGGA

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TGAGCTGCGCCAACTGTCATACCAC AACAACCACACTGTGGCGGAGAAA CGCCGAGGCGAGCCCGTGTGTAA CGCCTGCGGCCTTTACATGAAGCTG CACGGCGGCCCTTACATGAAGCC ATGAAGAGAAGA			Nat. Genet. 40, 862-870 (2008).
3LI1	ATGTTCAACTCGATGACCCCACCAC CAATCAGTAGCTATGGCGAGCCT GCTGTCTCCGGCCCCCCCAGTCA GGGGGCCCCCAGTGTGGGGACAGA AGGACTGTCTGGCCGCCCTCTGC CACCAAGCTAACCTCATGTCCGGCC CACCAAGTTATGGGCCAGCCAGAG AGACCAACAGCTGCACCGAGGGC AGTCAACTGCCACCGAGAGCCG CACTGTCCATCTCCCCGGAGTGC AGTCAAGTTGACCAAGAAGCGGCC ACTGTCCATCTCACCTCGTGCGGAT GCCAGCCTGGACCTGCAGCCCCCCG TAGCTTCATCAACTCGCGATGCCA ATCTCCCAGGAGGCCCCCCCGGTGCGAT GCCACCTGGCCCCCAGCCCAG	53	Involved in neural stem cell proliferation and neural tube development	Lee, J. et al. Gli1 is a target of Sonic hedgehog that induces ventral neural tube development. Development 124, 2537- 2552 (1997). Palma, V. et al. Sonic hedgehog controls stem cell behavior in the postnatal and adult brain. Development 132, 335-44 (2005).

'ABLE 1-continued	ABLE	1-continued
-------------------	------	-------------

		SEO ID	SEQ ID		
ENE	SEQUENCE	NO:	ROLE	REFERENCES	
	CAAGCCCTGGGGCCCAGTCATCCTG				
	CAGCAGTGACCACTCCCCGGCAGG				
	GAGTGCAGCCAATACAGACAGTGG				
	TGTGGAAATGACTGGCAATGCAGG				
	CTTGGACGAGGGACCTTGCATTGCT				
	GGCACTGGTCTGTCCACTCTTCGCC				
	GCCTTGAGAACCTCAGGCTGGACC				
	AGCTACATCAACTCCGGCCAATAG				
	GGACCCGGGGTCTCAAACTGCCCA				
	GCTTGTCCCACACCGGTACCACTGT				
	GTCCCGCCGCGTGGGCCCCCCAGTC				
	CCAGCAGCATCAGCTCTGCCTATAC				
	TGTCAGCCGCCGCTCCTCCCTGGCC				
	TCTCCTTTCCCCCCTGGCTCCCCAC				
	CAGAGAATGGAGCATCCTCCCTGC				
	CTGGCCTTATGCCTGCCCAGCACTA				
	CCTGCTTCGGGCAAGATATGCTTCA				
	ATAGGTGGTCTTCCCATGCCTCCTT				
	GGAGAAGCCGAGCCGAGTATCCAG				
	GATACAACCCCAATGCAGGGGTCA				
	CCCGGAGGGCCAGTGACCCAGCCC				
	AGGCTGCTGACCGTCCTGCTCCAGC				
	TAGAGTCCAGAGGTTCAAGAGCCT				
	GGGCTGTGTCCATACCCCACCACT				
	GIGGCAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG				
	ACTCACCACAGCCCCCCAGCATCA				
	CTGAGAATGCTGCCATGGATGCTA				
	GAGGGCTACAGGAAGAGCCAGAAG				
	TTGGGACCTCCATGGTGGGCAGTG				
	GTCTGAACCCCTATATGGACTTCCC				
	ACCTACTGATACTCTGGGATATGGG				
	GGACCIGAAGGGGCAGCAGCIGAG CCTTATGGAGCCGAGCGGCTCCAGCC				
	TCTCTGCCTCTTGGGCCTGGTCCAC				
	CCACCAACTATGGCCCCAACCCCTG				
	TCCCCAGCAGGCCTCATATCCTGAC				
	CCCACCCAAGAAACATGGGGTGAG				
	TTCCCTTCCCACTCTGGGCTGTACC				
	ACATTATCCACACTCCACTTCA				
	GCCAGAACAGGGGTGCCCAGTGGG				
	GTCTGACTCCACAGGACTGGCACCC				
	TGCCTCAATGCCCACCCCAGTGAGG				
	GGCCCCCACATCCACAGCCTCTCTT				
	TTCCCATTACCCCCAGCCCTCTCCT				
	TCCTTCAGAACCCAGCCTGATTATCT				
	GACTTTGATTCCCCCACCCATTCCA				
	CAGGGCAGCTCAAGGCTCAGCTTG				
	TGTGTAATTATGTTCAATCTCAACA				
	GGAGCTACTGTGGGAGGGTGGGGG				
	CAGGGAAGATGCCCCCGCCCAGGA				
	ACCTTCCTACCAGAGTCCCAAGTTT				
	ACATATGGACCTGGCTTTGGACCCA				
	ACTTGCCCAATCACAAGTCAGGTTC				
	CTATCCCACCCTTCACCATGCCAT				
	GAAAATTTTGTAGTGGGGGCAAAT				
	AGGGCTTCACATAGGGCAGCAGCA				
	CCACCTCGACTTCTGCCCCCATTGC				
	TCATCCTGAGGTGGGCAGGCTAGG				
	AGGGGGTCCTGCCTTGTACCCTCCT				
	CCCGAAGGACAGGTATGTAACCCC				

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CTCAGCTGGACTTTGTGGCTATTCT GGATGAGCCCCAGGGGCTGAGTCC TCCTCCTTCCCATGATCAGCGGGGC AGCTCTGGACATACCCCACCTCCCT CTGGGCCCCCCAACATGGCTGTGG GCAACATGAGTGTCTTACTGAGATC CCTACCTGGGGAAACAGAATTCCTC AACTCTAGTGCC			
IAND 2	ATGAGTCTGGTAGGTGGTTTTCCCC ACCACCCGGTGGTGCACCACGAGG GCTACCCGTTTGCCGCCGCCGCCGC CGCCAGCCGCTGCAGCCATGAGGA GAACCCCTACTTCCATGGCTGGCCC CCGACTACAGCATGGCCCTGTCCTA CAGCCCCGAGTATGCCAGCGGCAG GACTCAGAGCATCACACGGCGCAG GACTCAGAGCATCACACGGCGCAG GACTCAGAGCACGGAGTGCATCCC CAACGGCGCGGAGGGCGCCGCGCAG GCCACCGCCAGCGACGCCCAACT CTCCAAAATCAAGACCCTGCGCCTG GCCACCAGCTACACCAAGACACCCA TGGACCTGCGCGGAGGCGCAGC AGAATGGCGAGGCGA	54	Involved in cardiac development	Srivastava, D. et al. Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. Nat. Genet. 16, 154-160 (1997).
ΗF1Α	ATGGTTTCTAAACTGAGCCAGCTGC AGACGGAGCTCCTGGCGGCCCTGC TGGAGTCAGGGCTGAGCAAAGAGG CACTGCTCCAGGCACGGCAGGGGGG CGGGGCCCTACCTCCTGGCTGGAG GGGGCCCTCCTGGACGAGGGGGGG TGGCTGAGCTGCGCAATGGGGGAGC TGGCTGAGCTGCGGGCCCGAGGGAGC GGCGACCACCGGCGCCGTGGGGGAGGC GGCCACCAGAAGATGGCGAGGAC GACCACCAGAAGATGGCGAGGC GGCCCACCAGAAGATGGTCAAGTC CTACCTGCAGCAGCAGCGTGGTGGA GACCCCCCTCAGCAGAGGCCGTG GCGTGTGGCGAAGATGGTCAAGTC CTACCTGCAGCAGCAGCCCGTG GCGTGTGGCGAAGATGGTCAAGTC CTACCTGCAGCAGCACCACCTGTCC CAACCGGGAGGTGGTCCATACCAC TGGCCTCAACCAGGCGCGCCCC CTGTACCCCGCAGAAGAGCGCGCC CCGTGTAGCCAGAAGAGCGCGCC CCGTGTAGCCAGAAGAGCGGCCGCC CTGTACCCTGAACAAGGCGCACCCC ATGAAGACGCCAGAAGAGCGGCCGCC CTGTACCCTGAACAAGGGCGCGCC CCGTGCAAGCAGGAAGGCGGAGGA ACCCGTTCAACCAGGCAGGCCGCC CTGTACCCAGCAAGAGGGCCAGCA TGCAAGCGCAAGAAGGGCCGAGCA CCCATCCAAGAAGGGCCAGCATCA CCCATCCAAGAAGGGCCAGCATC CCCAGCAGCACAAGAGGGCCAGCA GAGGAGCAGAAGAACCCTAGCAG GAGGGCAGAAGAACCCTAGCAG GAGGAGCGAGAGAACCCTAGCAG GAGGAGCAGAGAGACGCTAGTGAG GAGGCCAACGAGGCGGCCAGCATC CCAGGGGCAGAGGACGCTAGTGAG GAGGCCAAGGGCGGGCCAGCATC CCAGGGGCAGAGGACGCTAGTGAG GAGGCCAAGGGCGGGCCAGCATC CCAGGGGCAGAGGACGCTAGTGAG GAGGCCAAGGGCGGGCCAGCATC CCAGGGGCAGAGGACCTAGCAG GAGGCCAAGGGCGGGCCCACCACCT CCAGGAGGCGGGGCCCACCACCACCAC GCCCACGGCCACAGGCGCAAGAACCCTACCC CAGGGCCAACGCGGCCCACCACCAC GCCCACCGCCACCACGCCCACCACCT GTCCCAACCACGCCGCCACACCT CCAGGAGGCGGGCCCCCCC CAGGCCCACCGCCCCCC CAGGCCCACCGCCCCCCC CAGGCCCACGCCCCCCC CAGGCCCACGGCCCCCCCC	55	Involved in liver, kidney, pancreatic and gut development	D' Angelo, A. et al. Hepatocyte nuclear factor lalpha and beta control terminal differentiation and cell fate commitment in the gut epithelium. Development 137,1573-82 (2010). Servitj a, JM et al. Hnf1 alpha (MODY3) controls tissue-specific transcriptional programs and exerts opposed effects on cell growth in pancreatic islets and liver. Mol. Cell. Biol. 29, 2945-59 (2009). Si-Tayeb, K.; Lemaigre, F. P.; Duncan, S. A.

of the Liver.

CAGCCTGCGACCAGTGAGACTGCA

TABLE 1-continued

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GAAGTACCCTCAAGCAGCGGCGGT CCCTTAGTGACAGTGTCTACACCCC TCCACCAAGTGTCCCCCAGGGCCT GGAGCCCAGCCACGCCTGCCCAGGGCCT GGAGCCCAGCCACGCCCCCCCTGTC AGCTGGGGGCCACGCACTGCCACAGC TTGGAGCAGACACCCCCAGGCCTC AGCACCCTGACAGCCTCCCCCGGCCAGGCCTC CCTGGGTCCTACCTCCTGGGGCCTGCCCC CCTGGGTCCTACGTTCACCAACACA GGTGCCTCCACGCTGGTGAGCCTGCCC CCTGGGTCCTACGTCCCCAGCAGGCG GGCCTCCACGCAGGCACAGAGTG TGCCGGTCATCACCTGGCCGCCCC CCGGGTCCTCCCCCGGCGCGCGCGCCCCCCCC			Dev. Cell 18, 175-189 (2010). Martovetsky, G., Tee, J. B. & Nigam, S. K. Hepatocyte nuclear factors 4α and 1α regulate kidney developmenta 1 expression of drug- metabolizing enzymes and drug transporters. Mol. Pharmacol. 84,808-23 (2013).
HNF1B	ATGGTTAGCAAACTGACATCCCTCC AGCAGGAACTTCTTTCTGCCCTCCT CTCCAGTGGGGTAACCAAAGAGGT ACTGGTCCAGGCTTTGGAGGAGTTG CTCCCCTCACCGAGTTTGGTGTAA AGTTGGAGACTCTCCCCCTCTCCCC TGGTTCTGGAGCAGAGCGGGGTCGCTT TCAGCGACGAAGGGCCGGATAC TAAACCGGTAATTTCATACGCTTACA AACGGACACGCAAGGGCCGGCGCC ATCCTCAAAGAACTGCAGGGCCCCTA ATTGCAGAGAAGGCGCGGAGCACC GAGCTGAAGATGACGCGGGGGACCACC GAGCTGAAGTGACGACGGCGGAGCACC GAGCTGAAGTTGACGAGAAGCTGCC AACATAACAT	56	Involved in liver, kidney, pancreatic and gut development	<pre>D' Angelo, A. et al. Hepatocyte nuclear factor lalpha and beta control terminal differentiation and cell fate commitment in the gut epithelium. Development 137,1573-82 (2010). Si-Tayeb, K.; Lemaigre, F. P.; Duncan, S. A. Organogenesis and Development of the Liver. Dev. Cell 18, 175-189 (2010). Clissold, R. L., Hamilton, A. J., Hattersley, A. T., Ellard, S. &amp; Bingham, C. HNFIB- associated renal and evtrarenol</pre>

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GGTGAGGGTATATAATTGCTTTCCC			disease-an
	AACAGGCGGAAGGAGGAAGCATTC			expanding
	CGGCAAAAGCTGGCGATGGATGCC			clinical
	TACTCAAGCAACCAGACACATAGC			spectrum.
	CTCAACCCTCTGTTGTCACACGGGT			Nat. Rev.
	CCCCTCATCACCAACCTTCTTCCTC			Nephrol. 11,
	TCCACCCAACAAACTTTCTGGTGTC			102-112
	CGATATTCCCAGCAGGGGAACAAC			(2014).
	GAGATAACATCTTCCTCTACTATAA			De Vas, M.
	GTCATCACGGAAATTCTGCAATGGT			G. et al.
	AACGTCACAGAGTGTGTTGCAACA			Hnf1b
	GGTATCACCCGCGTCTCTTGATCCA			controls
	GGCCACAATCTGTTGAGCCCTGACG			pancreas
	GAAGATGATCTCTGTTTCTGGTGG			morphogenesis
				and the
	ACCAACATACATAGTCTCAGTCATC			generation of
				NgH3+
				progonitorg
	ACCECTALIGCGCAALCIIIGAAC			progenitors.
	TCATAAGCACAAICIGIACCCG			142 971-92
	GGCGGCGCTCCAACCAGCTCAII			(2015)
	CTCCCAGCAGCTCCATTCACCCCAT			El-Khairi P
	CAACAGCCTCTGATGCAGCAGAGC			& Vallier L
	CCTGGTAGTCACATGGCTCAACAGC			The role of
	CGTTCATGGCAGCTGTCACTCAGCT			hepatocvte
	CCAGAACTCCCATATGTATGCCCAC			nuclear factor
	AAGCAAGAACCACCACAATACAGT			1β in disease
	CACACATCAAGATTCCCCAGTGCTA			and
	TGGTTGTTACTGACACATCCTCTAT			development.
	CTCAACTCTGACGAACATGTCCAGT			Diabetes,
	AGTAAACAATGTCCTCTGCAAGCAT			Obes. Metab.
	GG			18,23-32
				(2016).
HNF4A	ATGCGACTCTCCAAAACCCTCGTCG	57	Involved in	Si-Tayeb, K.;
	ACATGGACATGGCCGACTACAGTG		liver, kidney,	Lemaigre, F.
	CTGCACTGGACCCAGCCTACACCAC		pancreatic and	P.; Duncan, S.
	CCTGGAATTTGAGAATGTGCAGGT		gut	Α.
	GTTGACGATGGGCAATGACACGTC		development	Organogenesis
	CCCATCAGAAGGCACCAACCTCAA			and
	CGCGCCCAACAGCCTGGGTGTCAG			Development
	CGCCCTGTGTGCCATCTGCGGGGAC			of the Liver.
	CGGGCCACGGGCAAACACTACGGT			Dev. Cell 18,
	GCCTCGAGCTGTGACGGCTGCAAG			175-189
	GGCTTCTTCCGGAGGAGCGTGCGG			(2010).
	AAGAACCACATGTACTCCTGCAGA			Martovetsky,
	TTTAGCCGGCAGTGCGTGGTGGAC			G., Tee, J. B.
				& Nigam, S.
				K. Hepatocyte
				nuclear
				$1\alpha$ regulate
	ACACCUCAAGGICAAGCTATGAGG			id regulate
	CARCAGE IGEE IGEE ICAICAIGC			developments
	CGACAGATCACCTCCCCCCCCCCCCCCCCC			1 expression
	GGATCAACGCGACATTCCCCCC			of drug-
	AGAAGATTGCCACCATCCCACATC			metabolizing
	TGTGTGAGTCCATGAAGGAGCAGC			enzymes and
	TGCTGGTTCTCGTTGAGTGGGCCAA			drug
	GTACATCCCAGCTTTCTGCGAGCTC			transporters
	CCCCTGGACGACCAGGTGGCCCTG			Mol.
	CTCAGAGCCCATGCTGGCGAGCAC			Pharmacol.
	CTGCTGCTCGGAGCCACCAAGAGA			84,808-23
	TCCATGGTGTTCAAGGACGTGCTGC			(2013).
	TCCTAGGCAATGACTACATTGTCCC			Maestro, M.
	TCGGCACTGCCCGGAGCTGGCGGA			A. et al
	GATGAGCCGGGTGTCCATACGCAT			Distinct roles
	CCTTGACGAGCTGGTGCTGCCCTTC			of HNF1b eta,
	CAGGAGCTGCAGATCGATGACAAT			HNF1alpha,
	GAGTATGCCTACCTCAAAGCCATCA			and
	TCTTCTTTGACCCAGATGCCAAGGG			HNF4alpha in
	TCTTCTTTGACCCAGATGCCAAGGG GCTGAGCGATCCAGGGAAGATCAA			HNF4alpha in requlating

TABLE 1-continued

		SEQ ID		
GENE	SEQUENCE	NO :	ROLE	REFERENCES
	GAGCTTGGAGGACTACATCAACGA CCGCCAGTATGACTCGCGTGGCCGC TTTGGAGAGCTGCTGCTGCTGC CCACCTTGCAGAGACATCACCTGGCA GATGATCGAGCAGATCCAGTTCATC AAGCTCTTCGGCATGGCCAAGATTG ACAACCTGTTGCAGGAGATGCTGCT GGGAGGTCCGTGCCAGGAAGCCCAGGA GGGCGGGGTTGGAGTGGAG			development, beta-cell function and growth. Endocr. Dev. 12,33-45 (2007). Garrison, W. D. et al. Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon. Gastroenterol ogy 130, 1207-20 (2006).
HOXA1	ATGGACAACGCGCGGATGAATTCC TTCCTCGAGTACCCAATTTTGTCTA GTGGAGACAGTGGCACTTGCAGTG CCCGAGCCTATCCATCAGACACA GAATTACAACATTCCAAAGCTGTGC GATGTCAGCCAACAGTTGCGGCG AGACGACCGCTTCCTGGTCGGAAG AGGGGTCAACTCGCGCGCG CATCACCATCACCACCACCACCACCAC CACCCCCAACGCGCGCACTTACCAA CCAGCGGCAATTGGGCGTGAGC ATGGCCATCCTCATGGGCGTGAAC CAGCGGCAATTCGGCGCGACTTACCGAC CCTTATAGCCATCACCACACCA	58	Involved in neural and cardiovascular development	Tischfield, M. A. et al. Homozygous HOXA1 mutations disrupt human brainstem, inner ear, cardiovascular and cognitive development. Nat. Genet. 37, 1035- 1037 (2005).
HOXA10	ATGTGTCAAGGCAATTCCAAAGGT GAAAACGCAGCCAACTGGCTCACG GCAAAGAGTGGTCGGAAGAAGCGC TGCCCCTACAACGAAGCACCAGACA CTGGAGCTGGAGAAGGAGCTTCTG TTCAATATGTACCTTACTCGAGAGC GGCGCCTAGAGATTAGCCGCAGCG TCCACCTCACGGACAGACAAGTGA AAATCTGGTTTCAGAACCGCAGGA	59	Involved function in fertility, embryo viability, and regulation of hematopoetic lineage commitment	Buske, C. et al. Overexpression of HOXA10 perturbs human lympho- myelopoiesis in vitro and in

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TGAAACTGAAGAAAATGAATCGAG AAAACCGGATCCGGGAGCTCACAG CCAACTTTAATTTTTCC			vivo. Blood 97, 2286- 2292 (2001). Satokata, I., Benson, G. & Maas, R. Sexually dimorphic sterility phenotypes in Hoxalo- deficient mice. Nature 374, 460-463 (1995).
HOXA11	ATGGATTTTGATGAGCGTGGTCCT GCTCCTCTAACATGTATTGCCAAG TTGTACTTACTACGTCTCGGGTCCA GATTTCTCCAGCCTCCTTCTTTCT GCCCCAGACCCGTCCTTCCTCCCACC TGCCCCAGTCCAACCCGTGCGCG AAGTGACCTTCAGAGAGTACGCCA TTGAGCCGGCATCAGACCGTGCCCCCCGCGGCAACTCAGTCGCCACGAC GCCGGGGAGCAGCCGGCGCCCAGCGG GCCGGCGTGCCTGCCACGCCAC	60	Involved in kidney development	Patterson, L. T., Pembaur, M. & Potter, S. S. Hoxall and Hoxdl1 regulate branching morphogenesis of the ureteric bud in the developing kidney. Development 2153-2161 (2001).
HOXB6	ATGAGTTCCTATTTCGTGAACTCCA CCTTCCCCGTCACTCTGGCCAGCG GCAGGAGTCCTTCCTGGGCCAGCA CCGCTCTATTCGTCGGGCTATGCGG ACCCGCTGAGACATTACCCCGCGGC CTACGGGCCAGGGCCGGCCAGGA CAAGGCTTTGCCACTTCCTCCTAT TACCCGCCGGCGGCGGCGGCGGCTAC GGCCGAGCGGCGGCGCCTGCGACTAC GGCCGAGCGGCGGCGCCTGCGCACTC TCCGGCCGCGACGAGCCGCGCACTC TCCGGCCCGAGCGCGCGCGCCCCG GACTGCGCCGAGGCGGAGACCCG GACTGCGCCGAGCAGACAGAGCCGG TTCCGCCCCGGCCAGCACAGAGCGGA TGCCCCCCCCGTCTACCCGTGGA TGCCACCCGGATGAATTCGTGCAACA GTTCCTCCTTTTGGCCCAGCGCG	61	Involved in lung and epidermal development	<ol> <li>Patterson,</li> <li>T.,</li> <li>Pembaur, M.</li> <li>Potter, S. S.</li> <li>Hoxall and</li> <li>Hoxdl1</li> <li>regulate</li> <li>branching</li> <li>morphogenesis</li> <li>of the</li> <li>ureteric bud</li> <li>in the</li> <li>developing</li> <li>kidney.</li> <li>Development</li> <li>2153-2161</li> <li>(2001).</li> <li>Komuves, L.</li> </ol>

TABLE 1-continued

TABLE 1-continued

		SEQ ID		
GENE	SEQUENCE	NO :	ROLE	REFERENCES
	GCGAGGCCGCCAGACATACACACG TTACCAGACGCTGGAGCTGGAGAA GGAGTTTCACTACAATCGCTGCCTG ACGCGGCGCGCGCGCGCACCGAGATC GCGCACGCCCTGTGCCTGACCGAG AGGCAGATCAAGATATGGTTCCAG AACCGACGCATGAAGTGGAAAAAG GAGAGCAAACTGCTCAGCGCGCTCT CAGCTCAGTGCCGAGGAGGAGGAA GAAAAACAGGCCGAG			<pre>G. et al. Changes in HOXB6 homeodomain protein structure and localization during human epidermal development and differentiation. Dev. Dyn. 218, 636-647 (2000). Cardoso, W. V., Mitsialis, S. A., Brody, J. S. &amp; Williams, M. C. Retinoic acid alters the expression of pattern- related genes in the developing rat lung. Dev. Dyn. 207, 47- 59 (1996).</pre>
KLF4	ATGGCTGTCAGCGACGCGCTGCTCC CATCTTTCTCCACGTTCGCGTCTGG CCCGGCGGGAAGGAGAGAAGACACT GCGTCAAGCAGGTGCCCGAATAA CCGCTGGCGGGGGGGGCCCCGATAA CCGCTGGCGGGGGGGCACCAGGCCTGG CGGCGACCGTGGCCACAGACCTGG AGAGCGGCGGCAGCCCGAGGCCCC TACCTCGGAGAGAGACCTGGCGCCCC TACCTCGGAGTAGCACCTGGCGCCCC GTGTCTCCTCGAGTGGCGCCCCCG GTGTCCTCGTCAGCGGCCCCCG CCTCGGAGTCAGTGCCGCCCCCG TCTCCGGAGTCAGTGCCGCCCCCG GTGTCCTCGTCGCGCGCCCCCG GGGCACCGGCGGCGCGCCCCCCCCCC	62	Involved in regulation of pluripotency and development of skin. Reprogramming factor for induction of pluripotency.	Fuchs, E., Segre, J. A. & Bauer, C. Klf4 is a transcription factor required for establishing the barrier function of the skin. Nat. Genet. 22, 356-400 (1999). Jiang, J. et al. A core Klf circuitry regulates self- renewal of embryonic stem cells. Nat. Cell Biol. 10, 353- 360 (2008). Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-76 (2006). Takahashi, K. et al. Induction of pluripotent

TABLE 1-continued

GENE	SEQUENCE	NO:	ROLE	REFERENCES
	CCCGCCGCTCCATTACCAAGAGCTC ATGCCACCCGGTTCCTGCATGCCAG AGGAGCCCAAGCCAAG			from adult human fibroblasts by defined factors. Cell 131, 861-72 (2007). Yu, J. et al. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science (80). 318, 1917-1920 (2007).
THX3	ATGGAGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	63	Involved in pituitary gland development	(2007). Sheng, H. Z. et al. Multistep Control of Pituitary Organogenesis. Science (80). 278, 1809-1812 (1997).

	TABLE		Shermaea		
GENE	SEQUENCE	SEQ : NO:	ID ROLE	REFERENCES	
LMX1A	ATGGAAGGAATCATGAACCCCTAC ACGGCTCTGCCACCCCACAGCAG CTCCTGGCCATCGAGCAGGAGTGTT ACAGCTCAGATCCCTTCCGACAGG GTCTCACCCCACCC	64	Involved in neuronal development	Lin, W. et al. Foxal and Foxa2 function both upstream of and cooperatively with Lmx1a and Lmx1b in a feedforward loop promoting meso- diencephalic dopaminergic neuron development. Dev. Biol. 333, 386-396 (2009). Qiaolin, D. et al. Specific and integrated roles of Lmx1a, Lmx1b and Phox2a in ventral midbrain development. Development 138, 3399- 3408 (2011).	
4EF2C	ATGGGGAGAAAAAAGATTCAGATT ACGAGGATTATGGATGAACGTAAC AGACAGGTGACATTTACAAAGAG AAATTTGGGTTGATGAAGAAGGCT TATGAGCTGAGC	65	Involved in cardiac development	Lin, Q. et al. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. Science 276, 1404-7 (1997).	

CAGTGGTTTCCGTAGCAACTCCTAC TTTACCAGGACAAGGAATGGGAGG ATATCCATCAGCCATTTCAACAACA

TABLE 1-continued

BLE	1 -	cont	inue	F
1 D L L L		COILC	TITUEL	л

	TABL	1-continued			
GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES	
	TATGGTACCGAGTACTCTCTGAGTA GTGCAGACCTGTCATCTCTGTCTGG GTTTAACACCGCCAGCGCTCTCAC CTTGGTTCAGTAACTGGCTGGCAAC AGCAACACCTACATAACATGCCAC CATCTGCCCTCAGTCAGTTGGAAGC TTGCACTAGCACTCATTATCTCAG AGTTCAAATCTCTCCCTGCCTTCTA CTCAAAGCCTCAACATCAAGTCAG AACCTGTTTCTCCTCCTAGAGACCG TACCACCGCCCCCACGAGGCGGGG AGATCTCCTGTTGACAGCTTGAGCA GCTGTAGCAGTTCGTACGACGGGA GCGACCAGAGGGATCACCGGAACG AATTCCACTCCCCCATTGGACTCAC CAGACCTTCGCCCGACGAACGGA AAGTCCCTCAGTCAAGCGCAACG AAGTCCCTCAGTCAAGCGCATGCG AAGTCCCTCAGTCAAGCGCATGCG ACTTTCTGAAGGATGGCAACA				
MESP1	ATGGCCCAGCCCTGTGCCCGCCGC GCCTGGGCCCAACTCGGCGGCC GCCGCCTCCGACAACTCGGCGGCC GCCGCCCTCGACAAGGACTGCGG CCGCCCCTCGGCACAAGGACGCGG CCAGGCACCCTCGGGACCCCGC GCCCCTCGTAGGTAGGCGGGGC GCCCCTCCGTAGGTAGGCGGGCG GCCCCCTCCGTAGGTAGGCGGGCG GGCCAGAGCAGACGCCGGCAGGC GGCCAGAGGCAGACGCCAGTGAG GGCCAGAGCAGACGCCCGGCGGC GCCCCCCCCCC	66	Involved in cardiac development	Bondue, A. et al. Mesp1 Acts as a Master Regulator of Multipotent Cardiovascular Progenitor Specification. Cell Stem Cell 3,69-84 (2008).	
MITF	ATGCTGGAAATGCTAGAATATAAT CACTATCAGGTGCAGACCCACCTCG AAAACCCCACCAACGACACCACATAC AGCAAGCCCAACGGCAGCAGGTAA AGCAGTACCTTTCTACCACTTTAGC AAATAAACATGCCAACCAAGTCCT GAGCTTGCCATGTCCAAACCAAGTCCT GGCGATCATGTCATG	67	Involved in pigment cell and melanocyte differentiation	Widlund, H. R. & Fisher, D. E. Microphthala mia- associated transcription factor: a critical regulator of pigment cell development and survival. Oncogene 22, 3035-3041 (2003).	

ABLE	1-	cont	inued	
------	----	------	-------	--

	TABL	E 1-con	1-continued			
GENE	SEOUENCE	SEQ ID NO:	ROLE	REFERENCES		
	Carcererarcaaaaccaaccercerce					
	CCCCACCAGGCCTCACCATCAGCA					
	ACTCCTGTCCAGCCAACCTTCCCAA					
	CATAAAAAGGAGCTCACAGAGTC					
	TGAAGCAAGAGCACTGGCCAAAGA					
	GAGGCAGAAAAAGGACAATCACAA					
	CCTGATTGAACGAAGAAGAAGATT					
	TAACATAAATGACCGCATTAAAGA					
	ACTAGGTACTTTGATTCCCAAGTCA					
	AATGATCCAGACATGCGCTGGAAC					
	AAGGAACCATCTTAAAAGCATCC					
	GTGGACTATATCCGAAAGTTGCAA					
	CGAGAACAGCAACGCGCAAAAGAA					
	CTTGAAAACCGACAGAAGAAACTG					
	GAGCACGCCAACCGGCATTTGTTGC					
	TCAGAATACAGGAACTTGAAATGC					
	AGGCTCGAGCTCATGGACTTTCCCT					
	TATTCCATCCACGCGCTCTCTCCTCT					
	CCAGATTTCCTCA ATCCCATCA					
	AGCAAGAACCCGTTCTTCACAACT					
	GCAGCCAAGACCTCCTTCAGCATCA					
	TGCAGACCTAACCTGTACAACAACT					
	CTCGATCTCACGGATCGCACCATCA					
	CCTTCAACAACAACCTCCCCAACCA					
	GGACTGAGGCCAACCAAGCCTATA					
	GTGTCCCCCCCACAAAATCCCACCCTATA					
	AACTGGAAGACATCCTGATGGACG					
	ACACCCTTTCTCCCCCTCCCTCTCAC					
	TGATCCACTCCTTTCCTCAGTGTCC					
	CCCGGAGCTTCCAAAACAAGCAGC					
	CGGAGGAGCAGTATGAGCATGGAA					
	GAGACGGAGCACACTTGT					
IYC	ATGCCCCTCAACGTTAGCTTCACCA	68	Involved in cell	Pelengaris, S.,		
	ACAGGAACTATGACCTCGACTACG		proliferation,	Khan, M. &		
	ACTCGGTGCAGCCGTATTTCTACTG		differentiation	Evan, G. c-		
	CGACGAGGAGGAGAACTTCTACCA		and apoptosis.	MYC: more		
	GCAGCAGCAGCAGAGCGAGCTGCA		Reprogramming	than just a		
	GCCCCCGGCGCCCAGCGAGGATAT		factor for	matter of life		
	CTGGAAGAAATTCGAGCTGCTGCC		induction of	and death.		
	CACCCCGCCCTGTCCCCTAGCCGC		pluripotency.	Nat. Rev.		
	CGCTCCGGGCTCTGCTCGCCCTCCT			Cancer 2,		
	ACGTTGCGGTCACACCCTTCTCCCT			764-776		
	TCGGGGAGACAACGACGGCGGTGG			(2002).		
	CGGGAGCTTCTCCACGGCCGACCA			Takahashi, K.		
	GCTGGAGATGGTGACCGAGCTGCT			& Yamanaka,		
	GGGAGGAGACATGGTGAACCAGAG			S. Induction		
	TTTCATCTGCGACCCGGACGACGAG			of pluripotent		
	ACCTTCATCAAAAACATCATCATCC			stem cells		
	AGGACTGTATGTGGAGCGGCTTCTC			from mouse		
	GGCCGCCGCCAAGCTCGTCTCAGA			embryonic		
	GAAGCTGGCCTCCTACCAGGCTGC			and adult		
	GCGCAAAGACAGCGGCAGCCCGAA			fibroblast		
	CCCCGCCGCGGCCACAGCGTCTG			cultures by		
	CTCCACCTCCAGCTTGTACCTGCAG			defined		
	GATCTGAGCGCCGCCGCCTCAGAG			factors. Cell		
	TGCATCGACCCCTCGGTGGTCTTCC			126,663-76		
	CCTACCCTCTCAACGACAGCAGCTC			(2006).		
	GCCCAAGTCCTGCGCCTCGCAAGA			Takahashi, K.		
	CTCCAGCGCCTTCTCTCCGTCCTCG			et al.		
	GATTCTCTGCTCTCCTCGACGGAGT			Induction of		
	CCTCCCCGCAGGGCAGCCCCGAGC			pluripotent		
	CCCTGGTGCTCCATGAGGAGACAC			stem cells		
	CGCCCACCACCAGCAGCGACTCTG			from adult		
	AGGAGGAACAAGAAGATGAGGAA			human		
	GAAATCGATGTTGTTTCTGTGGAAA			fibroblasts by		
	AGAGGCAGGCTCCTGGCAAAAGGT			defined		
	CAGAGTCTGGATCACCTTCTGCTGG			factors. Cell		
	AGGCCACAGCAAACCTCCTCACAG			131,861-72		
	CCCACTGGTCCTCAAGAGGTGCCAC			(2007).		
	GTCTCCACACATCAGCACAACTACG			Yu, J. et al.		
	CAGCGCCTCCCTCCACTCGGAAGG			Induced		
	ACTATCCTGCTGCCAAGAGGGTCA			Pluripotent		
	AGTTGGACAGTGTCAGAGTCCTGA			Stem Cell		

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GACAGATCAGCAACAACCGAAAAT GCACCAGCCCCAGGTCCTCGGACA CCGAGGAGAATGTCAAGAGGCGAA CACACAACGTCTTGGAGCGCCAGA GGAGGAACGAGCTAAAACGGAGCT TTTTTGCCCTGCGTGACCAGATCCC GGAGTTGGAAAACAATGAAAAGGC CCCCAAGGTAGTTATCCTTAAAAAA GCCACAGCATACATCCTGTCGTCC AAGCAGGAGCATACATCCTGTCGTCC AAGCAGGAGCACAAAGCTCATTT CTGAAGAGGACTATGTGGGAAAC GCCGAGAACAGTTGAAACACAAAC TTGAACAGCTACGGAACTCTTGTGC G			Lines Derived from Human Somatic Cells. Science (80).318, 1917-1920 (2007).
MYCL	ATGGACTACGACTCGTACCAGCACT ATTTCTACGACTATGACTGCGGGGA GGATTTCTACCGCTCCACGGCGCCC AGCGAGGACATCTGGAAGAAATTC GAGCTGGTGGCCATCGCCCCCGG CGCCACCGGGGCCTGGGCCCCGG AGGGTGCACCGGAGCGTGGCCCGG AGGGTGCACCGGAGACGAAGCGGG ATCCCGGGGCCACTCGAAGGCTG GGCCAGGAACTACGCCTCCATCAT ACGCCGTGACTGCATGTGGAGCGG CTTCTCGGCCCGGGAACCGCCTGGA GAGAGCTGTGAGCGACCGGCCCGG ACGCCCCGGGCCCCGGGGAACCC CCCCAAGCGCCCCGGGGAACCCC GCCCCAAGCGCCCCGGCGCCCCGGA CTGCACTCCCAGCCCCGGCCCCGGA ACCCGGCGCCCCGGCGCCCCGGC AACCCGGGCCCCGGCCCCCGGC CCTGCTCCGGGCCCCGGCCCCCGGC CCTGCTCCGGGCCCCGGCCCCGGC CCTGCTCCGGGCCCCGGCCCCCGA GCGACTCGGGTAAGACCTCCCCG AGCCATCCAAGAGGGGGCCACCC ATGGGTGGCCAAAGCCTCCCCG AGCCATCCAAGAGGGGGCCCCCCG AGCCATCCAAGAGGGGCCCCCCCCCC	69	Involved in cell proliferation, differentiation and apoptosis.	Hatton, K. S. et al. Expression and activity of L-Myc in normal mouse development. Mol. Cell. Biol. 16, 1794-804 (1996).
MYCN	ATGCCGAGTTGTTCCACGTCTACGA TGCCAGGAATGATATGCAAGAACC CCGACTTGGAGTTTGACTCTTTGCA ACCATGCTTTTATTCGGGAGCAGAGAC GACTTTTATTCGGCGGCCCGGACA GCACCCCTCCTGGAGAGGACATCT GGAAAAATTCGAACTTTTGCCTAC ACCCCCACTCAGTCCTCTCGAGGA TTTGCGGAACACAGCAGTGAACCG CCGTCTTGGGTACAGAGAGATGCC TCGGTGGACCGAGTGACAGGAGAGCC CTGCGGAGGAAGACCGTTTCGGGC TCGGTGGACTGGAGCGCTTCCGGG GAACCAGTCATACTGCAGGATTG CATGTGGTCTGGATCTCACGCC GAACCCAGTCATACTGCAGGATTG CATGTGGTCTGGATCTCACGCC GAGAACCCAGTCATACTGCAGGATTG CATGTGGTCTGGATCTCACGCCG GAGAACTCCAACATGGCGGGGC CCTCCAACAGCGGGTCTACCGCCG GCGCCAGGAGCGGTTCTACCGCCAC AGTCCCCTGGTGCTGGAGCCCCTAG TCCCGCGGGGGAGAGCGCTAGGGCGG CGCTGCGTGCTGCTGCTGCGGCGGC CGCCGCGCGGCGGAGGGCCATGGGGC CGCCGCGCGCGCGGAGCGGTTCTCCGG CACCCCGCCGCTGCTGCGCCCGGT TAATAAGCGAGACCGGCCAGGCCCCGG GGGCACCAGGATACCCCCCCAG GGGCACCAGGATATGCCGCTCCCAG GGCCCGGCGAGAGGAACCCCCCCA GGCCCCGCGCGAGAGCACCCCCA	70	Involved in cell proliferation and differentiation	Malynn, B. A. et al. N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. Genes Dev. 14, 1390-9 (2000). Sawai, S. et al. Defects of embryonic organogenesis resulting from targeted disruption of the N-myc gene in the mouse. Development 117, 1445- 1455 (1993). Stanton, B. R., Perkins, A. S., Tessarollo, L., Sassoon, D. A. & Parada,

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
JENE	SEQUENCE ATAGTGATGACGAGGAGGACGAAG AGGAGGACGACGAGGAGGAGGAGAAG GATGTTGTCACGGTCGAGAAGCGA GTATTGTCACGGTCGAGAAGCGA GGACGTAAGAACGCAGCCCCCGGT CCAGGGCGGGCCCAGTCCAGT	NO:	ROLE	REFERENCES L. F. Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. Genes Dev. 6, 2235-47 (1992).
MYOD1	ATGGAGCTACTGTCGCCACCGCTCC GCGACGTAGACCTGACGGCCCCCG ACGGCTCTTCTGCTCCTTTGCCAC AACGGACGACTCTTGGCCCTTTGCCAC GTGTTCGACGCCCCGGACCTGCGC TTGTTCGAAGACCTGGACCTGGCC TGATGCACGTGGGGCGCCCCGG GCGCCCGGGGGGCACCGGCCCCGG GCGCACGTGAGGACGACGACCTGC CCGCGGCCGCGCCCGGCCCCGG GCGCCCACGTGCAACGCCACCACGG CGGCCGCCGCGCCCGCGCCCGG CAAGGCGGCCACGCGCACCACCAC GACCATGCGCAACGCGCACGCCC GACCATGCGCGACGCGCCCGC CACCATGCGCGACGCGCCCGCC GACCACTCAAGCGCTGCACGCCGC CACCATGCCGACGCGCCGCCGC GCCCCCGGCGCGCCGCCGCC CGCCCCGCGCGCCGC	71	Involved in skeletal muscle specification and differentiation Demonstrated to induce differentiation of hPSCs to skeletal muscle	Tapscott, S. J. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development 132, 2685- 2695 (2005). Abujarour, R. et al. Myogenic differentiation of muscular dystrophy- specific induced pluripotent stem cells for use in drug discovery. Stem Cells Transl. Med. 3,149-60 (2014).

		SEQ ID	SEQ ID				
GENE	SEQUENCE	NO :	ROLE	REFERENCES			
MYOG	ATGGAGCTGTATGAGACATCCCCCT	72	Involved in	Pownall, M.			
	ACTTCTACCAGGAACCCCGCTTCTA		skeletal muscle	E.,			
	TGATGGGGAAAACTACCTGCCTGTC		specification	Gustaisson,			
			and	M. K. &			
	CTCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		differentiation	Ellerson, C.			
				P. Myogenic Regulatory			
	CCCGAGCACTGTCCAGGCCAGTGC			Factors and			
	CTCCCCTCCCCTCTAACCTCTAA			the			
	AGAGGAAGTCGGTGTCCGTGGACC			Specification			
	GGCGGCGGGCGGCCACACTGAGGG			of Muscle			
	AGAAGCGCAGGCTCAAGAAGGTGA			Progenitors in			
	ATGAGGCCTTCGAGGCCCTGAAGA			Vertebrate			
	GAAGCACCCTGCTCAACCCCAACC			Embryos.			
	AGCGGCTGCCCAAGGTGGAGATCC			Annu. Rev.			
	TGCGCAGTGCCATCCAGTACATCGA			Cell Dev.			
	GCGCCTCCAGGCCCTGCTCAGCTCC			Biol. 18,747-			
	CTCAACCAGGAGGAGCGTGACCTC			783 (2002).			
	CGCTACCGGGGCGGGGGGGGGCCC			Shi, X. &			
	CAGCCAGGGGTGCCCAGCGAATGC			Garry, D. J.			
	AGCTCTCACAGCGCCTCCTGCAGTC			Muscle stem			
	CAGAGTGGGGCAGTGCACTGGAGT			cells in			
	TCAGCGCCAACCCAGGGGATCATC			development,			
	TGCTCACGGCTGACCCTACAGATGC			regeneration,			
	CCACAACCTGCACTCCCTCACCTCC			and disease.			
	ATCGTGGACAGCATCACAGTGGAA			Genes Dev.			
	GATGTGTCTGTGGCCTTCCCAGATG			20,1692-708			
	AAACCATGCCCAAC			(2006).			
NEURO	ATGACCAAATCGTACAGCGAGAGT	73	Involved in	Pataskar, A.			
D1	GGGCTGATGGGCGAGCCTCAGCCC		neuronal	et al.			
	CAAGGTCCTCCAAGCTGGACAGAC		specification	NeuroD1			
	GAGTGTCTCAGTTCTCAGGACGAG		and	reprograms			
	GAGCACGAGGCAGACAAGAAGGA		differentiation	chromatin and			
	GGACGACCTCGAAGCCATGAACGC		Demonstrated to	transcription			
	AGAGGAGGACTCACTGAGGAACGG		induce neuronal	factor			
	GGGAGAGGAGGAGGACGAAGATG		differentiation	landscapes to			
	AGGACCTGGAAGAGGAGGAAGAA		in hPSCs	induce the			
	GAGGAAGAGGAGGATGACGATCAA			neuronal			
				program.			
				EMBO J. 35,			
				24-45 (2016). Zhang V at			
	ACCCATCCACCCACTCAACCCCG			al Banid			
	CCCTACACGCACCCCCAACGCGG			air Rapid			
	TGCCTTGCTATTCTAAGACGCAGAA			induction of			
	GCTGTCCAAAATCGAGACTCTGCGC			functional			
	TTGGCCAAGAACTACATCTGGGCTC			neurons from			
	TGTCGGAGATCCTGCGCTCAGGCA			human			
	AAAGCCCAGACCTGGTCTCCTTCGT			pluripotent			
	TCAGACGCTTTGCAAGGGCTTATCC			stem cells.			
	CAACCCACCACCAACCTGGTTGCG			Neuron 78,			
	GGCTGCCTGCAACTCAATCCTCGGA			785-98			
	CTTTTCTGCCTGAGCAGAACCAGGA			(2013).			
	CATGCCCCCCCACCTGCCGACGGCC						
	AGCGCTTCCTTCCCTGTACACCCCT						
	ACTCCTACCAGTCGCCTGGGCTGCC						
	CAGTCCGCCTTACGGTACCATGGAC						
	AGCTCCCATGTCTTCCACGTTAAGC						
	CTCCGCCGCACGCCTACAGCGCAG						
	CGCTGGAGCCCTTCTTTGAAAGCCC						
	TCTGACTGATTGCACCAGCCCTTCC						
	TTTGATGGACCCCTCAGCCCGCCGC						
	TCAGCATCAATGGCAACTTCTCTTT						
	CAAACACGAACCGTCCGCCGAGTT						
	TGAGAAAAATTATGCCTTTACCATG						
	GGGGCCCAAAGCCACGGATCAATC						
	TTCTCAGGCACCGCTGCCCCTCGCT						
	GI CUTTUGATAGCCATTCACATCAT						
	GAGCGAGTCATGAGTGCCCAGCTC						
	AATGCCATATTTCATGAT						

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
NEURO G1	ATGCCAGCCCGCCTTGAGACCTGCA TCTCCGACCTCGACTGCGCCAGCAG CAGCGGCAGTGACTATCCGGCTTC CTCACCGACGAGGACTATCCGGCTTC CGGGGCCGCCAGGCAGGCCGCCGCA GGGCCGCCCCGCCGCCGGCCG	74	Involved in neuronal specification and differentiation	Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. Nat. Rev. Neurosci. 3, 517-530 (2002).
NEURO G3	ATGACACCACAACCATCTGGTGCTC CCACAGTCCAGGTGACGCGAGAGA CTGAAAGATCATTCCCACGCGGGTC CGAGGATGAGGTGACATGTCCAAC TAGCGCACCCCCCTCTCTCCCACCGG ACCCGCGGGAATGTGGAGAGCG ACCCAGGGAGGATGCAGAGGAGC ACCAAGGAAACTTCGAGCCGACG GGGTGGAAGAACTTCGAGCCGCACG GGGTGGAAGAACCTCCGAGGCG CCGCAGTCGGAGGAAAGCAACAG CGACCGGGAAAGGAATAGGATGCA TAATCTTAATTCTGCTCTGGACGCT CTGCGAGGCGTACTTCCTACTTTCC CGGATGACGCGAAATGACCAAGA TAACCTCGGGCTACTTCCTACATAA TTACATCTGAGCCGGTCACATAA TTACATCTGAGCCACCGGCCCCGCA CTGAGAATTGCCGATCACAGACT TGGGCTCTCTGAGCCACGCCCCGCA CTGTGGCGAGCTGGGTACCCCGGCCCCGCA CTGTGGCGAGCTGGGTACCCCGGCCCCGG CGGCTCTCTGAGCCACCGGCCCCGG CGGCTCTCTGAGCCACGGCTGCCAGCT TTGTATTCCCCGGTCGCCAGCCT CGAAGAAAGACCCGGACTCCTTGG AGCGACTTTTCAGCATGCCTGCC CCTGGCCAATGGCTTTCCCAGACT CGAGAAAGACCCGGACTCCTTGG AGCGACTTTTCAGCATGCCTGCC CCTGGCCCATGGCTTTCCCAGACT	75	Involved in pancreatic development, and neuronal specification and differentiation	<pre>Bertrand, N., Castro, D. S. &amp; Guillemot, F. Proneural genes and the specification of neural cell types. Nat. Rev. Neurosci. 3, 517-530 (2002). Arda, H. E. et al. Gene Regulatory Networks Governing Pancreas Development. Dev. Cell 25, 5-13 (2013).</pre>
NRL	ATGGCCCTGCCTCCCAGCCCGCTGG CCATGGAATATGTCAATGACTTTGA CTTGATGAAGTTTGAGGGCCGACCTGG GGAACCCTCTGAGGGCCCGACCTGG CCCACCTACAGCCTCACTGGGATCC ACACCTTACAGCTCAGTGCCTCCTT CACCCACCTTCAGTGAACCAGGCAT GGTAGGGCAACCGAGGGTACACG ACCAGGTTTGGAGGAGCTGTACTG GCTTGCTACCCTGCAGCAGCAGCTT GGGGCTGGGGAGGCATTGGGACTG AGTCCTGAAGAGGCCATGGGACTG AGTCCTGAAGAGGCCCAGTGGAGCTA CTGCAAGGTCAGGGCCCAGTGCCT GTTGATGGACCCCATGGTTACTACC CAGGGAGCCCCAGAGAGAGAGAG	76	Involved in photoreceptor development	Mears, A. J. et al. Nr1 is required for rod photoreceptor development. Nat. Genet. 29, 447-452 (2001).

TABLE 1-continued

TABLE	1-continued
	I CONCINCCO

		SEQ ID			
GENE	SEQUENCE	NO :	ROLE	REFERENCES	
	CCCAGCACGTTCAGTTGGCAGAGC GGTTTTCCGACGCGCGCGCTTGTCTC GATGTCTGTGCGAGAGACTAAACCG GCAGCTGCGGGGATGCGGGAGAGA CGAGGCTCTACGACTGAAGCAGAG GCGTCGAACGCTGAAGAACCGTGG CTATGCGCAAGCGTGGAGGCCGTGTT GAGGCCGAGCGCAGCGC				
ONECU T1	ATGAACGCGCAGCTGACCATGGAA GCGATCGGCGAGCTGCACGGGGTG AGCCATGAGCCGGCGCGCCCCCT GCCGACCTGCTGGGGCGCACCCCCCCCCGCGCCCCCCCGCGCCCCCGGCCCCCGGGCCCC	77	Involved in retinal, liver, gallbladder and pancreatic development	Chakrabarti, S. K., et al. Transcription factors direct the development and function of pancreatic β cells. Trends Endocrinol. Metab. 14, 78-84 (2003). Clotman, F. et al. The onecut transcription factor HNF6 is required for normal development 129,1819- 1828 (2002). Sapkota, D. et al. Onecut1 and Onecut2 redundant1y regulate early retinal cell fates during development. Proc. Natl. Acad. Sci. U. S. A. 111, E4086-95 (2014).	

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	~ CAACTTCTTCATGAACGCAAGAAG GAGGAGTCTGGACAAGTGGCAGGA CGAGGGCAGCTCCAATTCAGGCAA CTCATCTTCTTCATCAAGCACTTGT ACCAAAGCA			
OTX2	ATGATGTCTTATCTTAAGCAACCGC CTTACGCAGTCAATGGGCTGAGTCT GACCACTTCGGGTATGGACTTGCTG CACCACTCCGGGGTATGGACTTGCTG CACCCCCCGGGAACAGCCCGGCAGCCAC CCCCCGGGAACAGCGCGGGCAGCT AGATGTGCTGGAAGCACTGTTTGCC AAGACCCGGTACCCAGACATCTTC ATGCGAGAGGAGGTGGCACTGTTGCC AAGACCCGGTACCCAGACATCTTC ATGCGAGAGGAGTGCAAGAGAG GCTAAGTGCCGCAACAACAGCAA CAACATGCCCGCAACAACAGCAA CAACAGCAGAATGGAGGTCAAAAC AAAGTGAGACCTGCCAAAAAGAG ACATCTCCCAGCTCGGGAAGTGGCAAT TCACTCCCGCCTCTAGCACCCAGT CCGGACCATTGCCAGGACGCCAGT CCGGACCATTGCCAGCACGCAGT CCCGGCCAATGGCCAGCAGGGTG CAGGTCTATCTGGAGCCCAGT CCCGACCATTGCCAGCACGGCAGTGCT CCCGGCCCACTGCCAGACCCCAGT CCCGGCCCACTGCCAGCACGCAGC AGGTCCTATCCCAGCCCAG	78	Involved in photoreceptor differentiation, pineal gland development and induction and specification of forebrain and midbrain	Rhinn, M. et al. Sequential roles for Otx2 in visceral endoderm and neuroectoderm for forebrain and midbrain induction and specification. Development 125, 845-856 (1998). Nishida, A. et al. Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. Nat. Neurosci. 6,1255-1263 (2003).
PAX7	ATGGCGGCCCTTCCCGGCACGGTAC CGAGAATGATGCGGCCGGCTCCGG GGCAGAACTACCCCCGCACGGGAT TCCCTTTGGAAGTGTCCACCCGGCT TGGCCAAGGCCGGGTCCAATCAGCT GGGAGGGGCTTCATCAATGGGCG ACCCCTGCCTAACCACTCGGCCAC AAGATAGTGGAGATGGCCCACCAT GGCATCCGGCCTGGTCATCCCCCG GACAGCTGCGTGTCCTCCCACGGCTG CGTCTCCAAGATTCTTTGCCGGTAC CAGGAGACCGGGTCCATCCGGCCT GGGGCCATCGGCGCGCACCAGCGTG CAGCAGGGCACTCCGGCACCAAG AGGAAAAAGATTGAGAGTACAAG AGGGAAAACCCAGGCACGAGCTGCT GGGAGATCCGGGACAGGCTGCTG AAGGAAAACCCAGGACTGCTCGATGT GGGAGATCCGGGACAGGCTGCTG AAGGAAAACCCAGGACTGTCCGATGT GGGAAGACCGGGCCCCGAGCTGCTG AAGGAAAACCCAGGACTGTCCGACTG GGAAGAAAGAGGAGACGACGACCG ACTGTGCCCTCAGTAGGTCGATTA GCCGCGTCCTCAGAATCAAGTTCG GGAAGAAGAGGAGAGG	79	Involved in specification and differentiation of satellite cells Demonstrated to induce myogenic precursor differentiation in hPSCs	Darabi, R. et al. Human ES- and iPS- derived myogenic progenitors restore DYSTROPHIN and improve contractility upon transplantation in dystrophic mice. Cell Stem Cell 10, 610-9 (2012). Seale, P., et al. Pax7 Is Required for the Specification of Myogenic Satellite Cells. Cell 102, 777-786 (2000).

ABLE	1-cont	inued
------	--------	-------

		ana			
GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES	
	TACACCCGCGAGGAGCTGGCGCAG				
	AGGACCAAGCTGACAGAGGCGCGT				
	GTGCAGGTCTGGTTCAGTAACCGCC				
	GCGCCCGTTGGCGTAAGCAGGCAG				
	GAGCCAACCAGCTGGCGGCGTTCA				
	ACCACCTTCTGCCAGGAGGCTTCCC				
	GCCCACCGGCATGCCCACGCTGCC				
	CCCCTACCAGCTGCCGGACTCCACC				
	TACCCCACCACCATCTCCCAAG				
	ATGGGGGCAGCACTGTGCACCGGC				
	CTCAGCCCTGCCACCGTCCACCAT				
	GCACCAGGGCGGGCTGGCTGCAGC				
	GGCTGCAGCCGCCGACACCAGCTC				
	TGCCTACGGAGCCCGCCACAGCTTC				
	TCCAGCTACTCTGACAGCTTCATGA				
	ATCCGGCGGCGCCCTCCAACCACAT				
	GAACCCGGTCAGCAACGGCCTGTC				
	TCCTCAGGTGATGAGCATCTTGGGC				
	AACCCCAGTGCGGTGCCCCCGCAG				
	CCACAGGCTGACTTCTCCATCTCCC				
	CGCTGCATGGCGGCCTGGACTCGG				
	CCACCTCCATCTCAGCCAGCTGCAG				
	CCAGCGGGCCGACTCCATCAAGCC				
	AGGAGACAGCCTGCCCACCTCCCA				
	GGCCTACTGCCCACCCACCTACAGC				
	ACCACCGGCTACAGCGTGGACCCC				
	GTGGCCGGCTATCAGTACGGCCAG				
	TACGGCCAGAGTGAGTGCCTGGTG				
	CCCTGGGCGTCCCCGTCCCCATTC				
	CTTCTCCCACCCCAGGGCCTCCTG				
	CTTGTTTATGGAGAGCTACAAGGTG				
	GTGTCAGGGTGGGGAATGTCCATTT				
	CACAGATGGAAAAATTGAAGTCCA				
	GCCAGATGGAACAGTTCACC				
		00	Involved in	Turton T P	
OULFI		80	nituitary gland	C ot al	
	TCACCCCTCTCATCCCCCCCCATCCCCCCCCCCCCCCCC		development	U. EL AL.	
	ATAATCCATCACACTCIGCCICIG		geveropment	Mutationa	
	CTCTACCALCACAGIGCIGCCGAGI			mutations	
				within the	
				Cono	
	AUTTCATTATTCTGTTCCTTCCTGTC			Gene	
	ATTATGGAAACCAGCCATCAACCT			Associated	
	ATGGAGTGATGGCAGGTAGTTTAA			with Variable	
	CCCCTTGTCTTTATAAATTTCCTGA			Combined	
	CCACACCTTGAGTCATGGATTTCCT			Pituitary	
	CCTATACACCAGCCTCTTCTGGCAG			Hormone	
	AGGACCCCACAGCTGCTGATTTCAA			Deficiency. J.	
	GCAGGAACTCAGGCGGAAAAGTAA			Clin.	
	ATTGGTGGAAGAGCCAATAGACAT			Endocrinol.	
	GGATTCTCCAGAAATCAGAGAACT			Metab. 90,	
	TGAAAAGTTTGCCAATGAATTTAAA			4762-4770	
	GTGAGACGAATTAAATTAGGATAC			(2005).	
	ACCCAGACAAATGTTGGGGAGGCC				
	CTGGCAGCTGTGCATGGCTCTGAAT				
	TCAGTCAAACAACAATCTCCCCCATT				
	TGAAAATCTGCAGCTCAGCTTTAAA				
	ΔΔΤΩCΔΤΩCΔΔΔCTCAΔΔCCΔΔTA				
	TTATCCA AATCCCTCCACCA ACCT				
	AAAAGAAAACGAAGAACAACTATA				
	ACCA THOOTOGEN A A CARCONCE				
	AGCATTGCTGCTAAAGATGCTCTGG				
	AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA				
	AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA AACCTTCTTCTCAAGAGATCATGAG				
	AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA AACCTTCTTCTCAAGAGATCATGAG GATGGCTGAAGAACTGAATCTGGA				
	AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA AACCTTCTTCTCAAGAGATCATGAG GATGGCTGAAGAACTGAATCTGGA GAAAGAAGTAGTAAGAGTTTGGTT				
	AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA AACCTTCTTCTCAAGAGATCATGAG GATGGCTGAAGAACTGAATCTGGA GAAAGAAGTAGTAAGAGTTTGGTT TTGCAACCGGAGGCAGAGAGAAAA				
	AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA AACCTTCTTCTCAAGAGATCATGAG GATGGCTGAAGAACTGAATCTGGA GAAAGAAGTAGTAAGAGTTTGGTT TTGCAACCGGAGGCAGAGAAAA ACGGGTGAAAACAAGTCTGAATCA				
	AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA AACCTTCTTCTCAAGAGATCATGAG GATGGCTGAAGAACTGAATCTGGA GAAAGAAGTAGTAAGAGTTTGGTT TTGCAACCGGAGGCAAGAGAAAA ACGGGTGAAAACAAGTCTGAATCA GAGTTTATTTTCTATTTCTAAGGAA				
	AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA AACCTTCTCTCAAGAGAACAGAATA GATGGCTGAAGAACTGAATCTGGA GAAAGAAGTAGTAAGAGTTTGGTT TTGCAACCGGAGGCAGAGAGAAA ACGGGTGAAAACAAGTCTGAATCA GAGTTTATTTCTATTTCTAAGGAA CATCTTGAGTGCAGATCAGGCCTCA				
TABLE 1-continued

		SEQ ID		
GENE	SEQUENCE	NO :	ROLE	REFERENCES
POU5F1	ATGGCGGGACACCTGGCTTCAGATT TTGCCTTCTCGCCCCCTGCAGGTG TGGAGGTGATGGCCAGGGGGCC GGAGCCGGGCTGGGTTGATCCTCG GACCTGGCTAAGCTTCCAAGGCCG GGGGTTGGGCCAGGATCGGCCG CGTATGAGTTCTGTGGGGGATGG CGTACTGTGGGCCCAGGCTGGAGT GGGGCTACTGTGGGCCCAGGCGT GGAGACCTCTCAGCCTGAGGGCGA AGCAGGAGTCGGGGTGGAGAGCA CTCCGATGGGCCCCAGGCCG TGCACCGCAGGACGCCC TGCACCGCAGGGCTGGAGAGCA AGCAGGAGTCGGGGTGGAGAGCA CTCCGATGGGCCCCCGGAGCCG TGCACCGCAGGAGGCCC TGCACCGCAGGAGGCCC CGAAGTCGGGGGTGGAGAGCA AGCAGGAGTCGGGGTGGAGAGCA AGCAGGAGGATCACCGGGGCTGA AGCAGGAGAGGATCACCGGGCC TGCACCGCAGGAGGCCCC TGCACCGCAGGGGCCCAGGCA AAACCCGGAGGATCCCGGGGCCCAGCC TGCACCGCAGGCGAGGGGCCCCCC TGGGGGTCCTATTGGGAGAGGTATT CACGGCCGATGTGGGGCCCTGCT GCGGAAGTGGGGGGGGGG	81	Involved in regulation of pluripotency and embryogenesis. Reprogramming factor for induction of pluripotency	Boyer, L. A., et al. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell 122, 947-956 (2005). Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126,663-76 (2006). Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131,861-72 (2007). Yu, J. et al. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science (80). 318,
RUNX1	ATGGCTTCAGACAGCATATTTGAGT CATTTCCTTCGTACCCACAGTGCTT CATGAGAGAATGCATACTTGGAAT GAATCCTTCTAGAGACGTCCACGAT GCCAGCACGAGCGCCGCGCTCACG CCGCCTTCCACCGCGCCGCG	82	Involved in haematopoetic cell development	<pre>(2007). Woolf, E. et al. Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis. Proc. Natl. Acad. Sci. U. S. A. 100, 7731-6 (2003). Lacaud, G. et al. Runx1 is essential for</pre>

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CTGTGATGGCTGGCAATGATGAAA ACTACTCGGTGGGCTGAGCTGA			hematopoietic commitment at the hemangioblast stage of development in vitro. Blood 100, 458-66 (2002).
SIXI	ATGTCGATGCTGCCGTCGTTGGCT TTACGCAGGATCTGCAGCAAGTGGCGTGCG GAAACCTGGAGGTCTGCAGCAAGGCG GAAACCTGGACGACGCCAGGT TCCTGTGTCACTGCCGCCCCGCC	83	Involved in kidney, ear and olfactory epithelium development	Zheng, W. et al. The role of Six1 in mammalian auditory system development. Development 130, 3989- 4000 (2003). Xu, P. et al. Six1 is required for the early organogenesis of mammalian kidney. Development 130, 3085- 3094 (2003). Ikeda, K. et al. Six1 is essential for early neurogenesis in the development of olfactory epithelium. Dev. Biol.

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CAAACTATTCTCTCCCGGGCTTAAC AGCCTCGCAGCCCAGTCACGGCCT GCAGACCCACCAGCATCAGCTCCA AGACTCTCTGCTCGGCCCCCTCACC TCCAGTCTGGTGGACTTGGGGTCC			311, 53-68 (2007).
SIX2	ATGTCCATGCTGCCACCTTCGGCT TCACGCAGGAGCAAGTGCGCGGCG GCAACATCGAGCGCTGCGCCGCGCG GCAACATCGAGCGCTGGCCGCT TCCTGTGGTCGCTGCCCGCGCGCGCG GCACCTTCACAAGAATGAAAGCGT GCTCAAGGCCAAGACCACCCCGAG CTCTACAAGATCCTGGAGAGCCAC CAGTTCTCCCGCGCACAACCACCGCA AGCTGCACGAGCTGTGGCTCAAGG CACCTACATCGACGACCACCCCA AGCTGCAGCAGCTGTGGCTCAAGG CACCTACATCGACGAGCGCCGCA TGCGCGGCCGACCCCTGGGCGCCG TGGGCGAAGAGCAGCCCCTGGGCGCCG TGGGCGAAGAGCAGCCCCTGGGCGCCG CTTCAAGGAAAAGAGTCGCAGCGT GCTGCGCGAGGGGAGACCGCTGG GCTGCGCGAGTGGTACGCGCACAA CCCCTACCTTCACGAGCGGCCGCCG TGGGCGAAGTGGTACGCGCACAA CCCCTACCCTTCACGAGGCCACCAC GCTGCGCGAGTGGTACGCGCACAA CCCCTACCCTCACCACGGCGCCGC GCGGGGGGCGGCCGAGGCCAAGGAA AGGGAGAACAACGAGGCCAAGGAA AGGGAGAACAACGAGGAACTCCAAT TCTAACAGCCACAACCCCGCTGAAT GGCAGCGGCAGGCGAGCCCCATCATCA TCGGGGAGATGAGAAGACTCCCA TCGGGGGAGAGCGCAGCGC	84	Involved in kidney development	Kobayashi, A. et al. Six2 Defines and Regulates a Multipotent Self- Renewing Nephron Progenitor Population throughout Mammalian Kidney Development. Cell Stem Cell 3, 169- 181 (2008).
SNAI2	ATGCCGCGCTCCTTCCTGGTCAAGA AGCATTTCAACGCCTCCAAAAAGC CAAACTACAGCGAACTGGACACAC ATACAGTGATTATTTCCCCGTATCT CTATGAGAGTTACTTCCATGCCTGTC ATACCACAACCAGAGATCCTCAGC TCAGGAGCATACAGCGCCCATCACT GTGTGGACTACCGCTGCTCCATCTC TCCTCTTTCCGGATACTCCTCATCTT TGGGGCCAGGTGAGTCCCCCTCCTC ACGCCCAGCTACCCAATGGCCTCTC ATCTGACACCTCCTCCAAGGACCAC AGTGGCTCAGAAAGCCCCATTAGT GATGAAGAGGAAAGACTACAGTCC AAGCTGAAAAGGTTCAGTGCAATTT ATGCAATAAGACCTCATTCAGTGCAATTT TCTGGGCTGGCCAAACATAAGCAG CTGCACTACGATGCCAATAAGCAG CTGCACTACGATGCCAATAAGCAG CTGCACTACGATGCCAATAAGCAG CTGCACTACGATGCCAATAAGCAG CCTGCACTGCGATGCCCACTGGAA AATCTTTCAGCTGTAAATACTGTGA CAAGGAATAGTGTGAGCCTGGGCCC CCTGAAGATGCATATCGAGCCCAC ACATTACCTTGTGTTTGCAAGATCT GCGGCAAGGCGTTTTCCAGACCCTG GCTGCTCCAAGGCCACTAGAACT	85	Involved in neural crest development, epithelial- mesenchymal transition, and melanocyte stem cell development	Cobaleda, C., Perez-Caro, M., Vicente- Duelias, C. & Sanchez- Garcia, I. Function of the Zinc- Finger Transcription Factor SNA12 in Cancer and Development. Annu. Rev. Genet. 41, 41-61 (2007).

CACACGGGGGGAGAAGCCTTTTTCTT GCCCTCACTGCAACAGAGCATTTGC AGACAGGTCAAATCTGAGGGCTCA

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TCTGCAGACCCATTCTGATGTAAAG AAATACCAGTGCAAAAACTGCTCC AAAACCTTCTCCAGAATGTCTCTCC TGCACAAACATGAGGAATCTGGCT GCTGTGTAGCACAC			
SOX10	ATGGCGGAGGAGGAGCAGGACCTATCG GAGGTGGAGCTGAGCCCCGTGGCC TCGGAGGAGCCGCGCGCGCGCGCGGCGG GCTGCGGAGCAGGCGCGCGCGGGGGGGGGG	86	Involved in neuronal development	Southard- Smith, E. M., Kos, L. & Pavan, W. J. SOX10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. Nat. Genet. 18, 60-64 (1998). Britsch, S. et al. The transcription factor Sox10 is a key regulator of peripheral glial development. Genes Dev. 15, 66-78 (2001).
SOX2	ATGTACAACATGATGGAGACGGAG CTGAAGCCGCCGGGCCCGCAGCAA ACTTCGGGGGGCGCGCGGCGCG	87	Involved in regulation of pluripotency and embryogenesis, and in neuronal development. Reprogramming	Boyer, L. A., et al. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells.

TABLE 1-continued

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	AGCGCCTGGGCGCCGAGTGGAAAC TTTTGTCGAGGACGCGAGAGAGCGAC GCTCATCGACGAGGCGAGG		induction of pluripotency.	947-956 (2005). Graham, V. et al. SOX2 Functions to Maintain Neural Progenitor Identity. Neuron 39, 749-765 (2003). Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. Cell Stem Cell 10, 440- 454 (2012). Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-76 (2006). Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861-72 (2007). Yu, J. et al. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science (80). 318, 1917-1920 (2007).
SOX3	ATGCGACCTGTTCGAGAGAACTCAT CAGGTGCGAGAAGCCCGCGGGTTC CTGCTGATTTGGCGCGGAGCATTTT GATAAGCCTACCCTTCCCGCCGGAC TCGCTGGCCCACAGGCCCCCAAGCT CCGCTGCCCCCGGACTCCCCCCGGCCC	88	Involved in neuronal and pituitary development	Rizzoti, K. et al. SOX3 is required during the formation of

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TTTTCACCGTGGCCGCTCCAGCCCC GGGAGCGCCTTCTCCCTCCCGCCACG CTGGCGCACCTTCTCTCCCGCCCGG CAATGTACAGCCTTCTTGCGCCCGG CAATGTACAGCCTTCTTGGAGACTGA ACTCAAGAACCCCGTAGGGACCGG CACCACAGCGCGGCGCACCGGCGG CCCCGCGACCCGGGAGCGGCCAG CAGCGTGTGCGAACGCAGCCGG CGGCGCGAACTCGGCGGCGGCCAG CAGCGGTGGTGCGAGCGGGCGGG CAGCGGTGCCATGACGCCTTCAT GAAACGGCCCTGGAGACCGGCGG CAAAATGGCCCTGGAGAACCCTTCAT GATATGGTCCGCGGCGCACGGG CAAAATGGCCCTGGAGAACCCCAA GATGCACAATTCTGAGATCAGCAA GCTGTGCCGACGCGAGAAGCGACT TCGCGCCGTGCACATGAAGCGACC CGGCCTGGCCCAGAGAAGCGACC ATTCATCGACGACGCCGAGAAGCGACC GCGCCTGGCCCAGCAAGCGACC CGGCCTCCTGCCCCGGCGCCGGC CCGCAAGACCAAGACTCCCCCAG GCGCCTCCTGCCCCGGTGCCGCGG CCGCCGCGCCG			hypothalamo- pituitary axis. Nat. Genet. 36, 247-255 (2004).
SPI1	ATGTTACAGGCGTGCAAAATGGAA GGGTTTCCCCCTGTCCCCCCTAGC CATCAGAAGACCTGGTGCCCTATG ACACGGATCTATACCAACGCCAAA CGCACGAGTATTACCCACTATCTCAG CAGTGATGGGGAGAGCCATAGCGA CCATTACTGGGAGAGTTCCAACCCCCAC CACGTGCACAGCGAGTTCCAAGAGC TTCGCCGAGAACAACTTCACGGAG CTCCAGAGCGTGCAGCCCCCGCAG CTGCAGCAGCTGTACCACCCCCATGG AGCTGGAGCAGCTGCACCCCCATGG ATACCCCCATGGTGCCACCCCATCC CAGTCTTGGCCCCAGGCTCCCAGTACC CATCCCTGGCCCCCAGGCCCCCAGTCC CAGTCTTGCCCCAGGGAGCCCCCAGTCC CAGTCTTGCCCCAGGGCTCCCAGTACC CATCCCCGATGGGGAGGGGA	89	Involved in haematopoetic cell development	Scott, E. W. et al. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science 265, 1573-1577 (1994). Rosenbauer, F. & Tenen, D. G. Transcription factors in myeloid development: balancing differentiation

TABLE 1-continued

GENE	SEOUENCE	SEQ ID NO:	ROLE	REFERENCES
	ACAGCATCTGGTGGGTGGACAAGG ACAAGGCACCTTCCAGTTCTCGTC CAAGCACAAGAGGGCGCTGGCGCA CCGCTGGGGCATCCAGAAGGGCAA CCGCAAGAAGATGACCTACCAGAA GATGGCGCGCGCGCGCGCGCAACTA CGGCAAGAAGAAGACGCGCGAAGTCAAGAA GGTGAAGAAGAAGACCTACCA GTTCAGCGCGCAAGTCCACCTACCA GTTCAGCGCCGAAGTCTGGGCCG CGGGGGCCTGGCCGAGCGGCGCCA CCCGCCCCAC			with transformation. Nat. Rev. Immunol. 7, 105-117 (2007).
SPIB	ATGCTCGCCTGGAGGCTGCACAG CTCGACGGCCCACATTCAGCTGTC TGTACCCAGATGCGCTCTTCATGA CCTGGACAGCTGCAAGCATTCCAG GACTCCCTGGACAGGAGCTGGGC CCCCACCTGTCCAGGACTGGGCCCCGC CAGCTCTGGTACGACCCCCGGCAGCC CAGCTCTGCTACGACCCCCCAGC CAGCTCTGCTACGACCCCCCAGC GGCCTGGCCCCCGGGAGGCCCCGG GGCCTGGCCT	90	Involved in differentiation of lymphoid cells	Maroulakou, I. G. & Bowe, D. B. Expression and function of Ets transcription factors in mammalian development: a regulatory network. Oncogene 19, 6432-6442 (2000).
SPIC	ATGACGTGTGTTGAACAAGACAAG CTGGGTCAAGCATTTGAAGATGCTT TTGAGGTTCTGAGGCAACATTCAAC TGGAGATCTTCAGTACTCGCCAGAT TACAGAAATTACCTGGCTTTAATTCA ACCATCGTCCTCATGTCAAAGGAA ATTCCAGCTGCTATGAAGGTGTTGCC TACAGAGGAGACCTGTCTATAATTGG AGAACGGTAATTACAGTGCTGCG GACTTCTATTTTGAAGGAAATATTC ATCCAACTCTGCAGAACACAACCCACTCTT CTCCAGCTAGCTAGCAGACACAACTGA AAACCAGCTGGTACAACCCACTCTT CTCCAGCAAAAGGGGGAAAAGGC AGGAAGAGGCTCGGACCACTCTT CCCGGAGATGGCATCTGGTATAATC GGTAGATAAAACCAAAGGCATCTT TCCAGTTCACGAATCCCGGCATAACTGG GGTAGATAAAACCAAAGGCAACACA AAACCTGCCGAGCTCTTGGAGAA AAAACTTGCCGAGCTTTGGGGGAA AAAACTTGCCGAGCTTTGGGGGAA AAAAACTTGCCGAGCTTTGGGGGAA AAAAACTTGCCGAGCACAGGAAGACCCAT GACTTACCAGAAAATGGCCAGGGC ACTCACAGAAATACCGAAGAACTGG GGAAATTACCAAAATCCGGAGGAA GCTGACTTACCAGTCAGTGAGGCC ATTCTCCAAAGACTCTCTCCATCCT ATTTCCCAGGGAAAAGACTCTCT TCCACGTTGGGGAAAAGGACTCTT TCCCAGAGATCACCGAACACCAACCAA GCTGACTTACCAGTCAACTGGAGCC ATTCTCCAAAGACTCTCTCCATCCT	91	Involved in macrophage development	Kohyama, M. et al. Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. Nature 457, 318-321 (2009).

TABLE 1-continued

CENT	CEOUENCE	SEQ ID	DOLE	
GENE	SEQUENCE	NO:	ROLE	REFERENCES
	ATGCAAATTATAATTATACATATGC CAATTACCATGAGCTAAATCACCAT GATTGC			
SRY	ATGCAATCATATGCTTCTGCTATGT TAAGCGTATTCAACAGCGATGATTA CAGTCCACCTGTGCAAGAGAATAT TCCCGCTCTCCGGAGAAGCTGTAACT CTAAGTATCAGTGTGAAAGCTGTAACT CTAAGTATCAGTGTGAAACGGGAG AAAACAGTAAAGGCAACGTCCAGG ATAGAGTGAAGCGACCCATGAACG CATTCATCGTGTGGTCTCGCGATCA GAGGCGCAAGATGGCACTCAGAGA TCCCAGAATGCGAAACTCAGAGAT CAGCAACCAGCTGGAAACTCAGAGAT CAGCAACCAGCTGGAAACTCAGAGAT CAGCAACCAGCTGGAAACTCAGAGAT CAGCAACCAGCTGGAAACCAGAGA GAAAATGCTTACTGAAGCCGAAAA ATGGCCATTCTTCCAGGAGGCACA GAAAATGCCGAATTATAAGTATCG ACCTCGTCGGAAGGCGACAGAGT GCCGAAGATGCCATTGTACGAAGT GCCGAAGTGCAACTGGACACAGGT TGTACAGGGCAACTGGACAACAGGT TGTACAGGCAACTGGACAACAGGT TGTACAGGCAACTGGACAACAGGT TGTACAGGCCACTTACCGCCATCAA CGCAGCCACTCACGCAACGGCCACCA AGCTAGCCACTCACGCAACAGG GGACCGCTACCGCCACCGACAAC GCAGCCACTCACGCACACGGCACAAA GCTG	92	Involved in sex determination and spermatogenesis	Polanco, J. C. & Koopman, P. Sry and the hesitant beginnings of male development. Dev. Biol. 302,13-24 (2007). Koopman, P. et al. Male development of chromosomally female mice transgenic for Sry. Nature 351,117-121 (1991).
TBX5	ATGGCCGACGACGACGACGAGGGCTTT GGCCTGGCGCACACGCCTCTGGAG CCTGACGCAAAAGACCTGCCTGC GATTCGAAACCCGAGAGCGCCGCTC GGGGCCCCCAGCAAGTCCCCGTCG TCCCCGCAGGCGCCTTCACCCAGC AGGGCATGGAGGGGAATCAAAGTGT TTCTCCATGAAAGAGAACTGTGTGT TTCTCCATGAAAGAGAACTGTGGCA AAAATTCCACGAAGTGGCACGGA AATGATCATAACCAAGGCTGGAAG GCGGATGTTCCCCAGTAACAGGTG AAGGTGACGGGCCTTAATCCCAAA ACGAAGTACATTCTTCTCATGGACA TTGTACCTGCCGACGATCACAGATA CAAATTCGCAGATAATAAATGGTCT GTGACGGCCAAGCCCGGCCCCC ATGCCTGGCCGCCCTGTACGTGCACC CAGACTCATCCATCGGACGCC ATGGCTGGCCGCCCTGTACGTGCACC CAGACTCAAGGCCCGCTCCTT CCAGAAACTCAAGCTCACCACAATCC AGCCTGGCCGCCCTGTCCCTT CCAGAAACTCAAGCTCACCAAATACC AGCCTAGATTCCATGCGCACAATCC AGCCTGGACGCATTGGGCACATT CAAAAATACCATCGTGAAAG CGACTGACTCCTCCCCCCACGAGCCC ACGCTGGACCCATTGGGCATTTA AGCAGTGACTCCTCCCCCCACACCA CAACTCCCCCGCACCGGGACCC CAGAATGCAAATAATGGATTTGGCT CAAAAATACAGCGATTCGCCACC CAAGATCACGCAATTAAAGATTGA GAATAATCCCTTGCCAAAGATCG CACAGATGCCAATGCAACCA CAACGACGAGTGACAAGGCCGC ACGCCCGGGGCAATGAACCA CACCACCGGAGTGACAGGACTG CACAGATGCCAATGGACCGC CCCACCCCGGGCCAATGCCCCCAGG AGCCCCCGGGGCCAATACCACCA CAAGAATGCCCTTCCCCCCCAGG AGCACCCGTGGGCCCAATGCCCCCCC AGGACCCCCGGGGCAAAAGT AAAGAATGCCCTTCCCCCCCCCC	93	Involved in cardiac development	Bruneau, B. G. et al. A Murine Model of Holt-Oram Syndrome Defines Roles of the T-Box Transcription Factor Tbx5 in Cardiogenesis and Disease. Cell 106, 709-721 (2001).

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GATTCCTTCTACCGCTCTAGCTATC CACAGCAGCAGGGCCTGGGTGCCT CCTACAGGACAGAGTCGGCACAGC GGCAAGCTTGCATGTATGCCAGCTC TGCGCCCCCAGCGAGCCTGTGCACC AGCCTAGAGGACATCAGCTGCAAC ACGTGGCCAAGCATGCCTTCCTACA GCACCTGCACCGTCACCACCGTGC AGCCCATGGACAGGCTACCCTACC			
TFAP2 C	ATGTTGTGGGAAATAGCTTG ATGTTGTGGAAAATAACCGATAT GTCAAGTACGAAGAGGGACTGCGAG GATCGCCGCGCGCGCGCCACCTCT CCTCCGCCGGGCGCCCCCCCTCTCACAG CCCCGCGCCACCCCTCTCCCACAT GGAGTCGCCGACTCTCCCCACCAT GGAGTCGCCGACTCTCCCCACCAT CCCCCGCCCCCCCCCC	94	Involved in trophectoderm development	Cao, Z. et al. Transcription factor AP-2Y induces early Cdx2 expression and represses HIPPO signaling to specify the trophectoderm lineage. Development 142, 1606-15 (2015).

TABLE 1-continued

		SEQ ID	1			
GENE	SEQUENCE	NO:	ROLE	REFERENCES		
	AAATCCTACATGAACCCTGGAGAC CAGAGTCCAGCTGATTCTAACAAA ACCCTGGAGAAAATGGAGAAACAC					
	AGGAAA					

TABLE	2
	~

Sample_ID	Description	Media Condition	F N	Estimated fumber of M Cells	Mean Reads per Cell	Median Genes per Cell
UP_TF_1	HighMOI, (-) TRA-1-60 MACS sorted	Pluripotent stem cell		3,640	45,983	3,317
UP_TF_2	HighMOI, Unsorted	Pluripotent stem cell		3,505	49,750	3,843
UP_TF_3	HighMOI, Unsorted	Pluripotent stem cell		4,223	45,403	3,972
UP_TF_4	HighMOI, (-) TRA-1-60 MACS corted	Pluripotent stem cell		3,461	56,290	4,475
UP_TF_5	LowMOI, (-) TRA-1-60	Pluripotent stem cell		3,748	46,895	4,165
UP_TF_8	Library, Endothelial	Endothelial growth medium		3,563	41,056	3,698
UP_TF_10	Library, Multilineage	Multilineage differentiation		2,129	70,519	5,605
UP_TF_11	Library, Endothelial	Endothelial growth		6,574	23,250	3,105
UP_TF_12	Library, Multilineage	Multilineage differentiation		4,678	30,340	3,882
UP_TF_13	KLF Family, cMYC Mutants	Pluripotent stem cell medium		5,590	35,913	3,620
Sample_ID	Number of Reads	Valid Barcodes	Reads Mapped Confidently to Exonic Regions	Sequencin Saturation	Fraction ng Reads in n Cells	Median UMI Counts per Cell
UP_TF_1	167,381,505	97.90%	65.60%	17.00%	55.40%	11,785
UP_TF_2	174,376,238	98.40%	70.30%	20.80%	63.90%	15,985
UP_TF_3	191,740,141	98.10%	63.10%	18.90%	77.20%	16,090
UP_TF_4	194,819,799	98.20%	66.80%	25.00%	78.60%	19,132
UP_TF_5	175,765,276	98.10%	65.70%	17.70%	76.90%	17,349
UP_TF_8	146,283,407	98.20%	65.20%	16.60%	80.90%	15,049
UP_TF_10	150,135,344	98.20%	68.60%	20.20%	83.00%	27,785
UP_TF_11	152,847,871	98.20%	69.40%	11.20%	86.80%	10,681
UP_IF_I2	141,934,669	98.20%	/0.00%	11.00%	88.10%	14,526

66.20%

# TABLE 3

98.00%

UP\_TF\_13 200,756,922

## TABLE 3-continued

	Nun	iber of Genotyped	Cells		Nun	iber of Genotyped	Cells
Genotype	Stem cell media	Endothelial media	Multilineage media	Genotype	Stem cell media	Endothelial media	Multilineage media
ASCL1	186	78	21	ASCL5	140	64	51
ASCL3	471	150	89	ATF7	97	49	45
ASCL4	286	90	75	CDX2	267	192	103

78.70%

14,286

15.50%

	INDEL 5	-commucu			IT IDEL 5	continueu	
	Nun	nber of Genotyped	Cells	_	Nun	iber of Genotyped	Cells
Genotype	Stem cell media	Endothelial media	Multilineage media	Genotype	Stem cell media	Endothelial media	Multilineage media
CRX	292	107	54	MYC	291	113	36
ERG	62	30	7	MYCL	356	112	75
ESRRG	169	98	64	MYCN	50	33	12
ETV2	60	22	21	MYODI	197	68	40
FLI1	55	27	18	MYOG	284	122	81
FOXA1	53	27	14	NEUROD1	83	46	10
FOXA2	89	46	37	NEUROG1	154	103	23
FOXA3	255	90	61	NEUROG3	158	138	41
FOXP1	413	112	94	NRL	249	75	49
GATA1	288	111	72	ONECUT1	159	109	58
GATA2	62	81	60	OTX2	293	95	47
GATA4	71	101	58	PAX7	86	56	28
GATA6	44	44	35	POU1F1	126	61	50
GLI1	27	11	16	POU5F1	78	30	24
HAND2	310	113	81	RUNX1	139	47	43
HNF1A	88	45	39	SIX1	260	119	66
HNF1B	53	30	41	SIX2	295	103	84
HOXA1	166	67	57	SNAI2	485	96	50
HOXA10	344	111	66	SOX10	83	54	30
HOXA11	237	82	47	SOX2	137	53	27
HOXB6	166	95	44	SOX3	137	56	31
KLF4	298	259	145	SPI1	264	142	67
LHX3	175	76	45	SPIB	199	70	47
LMX1A	458	155	82	SPIC	147	80	35
mCherry	1689	689	495	SRY	166	61	65
MEF2C	87	49	51	TBX5	149	112	35
MESP1	227	70	55	TFAP2C	90	58	34
MITE	73	63	45				

TABLE 3-continued

TABLE 3-continued

TABLE 4

		Enrichment p-	value for each	genotype in clu	isters using Fis	sher's exact tes	t
	C6	C2	C5	C3	C1	C7	C4
CDX2	0.999581	0.502321	1	1	1	3.42E-58	1
KLF4	0.688329	1.12E-27	1	1	1	1	3.82E-21
FOXA1	0.848222	1	1	8.00E-08	1	1	1
FOXA2	0.559116	1	1	2.56E-15	1	0.788874	1
GATA2	0.002284	1	1.57E-10	1	1	0.91906	0.832613
GATA4	0.009787	0.781098	1.13E-09	1	0.553072	1	0.822422
GATA6	0.03266	0.23167	0.000147	1	1	1	1
SOX10	0.017774	0.043271	1	1	1	0.12661	1
NEUROD1	0.280233	1	1	1	1	0.34423	1
ETV2	0.016254	1	1	1	1	0.054486	1
SPIB	9.93E-07	1	0.29024	0.190193	1	1	1
SOX3	1.53E-05	1	1	1	1	1	0.063768
NEUROG3	6.23E-06	1	1	0.502271	1	0.50894	1
TBX5	1.71E-07	1	1	0.449045	1	1	1
MYOD1	3.73E-07	1	1	1	1	1	0.115324
MYC	9.91E-05	0.611641	1	1	0.394338	0.779857	1
ESRRG	5.02E-12	0.233929	1	1	0.58849	1	1
TFAP2C	6.90E-05	1	0.541387	1	1	1	0.638171
GLI1	0.017877	1	1	1	1	1	0.380973
NEUROG1	0.00162	1	1	1	1	0.620425	1
ASCL5	9.82E-08	0.737393	1	1	1	0.353463	1
FOXA3	3.08E-15	1	1	0.644816	1	1	1
ATF7	2.03E-09	1	1	0.534822	1	1	1
HOXA10	2.36E-09	1	0.4436	0.673452	0.599648	1	0.85978
SOX2	4.01E-06	1	0.461875	1	1	1	1
ONECUT1	2.98E-11	1	1	0.626421	1	1	0.822422
RUNX1	3.65E-07	1	1	1	0.450277	1	0.364314
SIX2	8.69E-16	0.888323	1	1	1	0.677188	0.710842
HOXA11	4.51E-09	1	1	1	1	0.860947	0.406197
SPIC	1.28E-06	1	1	1	1	1	0.648778
MYCL	2.52E-22	1	1	1	1	1	1
FOXP1	9.41E-17	0.702249	1	0.795614	0.374912	0.980162	1
SNAI2	4.89E-09	1	0.681398	1	1	0.616212	1
HNF1A	7.52E-11	1	1	1	1	1	1
LMX1A	2.74E-19	1	0.845485	1	1	1	0.912434

		Enrichment p-	value for each	genotype in clu	isters using Fig	sher's exact tes	t
	C6	C2	C5	C3	C1	C7	C4
ERG	0.164469	1	1	1	1	1	1
HAND2	7.41E-17	1	1	1	1	0.653393	1
MITF	2.07E-10	1	0.643049	1	1	1	1
PAX7	1.57E-05	1	1	1	1	0.692249	1
SIX1	1.58E-14	0.822135	1	1	0.599648	1	1
OTX2	3.17E-08	0.708559	1	1	1	1	0.754072
SPI1	5.65E-12	0.826686	1	1	1	0.767724	1
GATA1	2.36E-13	0.847734	1	1	1	1	0.629688
MYOG	7.41E-17	1	1	0.746058	1	0.966092	1
HNF1B	1.21E-06	1	1	1	0.434855	1	1
POU1F1	2.52E-14	1	1	1	1	1	1
FLI1	0.000193	1	1	1	1	1	1
HOXA1	3.20E-15	1	1	1	1	1	1
SRY	1.01E-17	1	1	1	1	1	1
CRX	4.15E-13	1	1	1	1	0.896121	1
ASCL1	0.000199	1	1	1	1	1	1
NRL	9.14E-09	1	1	1	0.494018	0.872071	1
LHX3	1.65E-11	1	1	1	1	1	1
MESP1	2.47E-11	1	1	1	0.534212	1	0.805949
HOXB6	3.05E-08	1	1	1	1	1	1
ASCL4	3.41E-17	1	1	1	0.646165	0.956545	1
MYCN	0.00932	1	1	1	1	1	1
MEF2C	3.40E-10	1	1	1	1	1	0.78156
POU5F1	3.21E-06	1	1	1	1	1	1
ASCL3	3.49E-19	1	1	1	0.707836	1	1
mCherry	1.64E-91	0.99443	0.961129	0.996934	0.263601	0.994961	0.947099

TABLE 4-continued

TABLE 5

TABLE 5-continued

Module	Description	n_genes	Module	Description	n_genes
GM1	Cytoskeleton and polarity	444	GM7	Embryonic development	509
GM2	Ion transport	973	GM8	Mitochondrial metabolism and translation	2242
GM3	Chromatin accessibility	1568	GM9	Ribosome biogenesis	190
GM4	Signaling pathways	873	GM10	Growth factor response	492
GM5	Neuron differentiation	444	GM11	Pluripotent state	234
GM6	Notch pathway	859			

.

7	DT	1.2	C
IA	. В Г	1 H.	n

Gene	Forward Primer (5'→3')	SEQ ID NO:	Reverse Primer (5'→3')	SEQ ID NO:
CDH5	AGACCACGCCTCTGTCATGTACCAAATC	95	CACGATCTCATACCTGGCCTGCTTC	113
PECAM1	GGTCAGCAGCATCGTGGTCAACATAAC	96	TGGAGCAGGACAGGTTCAGTCTTTCA	114
VWF	TCTCCGTGGTCCTGAAGCAGACATA	97	AGGTTGCTGCTGGTGAGGTCATT	115
KDR	AGCCATGTGGTCTCTCTGGTTGTGTATG	98	GTTTGAGTGGTGCCGTACTGGTAGGA	116
NANOG	TTTGTGGGCCTGAAGAAAACT	99	AGGGCTGTCCTGAATAAGCAG	117
POU5F1	CTTGAATCCCGAATGGAAAGGG	100	GTGTATATCCCAGGGTGATCCTC	118
SOX2	TACAGCATGTCCTACTCGCAG	101	GAGGAAGAGGTAACCACAGGG	119
DNMT3B	GAGTCCATTGCTGTTGGAACCG	102	ATGTCCCTCTTGTCGCCAACCT	120
SALL2	CAGCGGAAACCCCAACAGTTA	103	GAGGGTCAGTAGAACATGCGT	121
DPPA4	GACCTCCACAGAGAAGTCGAG	104	TGCCTTTTTCTTAGGGCAGAG	122
VIM	AGTCCACTGAGTACCGGAGAC	105	CATTTCACGCATCTGGCGTTC	123
CDH1	CGAGAGCTACACGTTCACGG	106	GGGTGTCGAGGGAAAAATAGG	124
CDH2	AGCCAACCTTAACTGAGGAGT	107	GGCAAGTTGATTGGAGGGATG	125

TABLE 6-continued

Gene	Forward Primer (5'→3')	SEQ ID NO:	Reverse Primer (5'→3')	SEQ ID NO:
EDGAM		1.00		100
EPCAM	IGAICCIGACIGCGAIGAGAG	108		126
LAMC1	GGCAACGTGGCCTTTTCTAC	109	AGTGGCAGTTACCCATTCCTG	127
SPP1	GAAGTTTCGCAGACCTGACAT	110	GTATGCACCATTCAACTCCTCG	128
THY1	ATCGCTCTCCTGCTAACAGTC	111	CTCGTACTGGATGGGTGAACT	129
TPM2	CTGAGACCCGAGCAGAGTTTG	112	TGAATCTCGACGTTCTCCTCC	130

#### REFERENCES

- [0153] 1. Xu, J., Du, Y. & Deng, H. Direct lineage reprogramming: strategies, mechanisms, and applications. *Cell Stem Cell* 16, 119-34 (2015).
- **[0154]** 2. Davis, Robert L; Weintraub, Harold; Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987-1000 (1987).
- [0155] 3. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663-76 (2006).
- [0156] 4. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861-72 (2007).
- [0157] 5. Yu, J. et al. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* 318, 1917-1920 (2007).
- [0158] 6. Wernig, M. et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 448, 318-324 (2007).
- **[0159]** 7. Maherali, N. et al. Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. *Cell Stem Cell* 1, 55-70 (2007).
- [0160] 8. Park, I.-H. et al. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 451, 141-146 (2008).
- **[0161]** 9. Pang, Z. P. et al. Induction of human neuronal cells by defined transcription factors. *Nature* 476, 220-223 (2011).
- **[0162]** 10. Sugimura, R. et al. Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* 545, 432-438 (2017).
- [0163] 11. Yang, N. et al. Generation of pure GABAergic neurons by transcription factor programming. *Nat. Methods* 14, 621-628 (2017).
- [0164] 12. Sugimura, R. et al. Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* 545, 432-438 (2017).
- [0165] 13. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78, 785-98 (2013).
- [0166] 14. Abujarour, R. et al. Myogenic differentiation of muscular dystrophy-specific induced pluripotent stem cells for use in drug discovery. *Stem Cells Transl. Med.* 3, 149-60 (2014).

- [0167] 15. Chanda, S. et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. *Stem Cell Reports* 3, 282-96 (2014).
- [0168] 16. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* 58, 610-20 (2015).
- [0169] 17. Mohr, S., Bakal, C. & Perrimon, N. Genomic screening with RNAi: results and challenges. *Annu. Rev. Biochem.* 79, 37-64 (2010).
- [0170] 18. Shalem, O., Sanjana, N. E. & Zhang, F. Highthroughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* 16, 299-311 (2015).
- [0171] 19. Adamson, B. et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 167, 1867-1882.e21 (2016).
- [0172] 20. Dixit, A. et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167, 1853-1866.e17 (2016).
- **[0173]** 21. Jaitin, D. A. et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* 167, 1883-1896.e15 (2016).
- [0174] 22. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell* 66, 285-299. e5 (2017).
- [0175] 23. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* 14, 297-301 (2017).
- [0176] 24. Macosko, E. Z. et al. Highly Parallel Genomewide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202-1214 (2015).
- [0177] 25. Nishiyama, A. et al. Uncovering Early Response of Gene Regulatory Networks in ESCs by Systematic Induction of Transcription Factors. *Cell Stem Cell* 5, 420-433
- [0178] 26. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. arXiv 1-12 (2008). doi:10.1088/1742-5468/ 2008/10/P10008
- [0179] 27. Orkin, S. H. & Hochedlinger, K. Chromatin connections to pluripotency and cellular reprogramming. Cell 145, 835 (2011).
- **[0180]** 28. Busskamp, V. et al. Rapid neurogenesis through transcriptional activation in human stem cells. *Mol Syst Blol* 10, (2014).
- [0181] 29. Velkey, J. M. & O'Shea, K. S. Expression of Neurogenin 1 in mouse embryonic stem cells directs the

differentiation of neuronal precursors and identifies unique patterns of down-stream gene expression. *Dev. Dyn.* 242, 230-53 (2013).

- [0182] 30. Castro, D. S. et al. A novel function of the proneural factor Ascll in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev.* 25, 930-45 (2011).
- **[0183]** 31. Tapscott, S. J. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* 132, 2685-2695 (2005).
- [0184] 32. Treutlein, B. et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNAseq. *Nature* 534, 391-5 (2016).
- **[0185]** 33. Niwa, H. et al. Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation. *Cell* 123, 917-929 (2005).
- [0186] 34. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer* 2, 764-776 (2002).
- [0187] 35. McConnell, B. B. & Yang, V. W. Mammalian Kruppel-like factors in health and diseases. *Physiol. Rev.* 90, 1337-81 (2010).
- [0188] 36. Tiwari, N. et al. Klf4 Is a Transcriptional Regulator of Genes Critical for EMT, Including Jnk1 (Mapk8). *PLoS One* 8, e57329 (2013).
- [0189] 37. Zhang, B. et al. KLF5 activates microRNA 200 transcription to maintain epithelial characteristics and

prevent induced epithelial-mesenchymal transition in epithelial cells. *Mol. Cell. Biol.* 33, 4919-35 (2013).

- [0190] 38. Gumireddy, K. et al. KLF17 is a negative regulator of epithelial-mesenchymal transition and metastasis in breast cancer. *Nat. Cell Biol.* 11, 1297-304 (2009).
- [0191] 39. Liu, Y.-N. et al. Critical and reciprocal regulation of KLF4 and SLUG in transforming growth factor (3-initiated prostate cancer epithelial-mesenchymal transition. *Mol. Cell. Biol.* 32, 941-53 (2012).
- **[0192]** 40. Li, R. et al. A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. *Cell Stem Cell* 7, 51-63 (2010).
- [0193] 41. Barrallo-Gimeno, A., Nieto, M. A. & Ip, Y. T. The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* 132, 3151-61 (2005).
- [0194] 42. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545-15550 (2005).
- [0195] 43. Morita, R. et al. ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. *Proc. Natl. Acad. Sci.* 112, 160-165 (2015).
- [0196] 44. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/ Cas9 knockout screens. *Genome Biol.* 15, 554 (2014)

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 130

<210> SEQ ID NO 1 <211> LENGTH: 1413 <212> TYPE: DNA <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

tteetgttge	agataagccc	agcttagccc	agctgacccc	agaccctctc	ccctcactcc	60
ccccatgtcg	caggatcgag	accctgaggc	agacagcccg	ttcaccaagc	cccccgcccc	120
gcccccatca	ccccgtaaac	ttctcccagc	ctccgccctg	ccctcaccca	gcccgctgtt	180
ccccaagcct	cgctccaagc	ccacgccacc	cctgcagcag	ggcagcccca	gaggccagca	240
cctatccccg	aggetggggt	cgaggctcgg	ccccgcccct	gcctctgcaa	cttgagcctg	300
gctgcgaccc	ctgctctgac	gtctcggaaa	attccccctt	gcccaggccc	ttgggggagg	360
gggtgcatgg	tatgaaatgg	ggctgagacc	cccggctggg	ggcagaggaa	cccgccagag	420
aaggagccaa	attaggette	tgtttccctg	atctggcact	ccaagggggac	acgccgacag	480
cgacagcaga	gacatgctgg	aaaggtacaa	gctcatccct	ggcaagcttc	ccacagctgg	540
actgggggctc	cgcgttactg	cacccagaag	ttccatgggg	ggcggagccc	gactctcagg	600
ctcttccgtg	gtccgggggac	tggacagaca	tggcgtgcac	agcctgggac	tcttggagcg	660
gcgcctcgca	gaccctgggc	cccgcccctc	tcggcccggg	ccccatcccc	gccgccggct	720
ccgaaggcgc	cgcgggccag	aactgcgtcc	ccgtggcggg	agaggccacc	tegtggtege	780
gcgcccaggc	cgccgggagc	aacaccagct	gggactgttc	tgtggggccc	gacggcgata	840
cctactqqqq	caqtqqcctq	qqcqqqqaqc	cqcqcacqqa	ctqtaccatt	tcataaaaca	900

74

				-001011	Iueu		
ggcccgcggg	cccggactgt	accacctcct	ggaacccggg	gctgcatgcg	ggtggcacca	960	
cctctttgaa	gcggtaccag	ageteagete	tcaccgtttg	ctccgaaccg	agcccgcagt	1020	
cggaccgtgc	cagtttggct	cgatgcccca	aaactaacca	ccgaggtccc	attcagctgt	1080	
ggcagttcct	cctggagctg	ctccacgacg	gggcgcgtag	cagctgcatc	cgttggactg	1140	
gcaacagccg	cgagttccag	ctgtgcgacc	ccaaagaggt	ggctcggctg	tggggcgagc	1200	
gcaagagaaa	gccgggcatg	aattacgaga	agctgagccg	gggccttcgc	tactactatc	1260	
gccgcgacat	cgtgcgcaag	agcgggggggc	gaaagtacac	gtaccgcttc	gggggccgcg	1320	
tgcccagcct	agcctatccg	gactgtgcgg	gaggcggacg	gggagcagag	acacaataaa	1380	
aattcccggt	caaacctcaa	aaaaaaaaaa	aaa			1413	
<210> SEQ : <211> LENG' <212> TYPE <213> ORGAI <220> FEATU <223> OTHEN polyn	ID NO 2 IH: 1275 : DNA NISM: Artif: JRE: R INFORMATIC nucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	ience: Synth	etic	
<400> SEQUI	ENCE: 2						
atgcccctca	acgttagctt	caccaacagg	aactatgacc	tcgactacga	ctcggtgcag	60	
ccgtatttct	actgcgacga	ggaggagaac	ttctaccagc	agcagcagca	gagcgagctg	120	
cageceeegg	cgggatcagg	tagcggtagc	cgccgctccg	ggetetgete	geeeteetae	180	
gttgcggtca	caccettete	ccttcgggga	gacaacgacg	gcggtggcgg	gagettetee	240	
acggccgacc	agctggagat	ggtgaccgag	ctgctgggag	gagacatggt	gaaccagagt	300	
ttcatctgcg	acccggacga	cgagaccttc	atcaaaaaca	tcatcatcca	ggactgtatg	360	
tggagcggct	tctcggccgc	cgccaagctc	gtctcagaga	agctggcctc	ctaccaggct	420	
gcgcgcaaag	acagcggcag	cccgaacccc	gcccgcggcc	acagcgtctg	ctccacctcc	480	
agcttgtacc	tgcaggatct	gagcgccgcc	gcctcagagt	gcatcgaccc	ctcggtggtc	540	
tteecetace	ctctcaacga	cagcagctcg	cccaagtcct	gcgcctcgca	agactccagc	600	
gccttctctc	cgtcctcgga	ttetetgete	tcctcgacgg	agteeteece	gcagggcagc	660	
cccgagcccc	tggtgctcca	tgaggagaca	ccgcccacca	ccagcagcga	ctctgaggag	720	
gaacaagaag	atgaggaaga	aatcgatgtt	gtttctgtgg	aaaagaggca	ggctcctggc	780	
aaaaggtcag	agtctggatc	accttctgct	ggaggccaca	gcaaacctcc	tcacagecea	840	
ctggtcctca	agaggtgcca	cgtctccaca	catcagcaca	actacgcagc	gcctccctcc	900	
actcggaagg	actatcctgc	tgccaagagg	gtcaagttgg	acagtgtcag	agtcctgaga	960	
cagatcagca	acaaccgaaa	atgcaccagc	cccaggtcct	cggacaccga	ggagaatgtc	1020	
aagaggcgaa	cacacaacgt	cttggagcgc	cagaggagga	acgagctaaa	acggagcttt	1080	
tttgccctgc	gtgaccagat	cccggagttg	gaaaacaatg	aaaaggcccc	caaggtagtt	1140	
atccttaaaa	aagccacagc	atacatcctg	tccgtccaag	cagaggagca	aaagctcatt	1200	
tctgaagagg	acttgttgcg	gaaacgacga	gaacagttga	aacacaaact	tgaacagcta	1260	
cggaactctt	gtgcg					1275	

<210> SEQ ID NO 3 <211> LENGTH: 1287

-continued

			- CONCIN	lued		
<pre><li><li><li><li><li><li><li><li><li><li< td=""><td>ificial Sequer TION: Descript e</td><td>nce tion of Art:</td><td>ificial Sequ</td><td>lence: Synth</td><td>etic</td><td></td></li<></li></li></li></li></li></li></li></li></li></pre>	ificial Sequer TION: Descript e	nce tion of Art:	ificial Sequ	lence: Synth	etic	
<400> SEQUENCE: 3						
atgcccctca acgttagc	tt caccaacagg	aactatgacc	tcgactacga	ctcggtgcag	60	
ccgtatttct actgcgac	ga ggaggagaac	ttctaccagc	agcagcagca	gagcgagctg	120	
cageceeegg egeceage	ga ggatatctgg	aagaaattcg	agctgctgcc	caccccgccc	180	
ctgtccccta gccgccgc	te egggetetge	tcgccctcct	acgttgcggt	cacacccttc	240	
teeetteggg gagacaac	ga cggcggtggc	gggagcttct	ccacggccga	ccagctggag	300	
atggtgaccg agctgctg	gg aggagacatg	gtgaaccaga	gtttcatctg	cgacccggac	360	
gacgagacct tcatcaaa	aa catcggatca	ggtagcggtc	tcgtctcaga	gaagetggee	420	
teetaccagg ctgegege	aa agacagcggc	agcccgaacc	ccgcccgcgg	ccacagcgtc	480	
tgctccacct ccagcttg	ta cctgcaggat	ctgagcgccg	ccgcctcaga	gtgcatcgac	540	
ccctcggtgg tetteccc	ta ccctctcaac	gacagcagct	cgcccaagtc	ctgcgcctcg	600	
caagactcca gegeette	tc tccgtcctcg	gattetetge	tctcctcgac	ggagtcctcc	660	
ccgcagggca gccccgag	cc cctggtgctc	catgaggaga	caccgcccac	caccagcagc	720	
gactctgagg aggaacaa	ga agatgaggaa	gaaatcgatg	ttgtttctgt	ggaaaagagg	780	
caggeteetg geaaaagg	tc agagtctgga	tcaccttctg	ctggaggcca	cagcaaacct	840	
cctcacagcc cactggtc	ct caagaggtgc	cacgtctcca	cacatcagca	caactacgca	900	
gegeeteeet ceactegg	aa ggactatcct	gctgccaaga	gggtcaagtt	ggacagtgtc	960	
agagteetga gacagate	ag caacaaccga	aaatgcacca	gccccaggtc	ctcggacacc	1020	
gaggagaatg tcaagagg	cg aacacacaac	gtettggage	gccagaggag	gaacgagcta	1080	
aaacggagct tttttgcc	ct gcgtgaccag	atcccggagt	tggaaaacaa	tgaaaaggcc	1140	
cccaaggtag ttatcctt	aa aaaagccaca	gcatacatcc	tgtccgtcca	agcagaggag	1200	
caaaagctca tttctgaa	ga ggacttgttg	cggaaacgac	gagaacagtt	gaaacacaaa	1260	
cttgaacagc tacggaac	tc ttgtgcg				1287	
<pre>&lt;210&gt; SEQ ID NO 4 &lt;211&gt; LENGTH: 1305 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Art &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMA polynucleotid</pre>	ificial Sequer TION: Descript e	nce tion of Art:	ificial Sequ	uence: Synth	etic	
<400> SEQUENCE: 4						
atgeceetca aegttage	tt caccaacagg	aactatgacc	tcgactacga	ctcggtgcag	60	
ccgtatttct actgcgac	ga ggaggagaac	ttctaccagc	agcagcagca	gagcgagctg	120	
cageeeeegg egeeeage	ga ggatatctgg	aagaaattcg	agctgctgcc	caccccgccc	180	
ctgtccccta gccgccgc	te egggetetge	tcgccctcct	acgttgcggt	cacacccttc	240	
teeetteggg gagacaac	ga cggcggtggc	gggagcttct	ccacggccga	ccagctggag	300	
atggtgaccg agctgctg	gg aggagacatg	gtgaaccaga	gtttcatctg	cgacccggac	360	
gacgagacct tcatcaaa	aa catcatcatc	caggactgta	tgtggagcgg	cttctcggcc	420	

	cogcoccaga	gaagetggee	tcctaccagg	ctgcgcgcaa	agacagcggc	480
agcccgaacc	ccgcccgcgg	ccacagcgtc	tgctccacct	ccagcttgta	cctgcaggat	540
ctgagcgccg	ccgcctcaga	gtgcatcgac	ccctcggtgg	tcttccccta	ccctctcaac	600
gacagcagct	cgcccaagtc	ctgcgcctcg	caagactcca	gcgcettete	teegteeteg	660
gattetetge	tctcctcgac	ggagtcctcc	ccgcagggca	gccccgagcc	cctggtgctc	720
catgaggaga	caccgcccac	caccagcagc	gactctgagg	aggaacaaga	agatgaggaa	780
gaaatcgatg	ttgtttctgt	ggaaaagagg	caggctcctg	gcaaaaggtc	agagtctgga	840
tcaccttctg	ctggaggcca	cagcaaacct	cctcacagcc	cactggtcct	caagaggtgc	900
cacgtctcca	cacatcagca	caactacgca	gcgcctccct	ccactcggaa	ggactatgga	960
tcaggtagcg	gtagtgtcag	agtcctgaga	cagatcagca	acaaccgaaa	atgcaccagc	1020
cccaggtcct	cggacaccga	ggagaatgtc	aagaggcgaa	cacacaacgt	cttggagcgc	1080
cagaggagga	acgagctaaa	acggagcttt	tttgccctgc	gtgaccagat	cccggagttg	1140
gaaaacaatg	aaaaggcccc	caaggtagtt	atccttaaaa	aagccacagc	atacatcctg	1200
tccgtccaag	cagaggagca	aaagctcatt	tctgaagagg	acttgttgcg	gaaacgacga	1260
gaacagttga	aacacaaact	tgaacagcta	cggaactctt	gtgcg		1305
<210> SEQ <211> LENG <212> TYPE <213> ORGA <220> FEAT <223> OTHE poly	ID NO 5 TH: 1290 : DNA NISM: Artif: URE: R INFORMATIC nucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	lence: Synth	etic
<400> SEQU	ENCE: 5					
<400> SEQU atgcccctca	ENCE: 5 acgttagctt	caccaacagg	aactatgacc	tcgactacga	ctcggtgcag	60
<400> SEQU atgcccctca ccgtatttct	ENCE: 5 acgttagctt actgcgacga	caccaacagg ggaggagaac	aactatgacc ttctaccagc	tcgactacga agcagcagca	ctcggtgcag gagcgagctg	60 120
<400> SEQU atgcccctca ccgtatttct cagcccccgg	ENCE: 5 acgttagctt actgcgacga cgcccagcga	caccaacagg ggaggagaac ggatatctgg	aactatgacc ttctaccagc aagaaattcg	tcgactacga agcagcagca agctgctgcc	ctcggtgcag gagcgagctg caccccgccc	60 120 180
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta	ENCE: 5 acgttagctt actgcgacga cgcccagcga gccgccgctc	caccaacagg ggaggagaac ggatatctgg cgggctctgc	aactatgacc ttctaccagc aagaaattcg tcgccctcct	tcgactacga agcagcagca agctgctgcc acgttgcggt	ctcggtgcag gagcgagctg caccccgccc cacacccttc	60 120 180 240
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg	ENCE: 5 acgttagctt actgcgacga cgcccagcga gccgccgctc gagacaacga	caccaacagg ggaggagaaac ggatatctgg cgggctctgc cggcggtggc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag	60 120 180 240 300
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg	ENCE: 5 acgttagctt actgcgacga cgcccagcga gccgccgctc gagacaacga agctgctggg	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgacccggac	60 120 180 240 300 360
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct	ENCE: 5 acgttagctt actgcgacga cgcccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa	caccaacagg ggaggagaaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc	60 120 180 240 300 360 420
<400> SEQU atgecectea cegtatttet cagececegg etgtececta tecetteggg atggtgaceg gaegagaeet geegecaage	ENCE: 5 acgttagctt actgcgacga cgcccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga	caccaacagg ggaggagaac ggatatctgg cggggtctgc cggcggtggc aggagacatg catcatcatc gaagctggcc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa	ctcggtgcag gagcgagctg caccccgccc cacacccttc ccagctggag cgacccggac cttctcggcc agacagcggc	60 120 180 240 300 360 420 480
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc	ENCE: 5 acgttagctt actgcgacga cgcccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg	caccaacagg ggaggagaaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat	60 120 180 240 300 360 420 480 540
<400> SEQU atgecectea cegtatttet cagececegg etgtececta tecetteggg atggtgaceg gaegagaeet geegecaage ageegaace	ENCE: 5 acgttagctt actgcgacga gccgccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcctcaga	caccaacagg ggaggagaaac ggatatctgg cggggtctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac	60 120 180 240 300 360 420 480 540 600
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcagct	ENCE: 5 acgttagctt actgcgacga gccgccagcga gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgccccaga	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct ccctcggtgg caagactcca	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctccaac	60 120 180 240 300 360 420 480 540 600 660
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcagct gattctctgc	ENCE: 5 acgttagctt actgcgacga gccgccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcctcaga cgcccaagtc tctcctcgac	caccaacagg ggaggagaaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac ctgcgcctcg	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct ccctcggtgg caagactcca ccgcagggca	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc	ctcggtgcag gagcgagctg caccccgccc cacacccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac tccgtcctcg	60 120 180 240 300 360 420 480 540 600 660 720
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcagct gattctctgc catgaggaga	ENCE: 5 acgttagctt actgcgacga gccgccagcga gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcctcaga cgcccaagtc tctcctcgac	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac ctgcgcctcg ggagtcctcc caccagcagc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct ccctcggtgg caagactcca ccgcagggca gactctgagg	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc gccccgagcc aggaacaaga	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac tccgtcctcg agatgggca	60 120 240 300 420 480 540 600 660 720 780
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcagct gattctctgc catgaggaga	ENCE: 5 acgttagctt actgcgacga gccgccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcccagatc tctcctcgac caccgcccac ttgtttctgt	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac ctgcgcctcg ggagtcctcc caccagcagc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct ccctcggtgg caagactcca gactctgagg caggctcctg	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc gccccgagcc aggaacaaga	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctccaac tccgtcctcg cctggtgctc agatgaggaa agagtctgga	60 120 180 240 300 420 480 540 600 660 720 780 840
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcagct gattctctgc catgaggaga gaaatcgatg tcaccttctg	ENCE: 5 acgttagctt actgcgacga gccgccagcga gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcccagtc tctcctcgac tctcctcgac	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ctgcgcctcg ggagtcctcc caccagcagc ggaaaagagg cagcaaacct	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct ccctcggtgg caagactcca gactctgagg caggctcctg caggctcctg	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc gccccgagcc aggaacaaga gcaaaaggtc cactggtcct	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgaccggac cttctcggcc agacagcggc cctgcaggat ccctgcacga tccgtcctca ac tccgtcctcg agatgaggaa agagtctgga	60 120 180 240 300 420 480 540 600 660 720 780 840 900
<400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcagct gattctctgc catgaggaga tcaccttctg cacgtctcca	ENCE: 5 acgttagctt actgcgacga gccgccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcccaga tctcctcgac tactcctcgac tactctcgac caccgcccac ttgtttctgt ctggaggcca cacatcagca	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc gtgcatcgac ctgcgcctcg ggagtcctcc caccagcagc ggaaaagagg cagcaaacct caactacgca	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga tcctaccagg tgctccacct ccctcggtgg caagactcca gactctgagg caggctcctg cctcacagcc gcgccccct	t cgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gccccgagcc aggaacaaga gcaaaaggtc cactggtcct	ctcggtgcag gagcgagctg caccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac tccgtcctcg cctggtgctc agatgaggaa agagtctgga ggactatcct	60 120 180 240 300 420 480 540 600 660 720 780 840 900 960

				concin	lucu		
aaatgcacca	gccccaggtc	ctcggacacc	gaggagaatg	tcggatcagg	tagcggtgag	1080	
ctaaaacgga	gctttttgc	cctgcgtgac	cagatcccgg	agttggaaaa	caatgaaaag	1140	
gcccccaagg	tagttatcct	taaaaaagcc	acagcataca	tcctgtccgt	ccaagcagag	1200	
gagcaaaagc	tcatttctga	agaggacttg	ttgcggaaac	gacgagaaca	gttgaaacac	1260	
aaacttgaac	agctacggaa	ctcttgtgcg				1290	
<210> SEQ 1 <211> LENG <212> TYPE <213> ORGAN <220> FEATU <223> OTHEN polyn	ID NO 6 TH: 1206 : DNA NISM: Artif JRE: R INFORMATI( nucleotide	icial Sequen DN: Descript	nce tion of Art:	ificial Seq	ience: Synthe	tic	
<400> SEQUE	ENCE: 6						
atgcccctca	acgttagctt	caccaacagg	aactatgacc	tcgactacga	ctcggtgcag	60	
ccgtatttct	actgcgacga	ggaggagaac	ttctaccagc	agcagcagca	gagcgagctg	120	
cageceeegg	cgcccagcga	ggatatctgg	aagaaattcg	agctgctgcc	caccccgccc	180	
ctgtccccta	gccgccgctc	cgggctctgc	tcgccctcct	acgttgcggt	cacacccttc	240	
tcccttcggg	gagacaacga	cggcggtggc	gggagcttct	ccacggccga	ccagctggag	300	
atggtgaccg	agctgctggg	aggagacatg	gtgaaccaga	gtttcatctg	cgacccggac	360	
gacgagacct	tcatcaaaaa	catcatcatc	caggactgta	tgtggagcgg	cttctcggcc	420	
gccgccaagc	tcgtctcaga	gaagctggcc	tcctaccagg	ctgcgcgcaa	agacagcggc	480	
agcccgaacc	ccgcccgcgg	ccacagcgtc	tgctccacct	ccagcttgta	cctgcaggat	540	
ctgagcgccg	ccgcctcaga	gtgcatcgac	ccctcggtgg	tcttccccta	ccctctcaac	600	
gacagcagct	cgcccaagtc	ctgcgcctcg	caagactcca	gcgccttctc	tccgtcctcg	660	
gattetetge	tctcctcgac	ggagtcctcc	ccgcagggca	gccccgagcc	cctggtgctc	720	
catgaggaga	caccgcccac	caccagcagc	gactctgagg	aggaacaaga	agatgaggaa	780	
gaaatcgatg	ttgtttctgt	ggaaaagagg	caggeteetg	gcaaaaggtc	agagtctgga	840	
tcaccttctg	ctggaggcca	cagcaaacct	cctcacagcc	cactggtcct	caagaggtgc	900	
cacgteteca	cacatcagca	caactacgca	gcgcctccct	ccactcggaa	ggactatcct	960	
gctgccaaga	gggtcaagtt	ggacagtgtc	agagteetga	gacagatcag	caacaaccga	1020	
aaatgcacca	gccccaggtc	ctcggacacc	gaggagaatg	tcaagaggcg	aacacacaac	1080	
gtettggage	gccagaggag	gaacggatca	ggtagcggtc	aaaagctcat	ttctgaagag	1140	
gacttgttgc	ggaaacgacg	agaacagttg	aaacacaaac	ttgaacagct	acggaactct	1200	
tgtgcg						1206	
<210> SEQ 1 <211> LENG <212> TYPE <213> ORGA <220> FEAT <223> OTHEN polyr	ID NO 7 FH: 1230 : DNA NISM: Artif RE: R INFORMATIO nucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Seq	lence: Synthe	tic	
<400> SEQUE	ENCE: 7						

ccgtatttct	actgcgacga	ggaggagaac	ttctaccagc	agcagcagca	gagcgagctg	120	
cagcccccgg	cgcccagcga	ggatatctgg	aagaaattcg	agctgctgcc	caccccgccc	180	
ctgtccccta	gccgccgctc	cgggctctgc	tcgccctcct	acgttgcggt	cacacccttc	240	
tcccttcggg	gagacaacga	cggcggtggc	gggagcttct	ccacggccga	ccagctggag	300	
atggtgaccg	agctgctggg	aggagacatg	gtgaaccaga	gtttcatctg	cgacccggac	360	
gacgagacct	tcatcaaaaa	catcatcatc	caggactgta	tgtggagcgg	cttctcggcc	420	
gccgccaagc	tcgtctcaga	gaagctggcc	tcctaccagg	ctgcgcgcaa	agacagcggc	480	
agcccgaacc	ccgcccgcgg	ccacagcgtc	tgctccacct	ccagcttgta	cctgcaggat	540	
ctgagcgccg	ccgcctcaga	gtgcatcgac	ccctcggtgg	tcttccccta	ccctctcaac	600	
gacagcagct	cgcccaagtc	ctgcgcctcg	caagactcca	gcgccttctc	tccgtcctcg	660	
gattctctgc	tctcctcgac	ggagtcctcc	ccgcagggca	gccccgagcc	cctggtgctc	720	
catgaggaga	caccgcccac	caccagcagc	gactctgagg	aggaacaaga	agatgaggaa	780	
gaaatcgatg	ttgtttctgt	ggaaaagagg	caggctcctg	gcaaaaggtc	agagtctgga	840	
tcaccttctg	ctggaggcca	cagcaaacct	cctcacagcc	cactggtcct	caagaggtgc	900	
cacgtctcca	cacatcagca	caactacgca	gcgcctccct	ccactcggaa	ggactatcct	960	
gctgccaaga	gggtcaagtt	ggacagtgtc	agagtcctga	gacagatcag	caacaaccga	1020	
aaatgcacca	gccccaggtc	ctcggacacc	gaggagaatg	tcaagaggcg	aacacacaac	1080	
gtcttggagc	gccagaggag	gaacgagcta	aaacggagct	ttttgccct	gcgtgaccag	1140	
atcccggagt	tggaaaacaa	tgaaaaggcc	cccaaggtag	ttatccttaa	aaaagccaca	1200	
gcatacatcc	tgtccgtcca	agcagaggag				1230	
<210> SEQ 3 <211> LENG' <212> TYPE <213> ORGAJ <220> FEATU <223> OTHEI polyn	ID NO 8 TH: 906 : DNA NISM: Artif: JRE: R INFORMATIC nucleotide	icial Seque DN: Descrip	nce tion of Art:	ificial Sequ	lence: Synth	netic	
<400> SEQU	ENCE: 8						
atgggatcag	gtagcggtct	cgtctcagag	aagetggeet	cctaccaggc	tgcgcgcaaa	60	
gacagcggca	gcccgaaccc	cgcccgcggc	cacagcgtct	gctccacctc	cagcttgtac	120	
ctgcaggatc	tgagcgccgc	cgcctcagag	tgcatcgacc	cctcggtggt	cttcccctac	180	
cctctcaacg	acagcagctc	gcccaagtcc	tgcgcctcgc	aagactccag	cgccttctct	240	
ccgtcctcgg	attctctgct	ctcctcgacg	gagtcctccc	cgcagggcag	ccccgagccc	300	
ctggtgctcc	atgaggagac	accgcccacc	accagcagcg	actctgagga	ggaacaagaa	360	
gatgaggaag	aaatcgatgt	tgtttctgtg	gaaaagaggc	aggctcctgg	caaaaggtca	420	
gagtctggat	caccttctgc	tggaggccac	agcaaacctc	ctcacagccc	actggtcctc	480	
aagaggtgcc	acgtetecac	acatcagcac	aactacgcag	cgcctccctc	cactcggaag	540	
gactatcctg	ctgccaagag	ggtcaagttg	gacagtgtca	gagtcctgag	acagatcagc	600	
aacaaccgaa	aatgcaccag	ccccaggtcc	tcggacaccg	aggagaatgt	caagaggcga	660	
acacacaacq	tcttggagcq	ccagaggagq	aacgagctaa	aacggagctt	ttttgccctg	720	
cgtgaccaga	tcccggagtt	ggaaaacaat	gaaaaqqccc	ccaaggtagt	tatccttaaa	780	
				55 5-			

aaagccacag catacatcct gtccgtccaa gcagaggagc aaaagctcat ttctgaagag	840
gacttgttgc ggaaacgacg agaacagttg aaacacaaac ttgaacagct acggaactct	900
tgtgcg	906
<210> SEQ ID NO 9 <211> LENGTH: 1062 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	etic
<400> SEQUENCE: 9	
atgcccctca acgttagctt caccaacagg aactatgacc tcgactacga ctcggtgcag	60
ccgtatttct actgcgacga ggaggagaac ttctaccagc agcagcagca gagcgagctg	120
cagcccccgg cgcccagcga ggatatctgg aagaaattcg agctgctgcc caccccgccc	180
ctgtccccta gccgccgctc cgggctctgc tcgccctcct acgttgcggt cacacccttc	240
tcccttcggg gagacaacga cggcggtggc gggagcttct ccacggccga ccagctggag	300
atggtgaccg agctgctggg aggagacatg gtgaaccaga gtttcatctg cgacccggac	360
gacgagacct tcatcaaaaa catcatcatc caggactgta tgtggagcgg cttctcggcc	420
gccgccaagc tcgtctcaga gaagctggcc tcctaccagg ctgcgcgcaa agacagcggc	480
agecegaaee eegeeegegg eeacagegte tgeteeaeet eeagettgta eetgeaggat	540
ctgagegeeg eegeeteaga gtgeategae eeeteggtgg tetteeeeta eeeteteaae	600
gacagcaget egeceaagte etgegeeteg caagaeteea gegeettete teegteeteg	660
gattetetge teteetegae ggagteetee eegeagggea geeeegagee eetggtgete	720
catgaggaga caccgcccac caccagcagc gactctgagg aggaacaaga agatgaggaa	780
gaaatcgatg ttgtttctgt ggaaaagagg caggctcctg gcaaaaggtc agagtctgga	840
teacettetg etggaggeea eageaaacet eeteaeagee eaetggteet eaagaggtge	900
cacgteteca cacateagea caaetaegea gegeeteeet ceaeteggaa ggaetateet	960
gctgccaaga gggtcaagtt ggacagtgtc agagtcctga gacagatcag caacaaccga	1020
aaatgcacca gccccaggtc ctcggacacc gaggagaatg tc	1062
<210> SEQ ID NO 10 <211> LENGTH: 1317 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	etic
<400> SEQUENCE: 10	
atgcccctca acgttagctt caccaacagg aactatgacc tcgactacga ctcggtgcag	60
ccgtatttct actgcgacga ggaggagaac ttctaccagc agcagcagca gagcgcgctg	120
cageceeegg egeceagega ggatatetgg aagaaatteg agetgetgee eaceeegeee	180
ctgtccccta gccgccgctc cgggctctgc tcgccctcct acgttgcggt cacacccttc	240
teeetteggg gagacaacga eggeggtgge gggagettet eeaeggeega eeagetggag	300
atggtgaccg agctgctggg aggagacatg gtgaaccaga gtttcatctg cgacccggac	360

	tcatcaaaaa	catcatcatc	caggactgta	tgtggagcgg	cttctcggcc	420
gccgccaagc	tcgtctcaga	gaagetggee	tcctaccagg	ctgcgcgcaa	agacagcggc	480
agcccgaacc	ccgcccgcgg	ccacagcgtc	tgctccacct	ccagcttgta	cctgcaggat	540
ctgagcgccg	ccgcctcaga	gtgcatcgac	ccctcggtgg	tcttccccta	ccctctcaac	600
gacagcagct	cgcccaagtc	ctgcgcctcg	caagactcca	gcgccttctc	teegteeteg	660
gattctctgc	tctcctcgac	ggagtcctcc	ccgcagggca	gccccgagcc	cctggtgctc	720
catgaggaga	caccgcccac	caccagcagc	gactctgagg	aggaacaaga	agatgaggaa	780
gaaatcgatg	ttgtttctgt	ggaaaagagg	caggeteetg	gcaaaaggtc	agagtctgga	840
tcaccttctg	ctggaggcca	cagcaaacct	cctcacagcc	cactggtcct	caagaggtgc	900
cacgteteca	cacatcagca	caactacgca	gegeeteeet	ccactcggaa	ggactateet	960
gctgccaaga	gggtcaagtt	ggacagtgtc	agagtcctga	gacagatcag	caacaaccga	1020
aaatgcacca	gccccaggtc	ctcggacacc	gaggagaatg	tcaagaggcg	aacacacaac	1080
gtcttggagc	gccagaggag	gaacgagcta	aaacggagct	ttttgccct	gcgtgaccag	1140
atcccggagt	tggaaaacaa	tgaaaaggcc	cccaaggtag	ttatccttaa	aaaagccaca	1200
gcatacatcc	tgtccgtcca	agcagaggag	caaaagctca	tttctgaaga	ggacttgttg	1260
cggaaacgac	gagaacagtt	gaaacacaaa	cttgaacagc	tacggaactc	ttgtgcg	1317
<212> TYPE <213> ORGAI	: DNA NISM: Artif:	icial Seque	220			
<220> FEAT <223> OTHE poly	URE: R INFORMATI( nucleotide	DN: Descript	tion of Art:	ificial Sequ	lence: Synth	etic
<220> FEAT <223> OTHE poly <400> SEQU	URE: R INFORMATIC nucleotide ENCE: 11	DN: Descript	tion of Art:	ificial Sequ	uence: Synth	etic
<220> FEAT <223> OTHE poly <400> SEQU atgcccctca	URE: R INFORMATIO nucleotide ENCE: 11 acgttagctt	DN: Descript	tion of Art: aactatgacc	ificial Sequ tcgactacga	lence: Synth ctcggtgcag	etic 60
<220> FEAT <223> OTHEN polyn <400> SEQU atgcccctca ccgtatttct	URE: R INFORMATIO nucleotide ENCE: 11 acgttagctt actgcgacga	Caccaacagg ggaggagaac	aactatgacc	ificial Sequ tcgactacga agcagcagca	lence: Synth ctcggtgcag gagcgagctg	etic 60 120
<220> FEAT <223> OTHEN polyn <400> SEQU atgcccctca ccgtatttct cagcccccgg	URE: R INFORMATIC nucleotide ENCE: 11 acgttagctt actgcgacga cgcccagcga	Caccaacagg ggaggagaac ggatatctgg	aactatgacc ttctaccagc aagaaattcg	ificial Sequ tcgactacga agcagcagca agctgctgcc	ence: Synth ctcggtgcag gagcgagctg cgccccgccc	etic 60 120 180
<220> FEAT <223> OTHE polyn <400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta	URE: R INFORMATIO nucleotide ENCE: 11 acgttagctt actgcgacga cgcccagcga gccgccgctc	caccaacagg ggaggagaaac ggatatctgg cgggctctgc	aactatgacc ttctaccagc aagaaattcg tcgccctcct	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacaccttc	etic 60 120 180 240
<220> FEAT <223> OTHEN polyn <400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg	URE: R INFORMATIC nucleotide ENCE: 11 acgttagctt actgcgacga cgcccagcga gccgccgctc gagacaacga	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacacccttc ccagctggag	etic 60 120 180 240 300
<220> FEAT <223> OTHE polyn <400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg	URE: R INFORMATIC nucleotide ENCE: 11 acgttagctt actgcgacga cgcccagcga gccgccgctc gagacaacga agctgctggg	Caccaacagg ggaggagaaac ggatatctgg cgggctctgc cgggcggtggc aggagacatg	aactatgacc ttctaccagc aagaaattcg tcgcctcct gggagcttct gtgaaccaga	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacacccttc ccagctggag cgacccggac	etic 60 120 180 240 300 360
<220> FEAT <223> OTHE polyi <400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct	URE: R INFORMATIC nucleotide ENCE: 11 acgttagctt actgcgacga gccgccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc	etic 60 120 180 240 300 360 420
<220> FEAT 223> OTHEN polyn <400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc	URE: R INFORMATIC nucleotide ENCE: 11 acgttagctt actgcgacga cgcccagcga gagacaacga agctgctggg tcatcaaaaa tcgtctcaga	Caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacacccttc ccagctggag cgacccggac cttctcggcc agacagcggc	etic 60 120 180 240 300 360 420 480
<220> FEAT <223> OTHE poly <400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc	URE: R INFORMATIC nucleotide ENCE: 11 acgttagctt actgcgacga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga	caccaacagg ggaggagaaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat	etic 60 120 180 240 300 360 420 480 540
<220> FEAT 2223> OTHEN polyn <400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg	URE: R INFORMATIC nucleotide ENCE: 11 acgttagctt actgcgacga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcccgcgg	Caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac	aactatgacc ttctaccagc aagaaattcg tcgcctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct ccctcggtgg	tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacaccettc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac	etic 60 120 180 240 300 360 420 480 540 600
<220> FEAT <223> OTHE polyi <400> SEQU atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcgccg	URE: R INFORMATIO nucleotide ENCE: 11 acgttagctt actgcgacga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcctcaga agccgccdagt	Caccaacagg ggaggagaac ggatatctgg cggggtcttgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga tcctaccagg tgctccacct ccctcggtgg caagactcca	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacacccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac tccgtcctcg	etic 60 120 180 240 300 360 420 480 540 600 660
<pre>&lt;220&gt; FEAT 2223&gt; OTHE poly atgcccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcgccg gacagcagct gattctctgc</pre>	URE: R INFORMATIC nucleotide ENCE: 11 acgttagctt actgcgacga gccgccagcga gcgcccagcga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcccagcgg tcatcaaaaa tcgtctcaga	Caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac ctgcgcctcg ggagtcctcc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct ccctcggtgg caagactcca ccgcagggca	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc gccccgagcc	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac tccgtcctcg cctggtgctc	etic 60 120 180 240 300 360 420 480 540 600 660 720
<pre>&lt;220&gt; FEAT &lt;223&gt; OTHE poly atgccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc ctgagcgccg gacagcagct gattctctgc catgaggaga</pre>	URE: R INFORMATIO nucleotide ENCE: 11 acgttagctt actgcgacga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcctcaga cgcccaagtc tctcctcgac	Caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac ctgcgcctcg ggagtcctcc caccagcagc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga tcctaccagg tgctccacct ccctcggtgg caagactcca caggactcca gggagctcca	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc gccccgagcc aggaacaaga	ence: Synth ctcggtgcag gagcgagctg cgccccgccc cacaccttc ccagctggag cgaccggac cttctcggcc agacagcggc cctgcaggat ccctgcacgat ccctgtcctcg cctggtgctc agatgaggaa	etic 60 120 180 240 300 360 420 480 540 600 660 720 780
<pre>&lt;220&gt; FEAT &gt; 223&gt; OTHE poly atgccctca ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc agcccgaacc ctgagcgccg gacagcgccg gacagcagct gattctctgc catgaggaga gaaatcgatg</pre>	URE: R INFORMATIO nucleotide ENCE: 11 acgttagctt actgcgacga gccgccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccagcgg ccgcccagcgg ccgcccagcgg tctcctcgac tctcctcgac	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ccacagcgtc gtgcatcgac ctgcgcctcg ggagtcctcc caccagcagc	aactatgacc ttctaccagc aagaaattcg tcgccctcct gggagcttct gtgaaccaga caggactgta tcctaccagg tgctccacct ccctcggtgg caagactcca caggactcagg caagactcca	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gcgccttctc gccccgagcc aggaacaaga gcaaaaggtc	Lence: Synth ctcggtgcag gagcgagctg cgccccgccc cacaccttc ccagctggag cgaccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac tccgtcctcg agatgaggaa agagtctgga	etic 60 120 180 240 300 360 420 480 540 600 660 720 780 840
<pre>&lt;220&gt; FEAT &lt;223&gt; OTHE poly atgccccta ccgtatttct cagcccccgg ctgtccccta tcccttcggg atggtgaccg gacgagacct gccgccaagc cgccgaacc ctgagcgccg gacagcagct gattctctgc catgaggaga tcaccttctg</pre>	URE: R INFORMATIO nucleotide ENCE: 11 acgttagctt actgcgacga gccgccagcga gccgccgctc gagacaacga agctgctggg tcatcaaaaa tcgtctcaga ccgcccgcgg ccgcccagtc tctcctcgac tactcctcgac tagtttctgt ctggaggcca	caccaacagg ggaggagaac ggatatctgg cgggctctgc cggcggtggc aggagacatg catcatcatc gaagctggcc ctgcgcctcg gtgcatcgac ctgcgcctcg ggagtcctcc caccagcagc ggaaaagagg cagcaaacct	aactatgacc ttctaccagc aagaaattcg tcgcctcct gggagcttct gtgaaccaga tcctaccagg tgctccacct ccctcggtgg caagactcca gactctgagg caggctcctg caggctcctg	ificial Sequ tcgactacga agcagcagca agctgctgcc acgttgcggt ccacggccga gtttcatctg tgtggagcgg ctgcgcgcaa ccagcttgta tcttccccta gccccgagcc aggaacaaga gcaaaaggtc cactggtcct	Lence: Synth ctcggtgcag gagcgagctg cgccccgccc cacaccttc ccagctggag cgacccggac cttctcggcc agacagcggc cctgcaggat ccctctcaac tccgtcctcg cctggtgctc agatgaggaa agagtctgga	etic 60 120 180 240 300 360 420 480 540 600 660 720 780 840 900

81

<pre>grigocaaga gggtcaagt ggacagtto agagtota gaagtota gacagtoa gacacacaa 1000 grottggac gocaagagg gacagacta aacgagot ittitgoot gotgacaa 1200 grattgac gocaagag gacagacta igaaagaca titutgaag ggattgtig 1260 cggaaacgac gagaacagt gaaacacaa cttgaacago taggaact tittgoot gagttig 1260 cggaaacgac gagaacagt gaaacacaa cttgaacago taggaact tittgoot 1317 </pre>			-concinded	
aaatgoaca geeeagge eteggaeae gaggagat teaggage aacacaa 1000 geetagge geeaggag gaacgaget aaacgaget tittegeet gegaacga 1140 atcoeggat tggaaacga gagaacgt gaacacaa etgaacgg tatteetta aaageeaea 1200 geatacate tgteogteea ageaggaga caaagete tittegaag ggaettgt 1260 cggaaacga gagaacgt gaacacaa etgaacge taeggaete tigtegee 1317 *100 SBO ID NO 12 *101 SBO ID NO 12	gctgccaaga gggtcaagtt	ggacagtgtc agagtcctga	gacagatcag caacaaccga	1020
<pre>gtcttgggc gccagagg gaacaggt gaacaggt a aacggagt tttttgcct gcgtgaccag 1140 atcccgggt tggaacaca tgaaaggce cccaggtag ttateetta aaagecca 1200 gcatacatee tgtcogtees ageagaggg caaagecta tttedgaag ggattgtg 1260 cggaaacgae gagaacagt gaacacaa ettgacage taeggaete ttgtgga 1217 *210 tBMFN: 1117 *212 tMFN: N117 *212 tMFN:NN: Attificial Sequence: Synthetic polynucleotide *200 FEATURE: *200 EEQUENCE: 12 #tgccecta gcggegge ggaggaae ttgacacaca ggagaatgt gagaatgg ageagaetg gagaacagg aggaggaet teatecaga ageagagg 120 ccgdattet actegga ggaggagaat ttgacaga agetggg teacagge caeagecg 120 ccgdatect actereaga ggaggagaat teacacag gagaacagg ageaggag acaagetg teggegga caeagecg 120 ccgdatect actereaa a cateated caegagag gttategg ageagaagg caeagecg ageacagg caeagecg tegeereag coccaage ggaggagat teaceaa caegtagot teatecaga agetggg taeeagae caeagecg aceagegg tegeereag tegeereag coccaage ggagagatg tgtatecea caeagetg attateg ageatetg tegeerea agetggge 480 agecegaace tegeereag gagaacags coggeggteg ggagette ceaegecga caeagecga 480 agecegaace cegeereag caeagecg tegeerea agetggeg agaacaga agetggg agaacaaga agatggga agettete ceaegtgg agaacaga agatggga 480 agecegaace tegeereaga ggagetege tegterea agetggeer tegeerea agatggeg 480 agecegaace tegeereaga ggagetege tegterea agetggeg 480 agecegaace tegeereaga ggagetege ceeeggg agetetege tegeerea agatggeg 480 agecegaace tegeereaga ggagetege gegegeteg tegeerea agetgggg 480 agecegaace tegeereaga ggagetege ceeeggg agetege gegeerea agatggeg 480 agecegaace tegeereaga ggagetege ceeeggg agecegaa gagaacaga agatggga 480 agecegaace tegeereaga ggagetege ceeeggg agecega agatateg agettege 480 agaacaga caegece agaaaagg caegetege agaacaga agatggga 480 agecegaace tegeereaga ggagetege ceeeggg agaacaga agatggga 480 agaecagaa caegecea agaagagg caagetege agecegaa agataggaga 780 gaatagaga caegecea agaagagg caegetege agaaaateg agaacaga agatggga 480 acaetteg tegeereaga agatatega agatgega agaatagg agaacaeaga agatgagaa 780 gaatagaga caegeceaga caeageeg caeagege agaeagaga gaacaeaga agatgagaa 780 gaatagaga caegeceaga caeageega aga</pre>	aaatgcacca gccccaggtc	ctcggacacc gaggagaatg	tcaagaggcg aacacacaac	1080
atocoggigt tiggaaacaa tigaaaggoo oocaaggtag tiatootta aaaagooca 1200 goataoatoo tigtoogtoo agooggigg caaaagoto tittotgaag gigatottigt 1260 coggaaacgoo gagaacagt gaaacaaca ottgaacago taoggaacto tiggoo 1317 *210 500000000000000000000000000000000000	gtcttggagc gccagaggag	gaacgagcta aaacggagct	tttttgccct gcgtgaccag	1140
<pre>geatacated tigtoogtood ageagagag daaageted tittigaag ggadtigtig 1260 cggaaacgae gagaacgit gaaacacaa ottgaacago taeggaacte tigtoog 1317 &lt;1100 SEQ ID NO 12 *1110 LENNET: 1317 *1120 TTPE: DNA *1200 FEATURE: *1200 FEATURE: *1200 FEATURE: *1200 FEATURE: *1200 FEATURE: *1200 SEQUENCE: 12 atgecoected aegitagete caecacagg actatigae togacagae gaegeagetg 120 caegeoeceg ogeceagea ggatatetg aagaatedg acteggege daacacete 240 teeeteggg gagacaagt ggagagaag tittaeceag egaegeage caegeoggae 300 atggigaege gegeegegg oggegetig togeceece acteceage 420 gaegagaeet teatecaaa at caecace aggagetig tigtoog caeaceette 240 eceteggg gagacaagg oggegeggg ggagaette caegeogeag caeaceege 180 ecegeaagae tegtoogeage ggagagaet tetteaceag egaegeage caeaceette 240 teeeteggg gagacaaga oggeggggg ggagaette caegeogeag caeaceege 180 ecegeaagae tegtotaga gaagagaeg tegtoecea caegeogea 300 acggtagaeet teatecaaaa catecate caggaetiga tigtogaeeggae 300 acggaagaeet testecaaaa caecacete caegtogge tegeageegga 300 gaegagaaeet testecaaaa caecace caegaetiga tigtogaeeggae 420 geegeeaage tegtotaga gaagetage testecaeet caegtogae caeaceegee 420 gaegeagaeet testecaaaa catecae caggaetiga tigtogaeega agaegegge 420 gaegeagaeet testecaaaa caecaceete caegtogae geeceetee 720 catgagege caeceecee gaegeeteg gaegetee caeceegge geeceese 720 catgagega caeceecee caeceege gaegetee geeceetee caeceese 720 catgagaga caeceecee caeceege gaegetee geeceaage geeceetee testecaage agaetagae agaetagae 720 caeceesag geeceaage tegeecee caeceege gaegetee caeceege caeageteg 720 caeceesag geeceaage caecaege testecaae caecee caeageteg agaetaeaa agaetageg 700 caeceesag geeceaage tegeecee gaegeagete gaegaetaga gaagetegae 120 geeteaate tegeeceaa gaegeete gaegeatege teatecaae caeagetega 120 aateceaage geeceaage tegeecee gaegagaete titteee gaegaetegae 120 geeteaate tegeeceaa tegeaagege caeagetege titteeeta aaaageece 1200 geetacate tegeeceaa tegaaagege caeagetege titteeeta aaaageecea 1200 geetacate tegeeceaa tegaaagege caeagetege titteeeta aaaageecea 1200 geetacate tegeeceaa tegaaagege caeagetege titte</pre>	atcccggagt tggaaaacaa	tgaaaaggcc cccaaggtag	ttatccttaa aaaagccaca	1200
cggaaacga gggaacgt ggaacacgt gaaccaca ottgaacgt taggaact ttgtg 1317	gcatacatcc tgtccgtcca	agcagaggag caaaagctca	tttctgaaga ggacttgttg	1260
(210) SEQ ID NO 12 (211) LENGTH: 1317 (212) TURE: (212) TURE: (212) TURE: (212) CURES: 12 atgraded: acgutaget cacacadage acttaged ctogatage ctoggtaged for cocgetated acgutaged tedeadage gagegaged gagegaged gagegaged gagegaged gagegaged cacacaded gageded acgutaged cacacadage gagegaged cacacaded caggtaged acgutaged tedeadage cacageded gagegaged cacageded gagegaged cacageded gagegaged tedeadage cacageded cacagedege aggegaged tedeadage cacageded gagegaged tedeadage cacageded gagegaged tedeadage cacageded gagegaged tedeadage cacageded cacagedege dagegaged tedeadage cacageded cacagedege dagegaged tedeadage cacageded tedeadage cacageded cacagedege gagededed gagededeg gagededed gagededeg gagededed cacagedege gagededed cacagedege gagededed cacagedege gagededed gagededeg gagededed gagededed gagededeg gagededed gagededeg gagedeged gagegaged	cggaaacgac gagaacagtt g	gaaacacaaa cttgaacagc	tacggaactc ttgtgcg	1317
<ul> <li>&lt;400&gt; SEQUENCE: 12</li> <li>atgodddda acgitagott dacaaagg aadtatgad togadtadga dtoggtogag 120</li> <li>cogtatttt actgogada ggaggagaa tottatacag agaagaaga gaggagadg 120</li> <li>cagodddag ggagaaaaga togadaattog agaagtagg caacacatte 240</li> <li>toettogg gagaaaaga eggeggtog ggagatte coacgogag caacacatte 240</li> <li>atgotgaca agetgetgg agagaaatg gtgaacaag attoggaggg ettee coaggegg caacacate agagaggg ettee coaggegg ettee caggacaga gatategg agaacatg togacacaga togacagag togacacag togeceagg togeceagg togeceagg togeceagg togeceagg ettee coaggegg totecee agetgegg ettee coagetgeg ettee coagetge 1400</li> <li>agacagaage togeceage toggoetee togecetee caggetgg tottee coagetgg 480</li> <li>agacagaage togeceagg togacacag togeceag agacateg togeceeaga togeceagag 150</li> <li>cagaagagag togeceagg gagteete coecgaggg tottee togecee 1600</li> <li>gaaagagag togeceaga ggagteete coecgaggg agacatag agatgagga 3780</li> <li>gaaategatg ttytteetg ggaaaagg cageteetg caacaget agaagtegg agaetatee 360</li> <li>gaatacaga ggetaagt ggaaagg agagette aagaged agaagaag agaagagg 300</li> <li>cacgtee a caacagag agacagg aggetee agagedeg agaagatag agaagagg 300</li> <li>cacgtee a caacagaa togeagaeg togacaga agacagga agaagagag 300</li> <li>cacgtee a caacagaa togeagag agacetag agaacaga agaagagga 300</li> <li>cacgtee a caacagaa togaaaaga togacaga agacagag agaagagaga 3780</li> <li>gedacacaa togaagag togacaga togacagaa tocaacaga agatgegg 3000</li> <li>cacgtee a caacagaa togaaaaga togagaaatg togaagaacaga agaagaggag 300</li> <li>cacgtee a caacagaa togaaaaga togagaaatg togaagaaga agaacagaa 360</li> <li>gedacacaa togaagag gaagacaga tottee toegaagag agaacaga agaagagaga 360</li> <li>gedacacaa togaaaaga togaagaag tottee togaagaga agaacagaa 360</li> <li>gedaaacaa togaaagag caacagat tttttee togaagag agaetgetg 3140</li> <li>ateceggag togecegea agaacaga taaagacaa tttee agaaggeg togaacacaaca 3200</li> <li>gedaacacaa togaaaaga caacagaagat tttee taaaagagaa 3220</li> <li>gedaacaca togecegaagag taaaagacaa tttee agaagage 1220</li> <li>gedaacaca togecegaaga taaagacac</li></ul>	<210> SEQ ID NO 12 <211> LENGTH: 1317 <212> TYPE: DNA <213> ORGANISM: Artific <220> FEATURE: <223> OTHER INFORMATION polynucleotide	cial Sequence N: Description of Art	ificial Sequence: Synthe	tic
atgccctcaacgttagctcaccacacgaacatatgactegactacgacagcagcogtattttactgcgcagaggagagaattetaccagaagcggagagagag120cagccccaggcgcccagagaggagatatetgaagaaattegagctgetgeecacccecce180ctggcccctagccgccgctecgggggggggggggggggggggggggggggggggggg	<400> SEQUENCE: 12			
ccqtattttactqcqcqagqqqqqqatttaccqqaqqqcqqqqqqq120cqqccccqqcqcqcqqqgqqqqqqqqqqqqqqq180ctqqccccqqcqcqcqqqcqqqqtqqqqqqqqqq240tcccttqqqqqqqqqqqqqqqqqqqqqqqqq300atqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq360atqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq360atqqqqqqqqqqqqqqqqqqqqqqcqqqqqqq360atqqqqqqqqqqqqqqqqqqqqqqcqqqqqqq460gcqcqaqqqtqtqqqqqqcqqqqqqqtttactqqqqqqqqqgcqqqqqqcqqqqqqqqqqqqqqqtttactqqqqqqqqqqgcqqqqqqcqqqqqqqtttactqqcqqqqqqq480aqcqqqqqcqqqqqqqtqtqqqqqtttactqqcqqqqqqqgqqqqqqcqqqqqqtqtqqqqtqtqqqqfqqqqqqgqqqqqqqcqqqqqqqtqtqqqqtqtqqqqfqqqqqqgqqqqqqqcqqqqqqqtqtqqqqcqqqqqqqfqqqqqqgqqqqqqqcqqqqqqqtqtqqqqcqqqqqqqfqqqqqqgqqqqqqqcqqqqqqqtqtqqqqcqqqqqqqqqqqfqqqqqqqgqqqqqqqcqqqqqqqtqtqqqqcqqqqqqqqqqqqqqqfqqqqqqqgqqqqqqqcqqqqqqqtqtqqqqcqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq	atgcccctca acgttagctt	caccaacagg aactatgacc	tcgactacga ctcggtgcag	60
cagececcagegececageeggatatetgagaaatetgagetgetgetcagecegete180ctggecectageegeegeteegggetetgtegecectetacagttgeggcagecegete240tecettegggagacaacgaegggetgggggagettetccagegeg300360atggtgaceteatetaaacatetatetcaggactgagtgagegg360geegeagaeteatetaaacatetatetcaggactgagtgagegg360geegeagaeteatetaaacatetatetcaggactgagtgagegg360geegeagaeteatetaaacatetatetcaggactgagtgegegg360geegeagaeteatetaaacatetatetcaggactgagtgegegg360geegeagaeteatetaaacatetatetcaggactga360360geegeagaeteatetaaacatetatetcaggactga340360geegeagaeteatetaaggaacteggctetaceagtegeegeag480agecegaaccegeeceggceacageggtegteceatcedeetaag540ctgagegeategeecegagegeetegcaggetegcettegg660gattetegtecectaggggateteggegeecegacedeetagg780gaaategagtegtetegcacacagegagetetegaagteteg900cadgettetegtegaceacageagagggageteteg960cadgeagaeggacaggggagegaegaegaggaegaegae1020aaatgeacegeceaggggaacgaegaegaggaegaegaeaaegetatategeagaegaegaacgaegae <td>ccgtatttct actgcgacga</td> <td>ggaggagaac ttctaccagc</td> <td>agcagcagca gagcgagctg</td> <td>120</td>	ccgtatttct actgcgacga	ggaggagaac ttctaccagc	agcagcagca gagcgagctg	120
ctggccctagcgcgcgtggggttggtggcgctctccacggcgccaccgttgg300atggtgacggagacaacgcggcggtgggggagcttccacgggcgccacggggg360gacgagaccttcatcaaaacatcatcatccaggactggcttccggc420gccgcaagctogtctagagaagtgggctatcacagctgcggggg480agccggaaccogcgcggccacgcgggcttccggc480agccgaaccogcgccaggtgactggtttcccctacctgcaggat540ctgagcgcgcogccaagggggatctgcttccggtgtttccccta600gaagagagatgggatctcccogcgggggggatctgcttccgdc660gattcttgttcccctagggatcaggggatctcccogcaggg780gaaacgaggcaccaccacggaacacagggaacacagggatctcg900cacgttctacacatacgcacacatacgcagcgccccag900cacgttccacacatacgcagcgcggg acctcgcacatacacag1020aaatgcacagccaggggaacagggaaaggggaaaggggaaaggagggtttgggggccagggggaacagggaaaggggaaaggagg1020aaatgcacagccggagggaacagggccaaggggaaaaggca1010gtttggggccaggggttttgcctgcggaccag1140atccggaggaacaggcacaagggcacaaggca1200gctacacttgtcgtcaaaaggggcacaggggaaaggcag1200gctggagggccagggggaacagggcacaagggg1210gctacaggga	cageceeegg egeceagega	ggatatctgg aagaaattcg	agetgetgee caeceegeee	180
teccttegggagacaacgaeggeggtgggggggttttccacggccgaccacggccgaccacggtgggg300atggtgaccactgttggaggagactggtgaacaaggtgaacagagtttattgcgacccgga360gacgagacctattaaaacatatatacaggactgatgtggagggcttatggg420gccgcaagatottacaaagaagtggtottacaagctggggggtdtotggg480agacggacacogcccgagcacacggtctgtgacgcttacaaa540ctgagcgcgcogccaagagtgcatcgacectcggggccaggttg540gaaagacagtcgcccaagagtgcatcgacectcgggggcectgtagg660gattettegcaccgcccagaggtettecogcaggggggatetteceggagga780gaaatcggggaaaagagcaggetceceatcaagaagagtetgg900cacgttetatatttegggacaggtaggaacaagaggatetee960gttggggcacadagaggaacggggacaaagagggaacagagg900cacgtetcacacatacgagagtcetcacacatacgaggatetee960gttggggcacadagaggaacggggacaagagegacaacacaga1020aaatgcacagoccaggggaacggggagaacggggacaacacaca1080gttttggaggccagggggaacggggacaaggegacaacacaca1080gttttggaggccagggggaacggggacaaggggacaacacaca1080gttttggaggccagggggaacggggacaaggegacaacacaca1080gttttggaggaacaggggaa	ctggccccta gccgccgctc	cgggetetge tegeceteet	acgttgcggt cacacccttc	240
atgggacceagetgetgggaggagacatggtgaaccagagtttcatctgcgaccegga360gacgagaccttcatcaaaacatcatcatccaggactgatgtggagegcttctoggc420gccgcaaactcgtctcagagaactggactcetaccaggctgtcggacaaagaccggac480agccggaacccgcccaggccacacggttgtctcacceccagcttgactgcagagat540ctgagogcgccgccaagatgtgacatggaccctcggggtcetcccac600gacagcagatcgcccaagtctggoccgcaggactgacccgcaggga660gattetetgtetececacgaggtetgeccgcagggaggcccgagga720catgaggagacaccgcccaccaccacgacggactetgggagaccagaaggtetgggaacagagattgtttetgggaaaagggcaccacgacgaggtetgg900cacgtetgagggetcaagtgagacagggacaggaggaacacaccg960gttgtggaccacatcagcacacatcagcagagtetggaacacaccg1020aaatgcacagggetcaagtgagacggggacaggaggaacacaccg1020aaatgcacagccgaaggggacaggggcacaaggggaacacacac1080gtettggaggccgaaggggaacagggcacaaggggtattettaaaaggcaca1200gcatacatetgtcggagatattettaaaaggcaca12001260gcaaacgggaacagggcacaaggggcacaaggggaatgtg1260ggaacaggggaacaggcacaaggagtattettaaaaggcaca1200gcaaacgactgtgaagaacgac <td< td=""><td>tcccttcggg gagacaacga</td><td>cggcggtggc gggagcttct</td><td>ccacggccga ccagctggag</td><td>300</td></td<>	tcccttcggg gagacaacga	cggcggtggc gggagcttct	ccacggccga ccagctggag	300
gaegagacet teateaaaa eateateatecaggactgta tgtgagegg etteteggee420geegagacet teateaaaa eateateateteateacaagetgegegeaaagaeceggae480ageeegaace eegeeeggg eeacagegtetgteecaectceagettgta eetgeaggat540etgagegeeg eegeetaagaetgegeetaeceteggegg660gaateetegeetgegeetaeegeeteegg660gatteetegeetgegeetagegeeteeta620gatteetegeetgegeetagegeeteeta620gatteetegeetgegeetagegeeteeta620gaategagaga eacegeetaetgegeetagegeeteeta720catgaggagaetgegeetagaateetagggegeeteeta780gaategaget tegegeetaetgegeetagegeeteetaagaggggg900cacetteetegetgegaetagaaggaetagggaateateeta960getgeetaagaggeeteetagaaggaetaggagaaggeeta960getgeetaagagggeetaagtgaagaetggagaaggeeta960getgeetaagagggeetaagtgaagaetggagaaggeeta960getgeetaagagggeetaagtgaagaetggagaaggeeta1020aaatgeacageecetagagaaggeetagaaggeetagaaggeeta1080getetteggaggeaagagegaaggeetatettegaaggeeta1140ateceggagttggaaacaattgaaaggeceaaggeetatettegaag1260ggaaacagtgaaacagaaetagaaggegaacttettetettegaet1260gegaaacagtgaaacacaaetagaagetacggaacttettetgeteeta1317	atggtgaccg agctgctggg	aggagacatg gtgaaccaga	gtttcatctg cgacccggac	360
gccgccaagetcgtctcagagaagctggcetcctaccaggctgcgcgcaaagacagcgge480agcccgaaceccgcccgggccaagcggetgctccaceccdgcggga540ctgagcgcaccgcctcagagtgcatcgaeccctcggggtcttcccca660gacagcagecgcccaagegagtcctceccggaggaeccggtggee720catgaggagcaccgccagegagtcctceccgcagggegagtcggaa780gaaatcgatgttgtttctgggaaaagggcaccagcagegagtctcecaagagte900cacgtttecacatcageacactacageaggecctececaagagte900cacgttccacacatcageagagtcctcacactacggaaggactace960gettctggagggcccaggggaacagggaagacacacaa1020gaatcggagggccagggggaacagggagaacgggga1140atcccggagtggcggaaggagcacaagggagtttttgccegggactgggcatacatetgtccgtcaagcaggggcaaaaggetttttgcce1200gcatacatetgtccgtcaagcaggggcaaaaggetttttgcce1200gcatacatetgtccgtcaagcaggggcaaaggetttttgcgg1317	gacgagacct tcatcaaaaa	catcatcatc caggactgta	tgtggagcgg cttctcggcc	420
agecegaacecegecegagceacaqegtetgetecacetceagettgtacettgagagat540ctgagegeegcegeceaagtectgegeetegcaagacteageceetegaa600gatatetegceceetegaeggagteetegcaagacteageceetegage720catgaggagcacegeeceacaceagegegatettgggggaacagaga780gaaategatgttgtttetgggaaaagagcagegeetegcaagaggeg900cacettegctggaggeecaceacageageceetegagggaatacaga960gettetegaaggacagtgggaacagteggaagteeteg960gettetgaagegggeteaggegaagteeteggaagaggg1020aaatgeacegeceagggggaacagaaggaaggagaa1020aaatgeacegeceagggggaacagagggaaggagaa1140ateceeggagtggeegeececaaggegcecaaggeg1260cggaaacgetgtecgteeageaggaggcaaaggeetettggage1260ggaaacgaegaacagttgaacagaetttetgaagggaettgteg1317	geegeeaage tegteteaga	gaagetggee teetaecagg	ctgcgcgcaa agacagcggc	480
ctgagogocgcogecteaggtgategaececteggtggtettecetae600gacageagetcgeceaagtectgegeetegcaagaeteeagegeetteeteegteeteggatetetegteteetegagggagteeteecegeagggaageeeeggagaareegaaategaagtegttetegggaaaagagcaggeeteegggagteeteggaaaaggegaaategatgtegtgaggeeacaecaageaggaaaaggeggagaecaagagagtetgg840teacettegctggaggeeacaecaageaggegeeteececaagaggtgg900caegteteeacaeateageaggacagteeceeteggagaaggaetaetee960getgeeaaggggteaagtggacagteegegeeteecegegeeteece960getgeeaaggggteaagtggacagteegaggeeteeggagaagaetaetee960getgeeaaggggteaagteggacagteegageeteeggagaateeggagaagaetaetee960getgeeaaggggteaagtegagacagteegaggagaateggagaagaetaeteeggaga1020aaatgeaceaggecaggeggaaaggagatteaaggaggeaaacacaaca1080gtetteggaggeaaaaggececaaggtegtetteecetgegteegga1140ateceggagtggaaacaategaaaggecaaaagetea <tttettegaag< td="">ggaettgge1260cggaaacgaegagaacagtgaaacacaatetteggaggaettgge1317</tttettegaag<>	agecegaace eegeeegegg	ccacagegte tgetecaeet	ccagcttgta cctgcaggat	540
gacagcagetcgecccaagtectgegectegcaagactecagegectteteteogtecteg660gattetetgeteteetegaggagteetecegeagggageccegaggeecetggggga720catgaggagacacegeccacacaceagcagegactetgaggaggaacaagaagatgggggaa780gaaategatgttgtttetgggaaaagaggcaggeetegggecaaaggteagagtetggga840teacettegctggaggeacaceacaacceteacageacaceacgeaggegeetegg900cacgteteeacacacacageaggacagtgegegeetegg960getgeccaagagggtecagtggacagtgeggacagtetagaggeetegg960getgeccaagagggtecagtggacagtgeggacagtetagaggeetegg960getgeccaagggggtecagtggacagtgegaggeggaatgteacagaggeg1020aaatgeaceageccaggggagaacgagetaaaacggagetteateggaggaa1080gtettggaggecagggaggaacgagetaaaacggagettttttgeetegggaggaa1200gcatacatetggaaacaatgaaaaggeccaaaggeg gaactatet1260ggaaacgaegagaacagtgaaacaaattttegaagaggacttgteg1317	ctgagegeeg eegeeteaga	gtgcatcgac ccctcggtgg	tetteeeeta eeeteteaae	600
gattetetgateteetegaaggagteeteecegeagggaageecegaggeaceteggaggaa720catgaggagacacegeceaacaceagcagagaatetggagaaggagaacaagaagagtaggagaa780gaaategatgttgtttetgggaaaagaggcaggeteetggeaaaaggteagagtetggaa840teacettetgctggaggecacagcaaacetcetcacageecacteggaaggactatetggaa900cacgteteecacateageacaactaegeaggagteetegceacteggaaggactateet960getgeceaagagggteaagttggacagtgteagagteetggacagategcaacaacega1020aaatgeaceageceagggaggaacagagetteaaggagegaaacagaacaa1080gtettggaggecagaggaggaacagagtececaaggagetettetgaaga1200geaacactetggaaaacaatgaaaaggececaaggtagtettetgaaga1260cggaaacgacgagaacagtgaaacacaatettgaacg1317	gacagcagct cgcccaagtc	ctgcgcctcg caagactcca	gcgccttctc tccgtcctcg	660
catgaggagacaccgcccaccaccagcagcgactctgaggaggaacaagaagattgaggaga780gaaatcgatgttgtttctgggaaaagggcaggctctggcaaaaggtcagagttgga840tcaccttctgctggaggccacagcaaacctcctcacagcecactggtcctcaagaggtgc900cacgtctccacacatcagcaggacagtgggggcctccctccactcggaaggactatcct960gctgccaagagggtcaagttggacagtgcagagtctgagacagatcagcaacaaccga1020aaatgcaccagccccaggtcctcggacaccgaggagaattcacagaggcaacacacacac1080gtcttggaggccagaggaggaacaggcagttttgccctgcgtgccacc1200gcatacatcctggaaaacaatgaaaaggccccaaggtagtttctgaagaggacttgttgggaaacagatggaaacaatgaaaggctttctgaagaggacttgttg1260cggaaacgacgagaacagtgaaacacaattttgccctttgtgcg1317	gattetetge teteetegae	ggagteetee eegeagggea	gccccgagcc cctggtgctc	720
gaaatcgatgttgtttctgtggaaaagaggcaggctcctggcaaaggtcagagtctgga840tcaccttctgctggaggcacagcaaacctcctcacagcecactggtcctcaagaggtgc900cacgtctccacacatcagcacaactacgcagggctcacctccactcggaaggactatcct960gctgccaagagggtcaagttggacagtgtcagagtcctgagacagatcagcaacaaccga1020aaatgcaccagcccaggtgctcggacaccgaggagatgtcaagaggcgaacacacacaa1080gtcttggagcgccagaggaggaacgagctaaaacggagctttttgccctgcgtgaccag1140atcccggagttggaaacaatgaaaggggcaaaagctcatttctgaagaggacttgttg1260cggaaacgacgagaacagttgaaacacaaatttgaacagttgtgcg1317	catgaggaga caccgcccac	caccagcagc gactctgagg	aggaacaaga agatgaggaa	780
tcaccttctgctggaggccacagcaaacctcctcacagcecactggtctcaagaggtgc900cacgtctccacaactacgcacggcctccctccactcggaaggactatcct960gctgccaagagggtcaagttggacagtgteagagtcctgagacagatcagcaacaaccga1020aaatgcaccagccccaggtectcggacacegaggagaatgtcaagaggggaacacaacacaa1080gtcttggagegccagaggaggaacgagctaaaacggagetttttgccctgcgtgaccaa1200gcatacatcetgtccgtccaagcagaggagccaaaggtcatttctgaagagacttgttg1260cggaaacgacgagaacagttgaaacacaaacttgaacagetacggaactttgtgcg1317	gaaatcgatg ttgtttctgt	ggaaaagagg caggctcctg	gcaaaaggtc agagtctgga	840
cacgteteeacaactacgeacaactacgeagegeeteeteceacteggaaggaetateet960getgeeaagagggteaagttggaeagtegagagteetgagaeagateagcaacaacega1020aaatgeaceageeeeagaggageteggaeaegaggagaatgteaagaggegaacacaaceaa1080gtettggagegeeeagaggaggaacggaetaaaaeggagetttttgeeetgegtgaeaga1140ateeeggagttggaaaacaatgaaaaggeeceeaaggtagttateettaaaaaageeae1200geatacateetgteegteeaageagaggagcaaaageetattetegaagaggaettgtg1260eggaaacgaegagaacagttgaaacacaaaeteggaacettetegeeg1317	tcaccttctg ctggaggcca	cagcaaacct cctcacagcc	cactggtcct caagaggtgc	900
gctgccaagagggtcaagttggacagtgtcagagtcctgagacagatcagcaacaaccga1020aaatgcaccagccccaggtcctcggacaccgaggagaatgtcaagaggcgaacacacacac1080gtcttggagcgccagaggaggaacgagctaaaacggagctttttgccctgcgtgaccag1140atcccggagttggaaaacaatgaaaaggcccccaaggtagttatccttaaaaaagccaca1200gcatacatcctgtccgtccaagcagaggagcaaaagctcatttctgaagaggacttgttg1260cggaaacgacgagaacagttgaaacacaaacttgaacagctacggaactcttgtgcg1317	cacgteteca cacateagea	caactacgca gcgcctccct	ccactcggaa ggactatcct	960
aaatgcacca geeecaaggte eteggacace gaggagaatg teaagaggeg aacacacaac 1080 gtettggage geeagaggag gaacgageta aaaeggaget tttttgeeet gegtgaceag 1140 ateeeggagt tggaaaacaa tgaaaaggee eecaaggtag ttateettaa aaaageeaca 1200 geatacatee tgteegteea ageagaggag caaaagetea tttetgaaga ggaettgttg 1260 eggaaaegae gagaacagtt gaaacacaaa ettgaacage taeggaaete ttgtgeeg 1317	gctgccaaga gggtcaagtt g	ggacagtgtc agagtcctga	gacagatcag caacaaccga	1020
gtettggage geeagaggag gaacgageta aaacggaget tttttgeeet gegtgaeeag 1140 ateeeggagt tggaaaacaa tgaaaaggee eecaaggtag ttateettaa aaaageeaca 1200 geataeatee tgteegteea ageagaggag eaaaagetea tttetgaaga ggaettgttg 1260 eggaaaegae gagaacagtt gaaacacaaa ettgaaeage taeggaaete ttgtgeeg 1317	aaatgcacca gccccaggtc	ctcggacacc gaggagaatg	tcaagaggcg aacacacaac	1080
atcccggagt tggaaaacaa tgaaaaggcc cccaaggtag ttatccttaa aaaagccaca 1200 gcatacatcc tgtccgtcca agcagaggag caaaagctca tttctgaaga ggacttgttg 1260 cggaaacgac gagaacagtt gaaacacaaa cttgaacagc tacggaactc ttgtgcg 1317	gtcttggagc gccagaggag	gaacgagcta aaacggagct	tttttgccct gcgtgaccag	1140
gcatacatee tgteegteea ageagaggag caaaagetea tttetgaaga ggaettgttg 1260 eggaaaegae gagaaeagtt gaaaeaeaa ettgaaeage taeggaaete ttgtgeg 1317	atcccggagt tggaaaacaa	tgaaaaggcc cccaaggtag	ttatccttaa aaaagccaca	1200
cggaaacgac gagaacagtt gaaacacaaa cttgaacagc tacggaactc ttgtgcg 1317	gcatacatcc tgtccgtcca	agcagaggag caaaagctca	tttctgaaga ggacttgttg	1260
	cggaaacgac gagaacagtt	gaaacacaaa cttgaacagc	tacggaactc ttgtgcg	1317

<210> SEQ ID NO 13 <211> LENGTH: 1086 <212> TYPE: DNA

<213> ORGANISM: Artificial Sequence <220> FEATURE:

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

<400> SEQUE	ENCE: 13					
atggcgactg	cggagacagc	acttccatca	atctcaacac	tcactgcact	ggggccattt	60
ccagataccc	aggacgattt	ccttaagtgg	tggcggtccg	aagaggctca	agacatggga	120
cctggtccgc	cggatcccac	cgaacctcct	ctgcatgtca	aaagtgaaga	tcagcctggc	180
gaggaagagg	atgacgaaag	gggtgccgac	gccacttggg	acttggatct	tctccttacc	240
aatttctctg	gtccggaacc	tggcggggca	ccacagacgt	gcgctctcgc	tccctcagaa	300
gcgagcgggg	ctcagtaccc	accccctccc	gaaactctgg	gagcctatgc	tgggggtcct	360
ggactggtgg	ctgggttgct	tggtagtgag	gaccattctg	gctgggtacg	ccccgctttg	420
agggcccgcg	ctccggacgc	ctttgtggga	ccggcgctcg	ctcctgcacc	ggctccggaa	480
ccaaaagccc	tcgcgctgca	gcccgtgtac	cccggacccg	gagccggatc	ctcagggggga	540
tacttcccac	ggaccggact	cagcgttcca	gcggcttccg	gggcgccata	cggattgttg	600
agcggctacc	cggctatgta	tcccgctccc	cagtaccaag	gacacttcca	attgttccgg	660
ggtcttcaag	ggcctgcgcc	cgggcctgct	accagtccca	gtttcctcag	ttgtctggga	720
ccgggaactg	ttggcactgg	acttggcggg	actgcagagg	acccaggcgt	tatagcagag	780
acagcgccaa	gtaaaagggg	ccgacgaagc	tgggccagga	aacgccaagc	tgcgcacact	840
tgtgcccatc	caggttgcgg	taaatcctac	acgaagagca	gtcatcttaa	agcacatctt	900
cgcacacaca	cgggcgagaa	gccctacgcc	tgtacttggg	aaggttgcgg	ctggagattc	960
gctagatctg	acgagctcac	ccggcattat	cgaaaacaca	ctggccagcg	accgttccgg	1020
tgccaactct	gcccaagggc	gttcagtcgc	tcagatcatc	tggctttgca	tatgaagcga	1080
cacctt						1086
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEAT &lt;223&gt; OTHEI polyn</pre>	ID NO 14 FH: 1065 : DNA NISM: Artif: JRE: R INFORMATI( nucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	lence: Synth	1086 etic
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG &lt;212&gt; TYPE &lt;212&gt; TYPE &lt;220&gt; FEATU &lt;220&gt; FEATU &lt;223&gt; OTHEI polyn &lt;400&gt; SEQUI</pre>	ID NO 14 TH: 1065 : DNA VISM: Artif: JRE: & INFORMATI( hucleotide ENCE: 14	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	ience: Synth	1086 etic
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG &lt;211&gt; TYGE &lt;213&gt; TYGE &lt;213&gt; ORGAI &lt;220&gt; FEATU &lt;223&gt; OTHEH polyn &lt;400&gt; SEQUI atggccctta</pre>	ID NO 14 TH: 1065 : DNA NISM: Artif: RE: R INFORMATIC NuCleotide ENCE: 14 gtgaacccat	icial Sequer DN: Descript tcttcccagc	nce tion of Art: ttttccacgt	ificial Sequ tcgcgtctcc	lence: Synth ttgccgagag	1086 etic 60
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEAT &lt;223&gt; OTHEI polyi &lt;400&gt; SEQUI atggccctta agaggccttc</pre>	ID NO 14 IH: 1065 : DNA VISM: Artif: JRE: R INFORMATIC hucleotide ENCE: 14 gtgaacccat aggaaaggtg	icial Sequer DN: Descript tcttcccagc gccgagggct	nce tion of Art: ttttccacgt gaacccgagt	ificial Sequ tcgcgtctcc ctggaggtac	ience: Synth ttgccgagag ggatgatgat	1086 etic 60 120
<pre>cacctt &lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAN &lt;220&gt; FEATU &lt;223&gt; OTHEN polyn &lt;400&gt; SEQUI atggccctta agaggccttc cttaacagtg</pre>	ID NO 14 TH: 1065 : DNA VISM: Artif: JRE: R INFORMATIC uucleotide ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt	icial Sequer DN: Descript tcttcccagc gccgagggct catactctca	nce tion of Art: ttttccacgt gaacccgagt atgggactgg	ificial Sequ tcgcgtctcc ctggaggtac acgggctggg	lence: Synth ttgccgagag ggatgatgat agcggaggca	1086 etic 60 120 180
<pre>cacctt &lt;210&gt; SEQ : &lt;211&gt; LENG: &lt;211&gt; LENG: &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEAT &lt;223&gt; OTHEI polyi &lt;400&gt; SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac</pre>	ID NO 14 IH: 1065 : DNA VISM: Artif: RE: R INFORMATIC nucleotide ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt caccaccacc	icial Sequer DN: Descript tcttcccagc gccgagggct catactctca ccctccgccc	nce tion of Art: ttttccacgt gaacccgagt atgggactgg ccagcgtttt	ificial Sequ tcgcgtctcc ctggaggtac acgggctggg actacccgga	tence: Synth ttgccgagag ggatgatgat agcggaggca gccaggtgcg	1086 etic 60 120 180 240
<pre>cacctt &lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;212&gt; TYPE &lt;220&gt; FEAT &lt;220&gt; FEAT c223&gt; OTHEI polyn &lt;400&gt; SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac ccgccgccat</pre>	ID NO 14 IH: 1065 : DNA UISM: Artif: JRE: A INFORMATI( bucleotide ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt caccaccacc attcagcccc	icial Sequer DN: Descript tottoccago googagggot catactotca cootcogoco ggogggtggo	nce tion of Art: ttttccacgt gaacccgagt atgggactgg ccagcgtttt ttggtgtccg	ificial Sequ tcgcgtctcc ctggaggtac acgggctggg actacccgga agctcctccg	lence: Synth ttgccgagag ggatgatgat agcggaggca gccaggtgcg gcctgaattg	1086 etic 60 120 180 240 300
<pre>cacctt &lt;210&gt; SEQ : &lt;211&gt; LENG; &lt;211&gt; LENG; &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEAT &lt;223&gt; OTHEI polyi &lt;400&gt; SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac ccgccgccat gatgccccgc</pre>	ID NO 14 IH: 1065 : DNA NISM: Artif: RE: R INFORMATIC bucleotide ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt caccaccacc attcagcccc tcggcccggc	icial Sequer DN: Descript tottoccago googagggot catactotca cootcogoco ggogggtggo gotgcatggt	nce tion of Art: gaacccgagt atgggactgg ccagcgtttt ttggtgtccg agatttctgc	ificial Sequ tcgcgtctcc ctggaggtac acgggctggg actacccgga agctcctccg tcgcgcctcc	ttgccgagag ggatgatgat agcggaggca gccaggtgcg gcctgaattg gggtcgactc	1086 etic 60 120 180 240 300 360
<pre>cacctt &lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEAT &lt;223&gt; OTHEI polyi &lt;400&gt; SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac ccgccgccat gatgccccgc gttaaggctg</pre>	ID NO 14 IH: 1065 : DNA VISM: Artif: JRE: 2 INFORMATIC bucleotide ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt caccaccacc attcagcccc tcggcccggc aacctcctga	icial Sequer DN: Descript tcttcccagc gccgagggct catactctca coctccgccc ggcgggtggc gctgcatggt ggctgatggt	nce tion of Art: gaacccgagt atgggactgg ccagcgtttt ttggtgtccg agatttctgc ggaggtggct	ificial Seq tcgcgtctcc ctggaggtac acgggctggg actacccgga agctcctccg tcgcgcctcc acggatgtgc	lence: Synth ttgccgagag ggatgatgat agcggaggca gccaggtgcg gcctgaattg gggtcgactc ccccgggctt	1086 etic 60 120 180 240 300 360 420
<pre>cacctt &lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGA' &lt;220&gt; FEATI &lt;223&gt; OTHEN polyn &lt;400&gt; SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac ccgccgccat gatgccccgc gttaaggctg acccgaggac</pre>	ID NO 14 FH: 1065 : DNA VISM: Artif: RE: RINFORMATIC bucleotide ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt caccaccacc atcagccccgc aacctcctga cgagaggtct	icial Sequer DN: Descript tottoccago googagggot catactotca cootcogoco ggogggtggo gotgcatggt ggotgatggt taagogggaa	nce tion of Art: gaacccgagt atgggactgg ccagcgtttt ttggtgtccg agatttctgc ggaggtggct ggggcacctg	ificial Sequ tcgcgtctcc ctggaggtac acgggctggg actacccgga agctcctccg tcgcgcctcc acggatgtgc gcccggctgc	ttgccgagag ggatgatgat agcggaggca gccaggtgcg gcctgaattg gggtcgactc ccccgggctt aagctgtatg	1086 etic 60 120 180 240 300 360 420 480
cacctt <210> SEQ : <211> LENG <212> TYPE <213> ORGAI <220> FEAT <223> OTHEI polyn <400> SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac ccgccgccat gatgccccgc gttaaggctg acccgaggac cgggggcccg	ID NO 14 IH: 1065 : DNA VISM: Artif: RE: RINFORMATIC bucleotide ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt caccaccacc attcagcccc tcggcccggc aacctcctga cgagaggtct gtgggaggcc	icial Sequer DN: Descript tottoccago googagggot catactotca cootcogoco googggtggo gotgcatggt ggotgatggt taagogggaa tococcogoco	nce tion of Art: ttttccacgt gaacccgagt atgggactgg ccagcgtttt ttggtgtccg agatttctgc ggaggtggct ggggcacctg cctgatacac	ificial Sequ tcgcgtctcc ctggaggtac acgggctggg actacccgga agctcctccg tcgcggctgc gcccggctgc cccccttag	ttgccgagag ggatgatgat agcggaggca gccaggtgcg gcctgaattg gggtcgactc ccccgggctt aagctgtatg tccagatgga	1086 etic 60 120 180 240 300 360 420 480 540
cacctt <210> SEQ : <211> LENG <212> TYPE <213> ORGAI <220> FEATU <220> FEATU <220> SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac ccgccgccat gatgccccgc gttaaggctg acccgaggac cgggggcccg ccagctcgac	ID NO 14 FH: 1065 : DNA VISM: Artif: RE: RINFORMATIC DUCLEOTIDE ENCE: 14 gtgaaaccat aggaaaggtg tgctcgattt caccaccacc attcagcccc tcggcccggc aacctcctga cgagaggtct gtgggaggcc ttcccgcacc	icial Sequer DN: Descript tcttcccagc gccgagggct catactctca ccctccgccc ggcgggtggc gctgcatggt taagcgggaa tcccccgccc tggccccaga	nce tion of Art: ttttccacgt gaacccgagt atgggactgg ccagcgtttt ttggtgtccg agatttctgc ggaggtggct ggggcacctg cctgatacac gcgagtttcc	ificial Seq tcgcgtctcc ctggaggtac acgggctggg actacccgga agctcctccg tcgcgcctcc acggatgtgc gcccggctgc cccccttag cccccttag	lence: Synth ttgccgagag ggatgatgat agcggaggca gcctgaattg gggtcgactc ccccgggctt aagctgtatg tccagatgga tggaggaccg	1086 etic 60 120 180 240 300 360 420 480 540 600
cacctt <210> SEQ : <211> LENG; <212> TYPE <213> ORGAI <220> FEAT <223> OTHEN polyn <400> SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac ccgccgccat gatgccccgc gttaaggctg acccgaggac cgggggcccg ccagctcgac gggtttggcg	ID NO 14 IH: 1065 : DNA VISM: Artif: RE: RINFORMATIC DUCLEOTIDE ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt caccaccacc tcggcccggc aacctcctga cgagaggtct gtgggaggcc ttcccgcacc ccccaggtcc	icial Sequer DN: Descript tottoccago googagggot catactotca cootcogoco ggogggtggo gotgatggt taagogggaa tococcogoco tggocccaga tggacttac	nce tion of Art: ttttccacgt gaacccgagt atgggactgg ccagcgtttt ttggtgtccg ggaggtggct gggggcacctg cctgatacac gcgagtttcc tacgccctc	ificial Sequ tcgcgtctcc ctggaggtac acgggctggg actacccgga agctcctccg tcgcgcctcc acggatgtgc gcccggctgc ccccccttag cccctccatt	ttgccgagag ggatgatgat agcggaggca gccaggtgcg gcctgaattg gggtcgactc ccccgggctt aagctgtatg tccagatgga tggaggaccg agcttttggt	1086 etic 60 120 180 240 300 360 420 480 540 600 660
cacctt <210> SEQ : <211> LENG' <212> TYPE <213> ORGAI <220> FEATU <220> FEATU <220> SEQUI atggccctta agaggccttc cttaacagtg gctcctgaac cgccgccat gatgccccgc gttaaggctg acccgaggac cgggggcccg ccagctcgac gggtttggcg ctttcgacg	ID NO 14 IH: 1065 : DNA VISM: Artif: JRE: 2 INFORMATIC nucleotide ENCE: 14 gtgaacccat aggaaaggtg tgctcgattt caccaccacc attcagcccc tcggcccggc aacctcctga cgagaggtct gtgggaggcc ttcccgcacc attcccgcacc	icial Sequer DN: Descript tcttcccagc gccgagggct catactctca coctccgccc ggcgggtggc gctgcatggt taagcgggaa tcccccgccc tggccccaga tggacttcac tgccgcagca	nce tion of Art: ttttccacgt gaacccgagt atgggactgg ccagcgtttt ttggtgtccg agatttctgc ggaggtggct gggggcacctg cctgatacac gcgagtttcc tacgccctc gccttgggcc	ificial Seq tcgcgtctcc ctggaggtac acgggctggg actacccgga gcccggctgc gcccggctgc cccccttag cccctccatt ctgcccccc	ttgccgagag ggatgatgat agcggaggca gcctgaattg gggtcgactc ccccgggctt aagctgtatg tccagatgga tggaggaccg agcttttggt	1086 etic 60 120 180 240 300 360 420 480 540 600 660 720

cgcagatcat	ggccgcgcaa	gcggacagct	acgcatacct	gctcatatgc	gggctgcgga	840
aaaacctaca	caaagagttc	acaccttaaa	gcgcaccttc	gcacacacac	aggcgagaaa	900
ccatatcatt	gtaactggga	cggatgtgga	tggaaatttg	ctcggtctga	tgagcttacg	960
agacattatc	gaaagcatac	cggacatcgg	ccctttcaat	gccatctttg	tgacagagct	1020
ttttcccggt	ctgaccacct	cgctctgcac	atgaagaggc	acatg		1065
<210> SEQ : <211> LENG <212> TYPE <213> ORGAI <220> FEATU <223> OTHEN polyn	ID NO 15 TH: 1035 : DNA NISM: Artif: JRE: & INFORMATIC Sucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	ience: Synthe	etic
<400> SEQUI	ENCE: 15					
atgctcatgt	ttgacccagt	tcctgtcaag	caagaggcca	tggaccctgt	ctcagtgtca	60
tacccatcta	attacatgga	atccatgaag	cctaacaagt	atggggtcat	ctactccaca	120
ccattgcctg	agaagttctt	tcagacccca	gaaggtetgt	cgcacggaat	acagatggag	180
ccagtggacc	tcacggtgaa	caagcggagt	tcaccccctt	cggctgggaa	ttcgccctcc	240
tctctgaagt	tecegtecte	acaccggaga	gcctcgcctg	ggttgagcat	gccttcttcc	300
agcccaccga	taaaaaaata	ctcaccccct	tctccaggcg	tgcagccctt	cggcgtgccg	360
ctgtccatgc	caccagtgat	ggcagctgcc	ctctcgcggc	atggaatacg	gagcccgggg	420
atcctgcccg	tcatccagcc	ggtggtggtg	cagcccgtcc	cctttatgta	cacaagtcac	480
ctccagcagc	ctctcatggt	ctccttatcg	gaggagatgg	aaaattccag	tagtagcatg	540
caagtacctg	taattgaatc	atatgagaag	cctatatcac	agaaaaaaat	taaaatagaa	600
cctgggatcg	aaccacagag	gacagattat	tatcctgaag	aaatgtcacc	ccccttaatg	660
aactcagtgt	cccccccgca	agcattgttg	caagagaatc	acccttcggt	catcgtgcag	720
cctgggaaga	gacctttacc	tgtggaatcc	ccggatactc	aaaggaagcg	gaggatacac	780
agatgtgatt	atgatggatg	caacaaagtg	tacactaaaa	gctcccactt	gaaagcacac	840
agaagaacac	acacaggaga	aaaaccctac	aaatgtacat	gggaagggtg	cacatggaag	900
tttgctcggt	ctgatgaact	aacaagacat	ttccgaaaac	atactggaat	caaacctttc	960
cagtgcccgg	actgtgaccg	cagcttctcc	cgttctgacc	atcttgccct	ccataggaaa	1020
cgccacatgc	tagtc					1035
<210> SEQ : <211> LENG' <212> TYPE <213> ORGAI <220> FEAT <223> OTHEI polyn	ID NO 16 TH: 1371 : DNA NISM: Artifi RE: & INFORMATIC nucleotide	icial Sequer DN: Descript	nce tion of Arti	ificial Sequ	nence: Synthe	etic
<400> SEQUI	ENCE: 16					
atggctacaa	gggtgctgag	catgagcgcc	cgcctgggac	ccgtgcccca	gccgccggcg	60
ccgcaggacg	agccggtgtt	cgcgcagctc	aagccggtgc	tgggcgccgc	gaatccggcc	120
cgcgacgcgg	cgctcttccc	cggcgaggag	ctgaagcacg	cgcaccaccg	cccgcaggcg	180
cagcccgcgc	ccgcgcaggc	cccgcagccg	gcccagccgc	ccgccaccgg	cccgcggctg	240

cctccagagg	acctggtcca	gacaagatgt	gaaatggaga	agtatctgac	acctcagctt	300
cctccagttc	ctataattcc	agagcataaa	aagtatagac	gagacagtgc	ctcagtcgta	360
gaccagttct	tcactgacac	tgaagggtta	ccttacagta	tcaacatgaa	cgtcttcctc	420
cctgacatca	ctcacctgag	aactggcctc	tacaaatccc	agagaccgtg	cgtaacacac	480
atcaagacag	aacctgttgc	cattttcagc	caccagagtg	aaacgactgc	ccctcctccg	540
gccccgaccc	aggccctccc	tgagttcacc	agtatattca	gctcacacca	gaccgcagct	600
ccagaggtga	acaatatttt	catcaaacaa	gaacttccta	caccagatct	tcatctttct	660
gtccctaccc	agcagggcca	cctgtaccag	ctactgaata	caccggatct	agatatgccc	720
agttctacaa	atcagacagc	agcaatggac	actcttaatg	tttctatgtc	agctgccatg	780
gcaggcctta	acacacacac	ctctgctgtt	ccgcagactg	cagtgaaaca	attccagggc	840
atgccccctt	gcacatacac	aatgccaagt	cagtttcttc	cacaacaggc	cacttacttt	900
cccccgtcac	caccaagctc	agagcctgga	agtccagata	gacaagcaga	gatgctccag	960
aatttaaccc	cacctccatc	ctatgctgct	acaattgctt	ctaaactggc	aattcacaat	1020
ccaaatttac	ccaccaccct	gccagttaac	tcacaaaaca	tccaacctgt	cagatacaat	1080
agaaggagta	accccgattt	ggagaaacga	cgcatccact	actgcgatta	ccctggttgc	1140
acaaaagttt	ataccaagtc	ttctcattta	aaagctcacc	tgaggactca	cactggtgaa	1200
aagccataca	agtgtacctg	ggaaggctgc	gactggaggt	tcgcgcgatc	ggatgagctg	1260
acccgccact	accggaagca	cacaggcgcc	aagcccttcc	agtgcggggt	gtgcaaccgc	1320
agettetege	actetaacea	cetaaceeta	catatraara	aacaccaaaa	~	1371
ageeeeege	goeoegacoa	corggooorg	cacacyaaya	ggeaceagaa	C	
<210> SEQ : <211> LENG: <212> TYPE <213> ORGAI <220> FEATU <223> OTHEN polyn	ID NO 17 TH: 849 : DNA NISM: Artifi RE: NIFORMATIC nucleotide	icial Sequer DN: Descript	nce	ificial Sequ	lence: Synth	etic
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAN &lt;220&gt; FEATU &lt;223&gt; OTHEN polyn &lt;400&gt; SEQUN</pre>	ID NO 17 TH: 849 : DNA IISM: Artifi JRE: NISM: Artifi JRE: INFORMATIC NUCLEOTIDE SNCE: 17	icial Sequer DN: Descript	nce	ificial Sequ	ience: Synth	etic
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEAT &lt;223&gt; OTHEI polyi &lt;400&gt; SEQUI atggacgtgc</pre>	ID NO 17 TH: 849 : DNA VISM: Artif: RE: NIFORMATIC Nucleotide ENCE: 17 tccccatgtg	cial Sequer DN: Descript cagcatcttc	caggagctcc	ificial Sequ agatogtgoa	cence: Synth	etic 60
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAN &lt;220&gt; FEATU &lt;223&gt; OTHEI polyn &lt;400&gt; SEQUN atggacgtgc tacttctcgg</pre>	ID NO 17 TH: 849 : DNA IISM: Artifi RE: NINFORMATIC DUCLEOTIDE ENCE: 17 tccccatgtg cgctgccgtc	cial Sequer DN: Descript cagcatcttc tctggaggag	caggagetee tactggcaac	ificial Seq agatcgtgca agacctgcct	ence: Synth cgagaccggc agagctggaa	etic 60 120
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAN &lt;220&gt; FEATU &lt;223&gt; OTHEN polyn &lt;400&gt; SEQUN atggacgtgc tacttctcgg cgttacctcc</pre>	D NO 17 TH: 849 : DNA UISM: Artifi JRE: NISFORMATIC LUCLEOTIDE ENCE: 17 tccccatgtg cgctgccgtc agagcgagcc	cial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt	caggagctcc tactggcaac tcagcctcag	agatcgtgca agacctgcct aaatcaaatt	lence: Synth cgagaccggc agagctggaa tgacagccag	etic 60 120 180
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEATU &lt;223&gt; OTHEI polyn &lt;400&gt; SEQUI atggacgtgc tacttctcgg cgttacctcc gaagatctgt</pre>	ID NO 17 TH: 849 : DNA NISM: Artifi RE: RE: INFORMATIC Ducleotide ENCE: 17 tccccatgtg cgctgccgtc agagcgagcc ggaccaaaat	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt cattctggct	caggagctcc tactggcaac tcagcctcag cgggagaaaa	ificial Sequ agatcgtgca agacctgcct aaatcaaatt aggaggaatc	ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag	etic 60 120 180 240
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAN &lt;220&gt; FEATU &lt;223&gt; OTHEN polyn &lt;400&gt; SEQUN atggacgtgc tacttctcgg cgttacctcc gaagatctgt atatcttcca</pre>	D NO 17 TH: 849 : DNA IISM: Artifi JRE: R INFORMATIC LUCLEOTIDE ENCE: 17 tccccatgtg cgctgccgtc agagcgagcc ggaccaaaat gtcctccaga	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt cattctggct ggacactctc	caggagctcc tactggcaac tcagcctcag cgggagaaaa atcagcccga	ificial Sequ agatogtgoa agacotgoot aaatoaaatt aggaggaato gottttgtta	ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag caacttagag	etic 60 120 180 240 300
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEATI &lt;223&gt; OTHEI polyi &lt;400&gt; SEQUI atggacgtgc tacttctcgg cgttacctcc gaagatctgt atatcttcca accaacagcc</pre>	ID NO 17 TH: 849 : DNA NISM: Artifi RE: RE: INFORMATIC DUCLEOTIDE ENCE: 17 tccccatgtg cgctgccgtc agagcgagcc ggaccaaaat gtcctccaga tgaactcaga	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt cattctggct ggacactctc tgtcagcagc	caggagctcc tactggcaac tcagcctcag cgggagaaaa atcagcccga gaatcctctg	ificial Sequ agatcgtgca agacctgcct aaatcaaatt aggaggaatc gcttttgtta acagctccga	ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag caacttagag ggaactttct	etic 60 120 180 240 300 360
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAN &lt;220&gt; FEATU &lt;223&gt; OTHEI polyn &lt;400&gt; SEQUN atggacgtgc tacttctcgg cgttacctcc gaagatctgt atatcttcca accaacagcc cccacggcca</pre>	ID NO 17 TH: 849 : DNA UISM: Artifi JRE: & INFORMATIC bucleotide ENCE: 17 tccccatgtg cgctgccgtc agagcgagcc ggaccaaaat gtcctccaga tgaactcaga agtttacctc	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt cattctggct ggacactctc tgtcagcagc cgaccccatt	caggagctcc tactggcaac tcagcctcag cgggagaaaa atcagcccga gaatcctctg ggcgaagttt	ificial Seq agatcgtgca agacctgcct aaatcaaatt aggaggaatc gcttttgtta acagctccga tggtcagctc	ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag ggaacttagag ggaacttct gggaaaattg	etic 60 120 180 240 300 360 420
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEAT &lt;223&gt; OTHEI polyi &lt;400&gt; SEQUI atggacgtgc tacttctcgg cgttacctcc gaagatctgt atatcttcca accaacagcc cccacggcca agctcctctg</pre>	DD NO 17 TH: 849 : DNA NISM: Artifi RE: 2 INFORMATIC Ducleotide ENCE: 17 tcccccatgtg cgctgccgtc agagcgagcc ggaccaaaat gtcctccaga tgaactcaga agtttacctc tcacctccac	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt cattctggct ggacactctc tgtcagcagc cgaccccatt gcctccatct	cacaggagctcc tactggcaac tcagcctcag cgggagaaaa atcagcccga gaatcctctg ggcgaagttt tctccggaac	ificial Sequ agatcgtgca agacctgcct aaatcaaatt aggaggaatc gcttttgtta acagctccga tggtcagctc tgagcaggga	ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag caacttagag ggaactttct gggaaaattg accttctcaa	etic 60 120 180 240 300 360 420 480
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;211&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEATU &lt;223&gt; OTHEI polyn &lt;400&gt; SEQUI atggacgtgc tacttctcgg cgttacctcc gaagatctgt atatcttcca accaacagcc cccacggcca agctcctctg ctgtggggtt</pre>	ID NO 17 TH: 849 : DNA NISM: Artifi RE: NIFORMATIC DUCLEOTIDE ENCE: 17 tccccatgtg cgctgccgtc agagcgagcc ggaccaaaat gtcctccaga tgaactcaga agtttacctc tcacctccac gcgtgcccgg	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt cattctggct ggacactctc tgtcagcagc cgacccatt gcctccatct ggagctgccc	caggagctcc tactggcaac tcagcctcag cgggagaaaa atcagcccga gaatcctctg ggcgaagttt tctccggaac tcgccaggga	ificial Sequ agatcgtgca agacctgcct aaatcaaatt aggaggaatc gcttttgtta acagctccga tggtcagctc tgagcaggga aggtgcgcag	ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag ggaacttag gggaaaattg accttctcaa cgggacttcg	etic 60 120 180 240 300 360 420 480 540
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEAT &lt;223&gt; OTHEI polyn &lt;400&gt; SEQUI atggacgtgc tacttctcgg cgttacctccg gaagatctgt atatcttcca accaacagcc cccacggcca agctcctctg ctgtggggtt gggaagccag</pre>	DD NO 17 TH: 849 : DNA VISM: Artif: RE: 2 INFORMATIC ucleotide ENCE: 17 tccccatgtg cgctgccgtc agagcgagcc ggaccaaaat gtcctccaga tgaactcaga agtttacctc tcacctccac gcgtgcccgg gtgacaaggg	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt cattctggct ggacactctc tgtcagcagc cgaccccatt gcctccatct ggagctgccc aaatggcgat	caggagctcc tactggcaac tcagcctcag cgggagaaaa atcagcccga gaatcctctg ggcgaagttt tctccggaac tcgccaggga	ificial Sequ agatcgtgca agacctgcct aaatcaaatt aggaggaatc gcttttgtta acagctccga tggtcagctc tgagcaggga aggtgcgcag	ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag ggaacttagag ggaactttct gggaaaattg accttctcaa cgggacttcg gagggtgcac	etic 60 120 180 240 300 360 420 480 540 600
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEATI &lt;223&gt; OTHEI polyn &lt;400&gt; SEQUI atggacgtgc tacttctcgg cgttacctcc gaagatctgt atatcttcca accaacagcc cccacggcca agctcctctg ctgtggggtt gggaagccag cggtgccact</pre>	D NO 17 TH: 849 : DNA NISM: Artifi RE: NICE: 17 tccccatgtg cgctgccgtc agagcgagcc ggaccaaaat gtcctccaga tgaactcaga agtttacctc tcacctccac gcgtgcccgg gtgacaaggg	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt cattctggct ggacactctc tgtcagcagc cgacccatt gcctccatct ggagctgccc aaatggcgat caggaaagtt	caggagctcc tactggcaac tcagcctcag cgggagaaaa atcagcccga gaatcctctg ggcgaagttt tctccggaac tcgccaggga gcctcccccg	ificial Sequ agatcgtgca agacctgcct aaatcaaatt aggaggaatc gcttttgtta acagctccga tgggcagcag aggtgcgcag acggcaggag gctcccactt	ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag ggaacttct gggaaaattg accttctcaa cgggacttcg gagggtgcac gaaggtgcac	etic 60 120 180 240 300 360 420 480 540 600 660
<pre>&lt;210&gt; SEQ : &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGAI &lt;220&gt; FEATT &lt;223&gt; OTHEI polyn &lt;400&gt; SEQUI atggacgtgc tacttctcgg cgttacctccg aagatctgt atacttctcca accaacagcc cccacggcca agctcctctg ctgtggggtt gggaagccag cggtgccact cagcggacgc</pre>	ID NO 17 TH: 849 : DNA VISM: Artif: RE: RE: INFORMATIC oucleotide INCE: 17 tccccatgtg cgctgccgtc agagcgagcc agagcgagcc ggaccaaaat gtcctccaga tgaactcaga agtttacctc tcacctccac gcgtgcccgg gtgacaaggg ttaacggctg acacaggaga	icial Sequer DN: Descript cagcatcttc tctggaggag ctgctatgtt ggacactctc tgtcagcagc cgacccatt gcctccatct ggagctgccc aaatggcgat caggaagtt aaagccttac	caggagctcc tactggcaac tcagcctcag cgggagaaaa atcagcccga gaatcctctg ggcgaagttt tctccggaac tcgccaggga gcctcccccg tacaccaaaa	ificial Sequ agatcgtgca agacctgcct aaatcaaatt aggaggaatc gctttgtta acagctccga tggtcagctc tgagcaggga aggtgcgcag acggcaggag gctcccactt gggaagggtg	e ence: Synth cgagaccggc agagctggaa tgacagccag cgaactgaag ggaacttagag ggaacttct gggaaaattg accttctcaa cgggacttcg gagggtgcac gaaagcacac tgagtggcgt	etic 60 120 180 240 300 360 420 480 540 600 660 720

aaatgeteee actgtgacag gtgtttttee aggtetgaee acetggeeet geacatgaag	j 840
aggcacete	849
<pre>&lt;210&gt; SEQ ID NO 18 &lt;211&gt; LENGTH: 906 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synt polynucleotide</pre>	chetic
<400> SEQUENCE: 18	
atggacgtgt tggctagtta tagtatattc caggagctac aacttgtcca cgacaccggc	60
tacttctcag ctttaccatc cctggaggag acctggcagc agacatgcct tgaattggaa	a 120
cgctacctac agacggagcc ccggaggatc tcagagacct ttggtgagga cttggactgt	180
tteeteeacg etteeeetee eeegtgeatt gaggaaaget teegtegett agaeeeeetg	g 240
ctgeteeeeg tggaagegge catetgtgag aagagetegg cagtggacat ettgetetet	300
cgggacaagt tgctatctga gacctgcctc agcctccagc cggccagctc ttctctagac	360
agctacacag ccgtcaacca ggcccagctc aacgcagtga cctcattaac gcccccatcg	420
tcccctgagc tcagccgcca tctggtcaaa acctcacaaa ctctctctgc cgtggatggc	2 480
acggtgacgt tgaaactggt ggccaagaag gctgctctca gctccgtaaa ggtgggaggg	540
gtcgcaacag ctgcagcagc cgtgacggct gcggggggccg ttaagagtgg acagagcgac	c 600
agtgaccaag gagggctagg ggctgaagca tgtcccgaaa acaagaagag ggttcaccgc	660
tgtcagttta acgggtgccg gaaagtttat acaaaaagct cccacttaaa ggcccaccag	720
aggactcaca caggtgagaa gccttataag tgctcatggg agggatgtga gtggcgtttt	780
gcacgaagcg atgagctcac gaggcactac aggaaacaca caggtgcaaa gcccttcaaa	a 840
tgcaaccact gcgacaggtg tttttccagg tctgaccatc ttgccctcca catgaagaga	a 900
catatc	906
<210> SEQ ID NO 19 <211> LENGTH: 1077 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synt polynucleotide	hetic
<400> SEQUENCE: 19	
atggtcgata tggataaact cataaacaac ttggaggtcc aacttaattc agaaggtggc	60
tcaatgcagg tattcaagca ggtcactgct tctgttcgga acagagatcc ccctgagata	a 120
gaatacagaa gtaatatgac tteteeaaca eteetggatg eeaaceecat ggagaaeeca	a 180
gcactgttta atgacatcaa gattgagccc ccagaagaac ttttggctag tgatttcagc	240
ctgccccaag tggaaccagt tgacctctcc tttcacaagc ccaaggctcc tctccagcct	300
getageatge tacaagetee aataegteee eecaageeae agtettetee eeagaeeett	360
gtggtgtcca cgtcaacatc tgacatgagc acttcagcaa acatteetac tgttetgace	420
ccaggetetg teetgacete eteteagage aetggtagee ageagatett acatgteatt	480
cacactaton optoanton totoccaat aanatomoto optocaato	4 540

				-contin	nued	
gtagtgcagt	ctctgcccat	ggtgtatact	actttgcctg	cagatggggg	ccctgcagcc	600
attacagtcc	cactcattgg	aggagatggt	aaaaatgctg	gatcagtgaa	agttgacccc	660
acctccatgt	ctccactgga	aattccaagt	gacagtgagg	agagtacaat	tgagagtgga	720
tcctcagcct	tgcagagtct	gcagggacta	cagcaagaac	cagcagcaat	ggcccaaatg	780
cagggagaag	agtcgcttga	cttgaagaga	agacggattc	accaatgtga	ctttgcagga	840
tgcagcaaag	tgtacaccaa	aagctctcac	ctgaaagctc	accgcagaat	ccatacagga	900
gagaagcctt	ataaatgcac	ctgggatggc	tgctcctgga	aatttgctcg	ctcagatgag	960
ctcactcgcc	atttccgcaa	gcacacaggc	atcaagcctt	ttcggtgcac	agactgcaac	1020
cgcagctttt	ctcgttctga	ccacctgtcc	ctgcatcgcc	gtcgccatga	caccatg	1077
<210> SEQ 1 <211> LENGT <212> TYPE <213> ORGAN <220> FEATU <223> OTHEN polyr	D NO 20 TH: 732 DNA HISM: Artif: RE: NIFORMATIC NUCLEOTIDE	icial Sequer DN: Descript	nce tion of Arti	ificial Sequ	ience: Synth	etic
<400> SEQUE	ENCE: 20					
atgtccgcgg	ccgcctacat	ggacttcgtg	gctgcccagt	gtctggtttc	catttcgaac	60
cgcgctgcgg	tgccggagca	tggggtcgct	ccggacgccg	agcggctgcg	actacctgag	120
cgcgaggtga	ccaaggagca	cggtgacccg	ggggacacct	ggaaggatta	ctgcacactg	180
gtcaccatcg	ccaagagctt	gttggacctg	aacaagtacc	gacccatcca	gaccccctcc	240
gtgtgcagcg	acagtctgga	aagtccagat	gaggatatgg	gatccgacag	cgacgtgacc	300
accgaatctg	ggtcgagtcc	ttcccacagc	ccggaggaga	gacaggatcc	tggcagcgcg	360
cccagcccgc	tctccctcct	ccatcctgga	gtggctgcga	aggggaaaca	cgcctccgaa	420
aagaggcaca	agtgccccta	cagtggctgt	gggaaagtct	atggaaaatc	ctcccatctc	480
aaagcccatt	acagagtgca	tacaggtgaa	cggccctttc	cctgcacgtg	gccagactgc	540
cttaaaaagt	tctcccgctc	agacgagctg	acccgccact	accggaccca	cactggggaa	600
aagcagttcc	gctgtccgct	gtgtgagaag	cgcttcatga	ggagtgacca	cctcacaaag	660
cacgcccggc	ggcacaccga	gttccacccc	agcatgatca	agcgatcgaa	aaaggcgctg	720
gccaacgctt	tg					732
<210> SEQ 1 <211> LENG <212> TYPE <213> ORGAN <220> FEAT <223> OTHEN polyr	D NO 21 TH: 1440 DNA HISM: Artif: RE: NINFORMATIC nucleotide	icial Sequer DN: Descript	nce tion of Arti	ificial Sequ	ience: Synth	etic
<400> SEQUI	ENCE: 21					
atgctcaact	tcggtgcctc	tctccagcag	actgcggagg	aaagaatgga	aatgatttct	60
gaaaggccaa	aagagagtat	gtattcctgg	aacaaaactg	cagagaaaag	tgattttgaa	120
gctgtagaag	cacttatgtc	aatgagctgc	agttggaagt	ctgattttaa	gaaatacgtt	180
gaaaacagac	ctgttacacc	agtatctgat	ttgtcagagg	aagagaatct	gcttccggga	240
acacctqatt	ttcatacaat	cccagcattt	tgtttgactc	caccttacaq	tccttctgac	300

tttgaaccct	ctcaagtgtc	aaatctgatg	gcaccagcgc	catctactgt	acacttcaag	360	
tcactctcag	atactgccaa	acctcacatt	gccgcacctt	tcaaagagga	agaaaagagc	420	
ccagtatctg	cccccaaact	ccccaaagct	caggcaacaa	gtgtgattcg	tcatacagct	480	
gatgcccagc	tatgtaacca	ccagacctgc	ccaatgaaag	cagccagcat	cctcaactat	540	
cagaacaatt	cttttagaag	aagaacccac	ctaaatgttg	aggctgcaag	aaagaacata	600	
ccatgtgccg	ctgtgtcacc	aaacagatcc	aaatgtgaga	gaaacacagt	ggcagatgtt	660	
gatgagaaag	caagtgctgc	actttatgac	ttttctgtgc	cttcctcaga	gacggtcatc	720	
tgcaggtctc	agccagcccc	tgtgtcccca	caacagaagt	cagtgttggt	ctctccacct	780	
gcagtatctg	caggggggagt	gccacctatg	ccggtcatct	gccagatggt	teccettect	840	
gccaacaacc	ctgttgtgac	aacagtcgtt	cccagcactc	ctcccagcca	gccaccagcc	900	
gtttgccccc	ctgttgtgtt	catgggcaca	caagtcccca	aaggcgctgt	catgtttgtg	960	
gtaccccagc	ccgttgtgca	gagttcaaag	cctccggtgg	tgagcccgaa	tggcaccaga	1020	
ctctctccca	ttgcccctgc	tcctgggttt	tccccttcag	cagcaaaagt	cactcctcag	1080	
attgattcat	caaggataag	gagtcacatc	tgtagccacc	caggatgtgg	caagacatac	1140	
tttaaaagtt	cccatctgaa	ggcccacacg	aggacgcaca	caggagaaaa	gcctttcagc	1200	
tgtagctgga	aaggttgtga	aaggaggttt	gcccgttctg	atgaactgtc	cagacacagg	1260	
cgaacccaca	cgggtgagaa	gaaatttgcg	tgccccatgt	gtgaccggcg	gttcatgagg	1320	
agtgaccatt	tgaccaagca	tgcccggcgc	catctatcag	ccaagaagct	accaaactgg	1380	
cagatggaag	tgagcaagct	aaatgacatt	gctctacctc	caacccctgc	tcccacacag	1440	
<pre>&lt;211&gt; LENG' &lt;212&gt; TYPE &lt;213&gt; ORGAJ &lt;220&gt; FEATU &lt;223&gt; OTHEJ polyu &lt;400&gt; SEOUU</pre>	TH: 1536 : DNA NISM: Artif: JRE: R INFORMATIC nucleotide	icial Seque DN: Descrip	nce tion of Art.	ificial Sequ	lence: Synth	netic	
atgcatactc	ctgatttcgc	tggacctgac	aacacccaaa	ccatagacat	tatggagatt	60	
tqtqaatcta	tactcgaaag	aaadadadat	gattcadado	qaaqtacatq	ctctatcete	120	
qaqcaaacaq	acatogadad	qqtaqaaqct	ctqatataca	tqtccaqttq	qqqtcaqaqa	180	
tcccagaagg	qqqacttact	tagaatccga	ccqcttactc	caqtttccca	taqcqqcqaq	240	
gtaacaacta	ctgttcatat	ggacqcaqcc	acgcctgage	tgcccaaaqa	ctttcacaqc	300	
ctctcaactc	tttgcatcac	tccaccacaq	tecceqate	ttgtcgaacc	atcaacccqq	360	
accectgtta	- gcccgcaagt	tacagattca	aaggegtgta	ccgcgaccga	tgttctgcag	420	
aqttcagcgg	ttqtaqcqcq	qqcattgage	qqaqqqqctq	aacqaqqtct	qttaaatctt	480	
gaaddedtad	cgagttetee	ttatagage	aagggtacta	gtgttattcg	gcataccoor	540	
gagagt.cccc	cagettett	ccccaccata	caaaccccac	actotococt	tagtgattcc	600	
aaaaaaaaa	aggeoget	attenter	ttagenera	ttagarate	agegatter	660	
cyyyaagggg	ayyaacaget	yrrgggccad	LLUYAYACAC		acactigaca	66U	
gatagettge	tgtccaccaa	cctggtgtca	tgtcaacctt	gtttgcacaa	gteegggggt	/20	
ctccttctga	ctgacaaagg	tcaacaagcg	ggatggcctg	gcgctgtcca	aacatgcagt	780	
cctaaaaact	acqaaaatqa	tttgcctagg	aaaaccacqc	cacttatcag	tgtgagtgtt	840	

cccccccccccccccccccccccccccccccccccc						
gegetettiga         ageceeeeee         acaactigtee         gittiggetettiga         gegegeoorg         cecegeaaace         opgegegeaaa         eggedgeoorg         cecegeaaaca         gegetettiga         gegetettiga         gegetettiga         eggetettiga	cccgctccac ctgtcct	gtg ccagatgatc (	cctgtaaccg	ggcaatcatc	tatgttgcct	900
gcagogocog occegocaco ogtgttogtg gggcocogt toccogoagg tgoagtoatg         1020           ttggttotte occagoggge octocogoa caqeteceg gtgoagoaa tgtaatget         1080           gcoggaaaca cgaaattgt gcocttgaa ceqetecag ttttoataac gagetecagg         1140           aattgtgtge cacaagtega ottotcaag agaeggaact atgtgtget tttocaagga         1200           tgoagaaaaa catatttaa atectecat otgaaagcaa atetteggaa catacagga         1200           tgoagaaaaa catattega otggaagg ggaaaaaaat tegttgtee agttgtgaa         1320           caagtegga ogcagogg gaaaaaaaat tegttgtee agttgtgaa         1380           aagaagattta tgaggteega ccateceag agtegee gaagegaaca gaaeggaaca gaaeggaac         1400           aaaatteetig getggeaage cgaggtggg aaaacteaace gaaegeete egetgaace         1530           ccalo SEQ ID NO 23         1536           c210 > SEQ ID NO 23         1533           c210 > SEQ UD NO 23         1530           c210 > SEQ UD NO 23         1533           c210 > SEQ UD NO 23         1530           c210 > SEQ UD NO 23         1500           c210 > SEQ UD NO 23         1500           c210 > SEQUENCE: 23         110           atgaacatte acatgaage cagaeggaa aagaegate atacattega gaaecegaatg         60           ttgaagttgg atggee dgaecegaa acateeee atacattega gaaecgaatg         120           gdgaaeggaa gaeceeedaa etacaecega acateee ateeedaa         120	gcgttcttga agccccc	ccc acaactgtcc 🤉	gttggtactg	ttcgcccgat	ccttgcgcaa	960
ttggttette eecagggge eeteegea eagetegg gtgeagega tgteatgget 1000 geoggaaaca egaattgt geoettgea eegetega atgtgaget tteetaac gageteadag 1140 aattgtgtg eacaagtega etteeaga agaeggaaet atgtggae etteeagag agtgatgag 1200 tgeagaaaaa eatatteaa ateeteea etgaaageae atetteegga eetaeagga 1200 eteagtegae ategeaggae geataeeggg gaaaaaaaat tegttgtee agttggea 1380 agaagattta tgaggteega eeateeagg gaaaaaaaat tegttgtee agttggea 1380 agaagatta tgaggteega eeateeagg gaaaaaaaat tegttgtee agttggea 1380 aaatteetg getggeaage egaggtgga aaeteaaee gaategete egetgaaee 1500 coeggeagee egetggtaag tageetgee agteee agteee eeses synthetic 2210 > SEQ ID NO 23 2211 > LENGTH: 1206 2212 > TTFE DNA 2213 > ORGNISM: Artificial Sequence 2223 > OTHEN INFORMATION: Description of Artificial Sequence: Synthetic 2213 > ORGNISM: Artificial Sequence 2223 > OTHEN INFORMATION: Description of Artificial Sequence: Synthetic 2213 > ORGNISM: artificial Sequence 2223 > OTHEN INFORMATION: Description of Artificial Sequence: Synthetic 210 > SEQUENCE: 23 atgaaeatte acatgaageg cagagegat aggaegat geegetett geteaaeaat 180 gtgaaggag ageegeetga ggaetetee teegtagae atteeaga gaaeaaaat 120 ggateeceaa acgteeaea etaecegga ageeteta etteeaga eagaetgag 240 coegtagate tteeaattaa caaageeaga eateteea etgeggeag geeaeaatg eacaagett 340 agtaagtatg eageaagte ategtgee ageeteag geeaeaag eacaeaget 540 agtaagtatg eageaagte ategtgea ageeteag geeaeaag eacaeaget 540 agtaagtatg eagaagtge ateagtgea teeaagtgt eeceaga gaeaceaag teeaaagt 540 agtaagtatg eagaagtge eecegtee agteegata geeagaeeg agaeaeagt 540 agtaagtatg eagaagtge eecegteegee eagteagta acaeageeg agaeaeagt 540 agtaagtatg eagaagtge eeceegee eagteaga ageeteaga gaaeaeagt 540 agtaagtatg eagagatge eeceegee eagteaga ageeteaga gaaeaeagt 540 agtaagtatg eagaagtge eeceegee eagteaga ageeteaga gaaeaeagt 540 agtaeagta eagaagtge eeceegee eagteaga acaeageeg eagaaeage 320 aggaagatta tegeeteege eeceega ageeteaga gaaeaeaga 322 aggaaeaga eecagagga eeceegee eagteaga acaeageeg eagaaeate 320 aggaagaattg eaggaaggeegeeteg eageeagaeg eagaaeaeageeg ageeegeegeegeegeegee	gcagcgcccg ccccgca	acc cgtgttcgtg 🤅	gggcccgctg	tcccgcaggg	tgcagtcatg	1020
<pre>gccggaaca cgaactgt gcccctga ccgctcag ttttcatac gagctcaag 1140 aattgtgtg cacaagtcg cttctcacg agacggaact atgtgtgct tttcccaggt 1200 tgcagaaaaa catattcaa atcctcat ctgaaagca atctcgga catacagga 1260 gagaagctt ttaattgtag ctgggatgg tgtgataaaa aattcgcaag aagtgatgag 1320 ctcagtcgac atcgcagga gcataccggg gaaaaaaat togttgtc agttgtga 1380 aqaagatta tgaggtccg ccatctcac aagcacgge gacgccaat gactacaag 1440 aaattcctg gctggcaag cqaggtgga aactcaac gaacgctc cgctgaace 1500 cccgggaagc cgctggtag tatgctgc agtgcc</pre>	ttggttcttc cccaggg	ggc cctcccgcca (	ccageteegt	gtgcagcgaa	tgtcatggct	1080
aattgtgtge         caaagtega         cttecaega         ageggaact         attgtgtge         tteceagtg         1200           tgcagaaaaa         catattteaa         atceetetaa         cgtagaagac         atteggaga         1200           ctcagtegaa         catattteaa         atceetetaa         atteggaga         1200           ctcagtegaa         ctagtegga         gataccaegg         gatacaaga         atteggaga         1300           aaaateetg         geggeaage         cgaggggg         aaaataaat         tegtttgte         agadattt           caagtegga         cgeggaage         cgaggtgga         aaaetecaace         gadegeteta         1500           ceeggeagee         ceegggage         ceetgggaage         agategetee         1536           c210>         SEQ ID NO 23         ceetgaaga         ceetggaage         1536           c211>         LENGTH:         1200         ceetgaagaeg         gaacegaatg         60           c113>         SRQUENCE:         23         atgaacatte         acaageegg         acacegaatg         120           ggatececaa         acgteceaca         ceeteggaag         acacegaatg         120           ggatececaa         acgteceaca         ceetegaacage         acacegaatg         20 <t< td=""><td>gccggaaaca cgaaatt</td><td>gtt gccccttgca</td><td>cccgctccag</td><td>ttttcataac</td><td>gageteacag</td><td>1140</td></t<>	gccggaaaca cgaaatt	gtt gccccttgca	cccgctccag	ttttcataac	gageteacag	1140
tgcagaaaaa catatttcaa atcctcat ctgaaagca atcttcgga catacagga 1260 gagaagcctt ttaattgtag ctggatgge tgtgataaaa aattegaag aagtgatgag 1320 ctcagtegaa ategeaggae geataceggg gaaaaaaaat tegtttgtee agtttgtgae 1380 agaagattta tgaggteega ecateteae aageaegee gaegeeaet gaetaeaaag 1440 aaaatteetg getggeaage egaggtggga aaacteaaee gaategette egetgaatee 1500 eeeggeagee egetggtaag tatgeetgee agtgee 1536 c113 > LEMOTH: 1206 c212 > TYPE: DNA c213 > DROAMISM: Artificial Sequence c200 > FENTORE: c202 > FENTORE: c202 > FENTORE: c203 > GAURISM: Artificial Sequence c304 > SEQUENCE: 23 atgaacatte acatgaagee cagetgagg gaaaaaceg ageteetga gaeedaat 120 ggateeceaa aegteeae etaeegg atggee atggee atteetga gaacegaatg 60 ttgatgttgg atggeatgee egetgaeg gtaaaaaceg ageteetga getegaaeaa 120 ggateeceaa aegteeaeae etaeegga atggeeggagga geegeetga tetteetee 300 ggaaggagg geegeetga ggaeteete teegtagae tatteeagae acagetgag 240 eegtagate ttreaattaa caaagecaga acateteea etgeggaag tetteteee 300 gtaagtatga cageaagtge atetagteea agtetaega geetaegga tettetee 300 gtaagtatga cageaagtge atetagteea agtetaega geetagaeg tettetee 300 gtaagtatga cteeeggtee atetagteea agtetaega geetageg eacaaegt 540 agtgaagtag cteeeggtee atetagteea ageetagtg geetgggtge caacaatt 480 eteeatatta tteaaeeegg geetegge ageetagtg geetgggtg eacaaatt 540 agteaegtae ataggateee egetggee ageetagtg geetgggtg eacaaatt 540 agteaegtae ataggateea eacgetgge tetgeeaga aceesgetga tetgaaga 1540 eetgeaagga caeagatgga ceeeegge tetgeeaga acegatee ateegatgg 660 eatgggaaag caeagatgga ceeeegege tetgeeaga gaeagteea ateegatagt 720 gaeegaegta attageea egaagtee aceesgee tetgeeaga aegaegeeg agedeede 780 etgeaagga ataaceaga atteeeega gaeeteega ageageegg agedeege 780 etgeaaagg ataaaeaga atteeeega eeeegae eeegaagaeeg gaeegaegg 660 eatgggaaag caeagatgga eeeeegae tegaaeaee eeegaaeegae eeegaege 780 etgeaaagg ataaaeaga atteeeega eeeeaaee eeegaaeeegae eeegaege 780 etgeaaagg ataaaeaga atteeeegae eeeeaaee eeegaaeeegae tegaegaeegae 780 etgeaaagg ataaaeaega ataeeeea eeeeaaeeeeeeaaeeeeeeeeee	aattgtgtgc cacaagt	cga cttctcacga a	agacggaact	atgtgtgctc	tttcccaggt	1200
<pre>gagaagcett ttaattgtag etggatgge tgtgataaaa aattegeaag aagtgatgag 1320 ctcagtegae ategeaggae gedaecegg gaaaaaaaat tegtttgtee agtetgae 1380 agaagattta tgaggteega eedaecegg gaaaaeaeaat tegtttgtee agtetaeaag 1440 aaaatteetg getggeaage egaggtgga aaacteaee gaategette egetgaatee 1500 cceeggeagee egetggtaag tatgeetge agtege gaegeeaeat gaetaeaaag 1440 aaaatteetg getggeaage egaggtgga aaacteaee gaategette egetgaatee 1500 cceeggeagee egetggtaag tatgeetge agtege gaegeeaeat gaetaeaaag 1440 aaaatteetg getggeaage egaggtgga aaacteaee gaategette egetgaatee 1500 cceeggeagee egetggtaag tatgeetge agtege &lt;11&gt; LEKONTM: 1206 &lt;11&gt; UKONTM: Artificial Sequence &lt;220&gt; FEATURE! &lt;220&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynueleotide &lt;400&gt; SEQUENCE: 23 atgaacatte acatgaageg caagaegat aggaegeag geetetga gaeeegaag 60 ttgatgttgg atggeatge egetgaeg gtaaaaaeeg ageeeetga geetegaaeaa 120 ggateeeeaa aegteeaeaa etaeegga atggaggag tgeegeett geteaaeaat 180 gtgaaggag ageeeetga ggaetetet teegtagat atteeagae acagaetgag 240 ceegtagate ttreaattaa caaageeaga acateeeaa etteeaga gaeetageag tteetteee 300 gtaagtatga cageaagtge atetagteea agtetaega geetaega teetteeea 300 gtaagtatga cageaagtge atetagteea agtetaega geetaega teetteeea 300 gtaagtatga caegaagtge atetagteea agtetaega geetaega teetteeea 300 gtaagtatga etteeeggtee atetagteea ageegtagt geetgggtge caaeaattt 480 eteeatatt tteaaeeeg geetegte ageegtagt geetgggtge caaeaattt 480 eteeatatt tteaeeeeg geetegte ageegtagt geetgggetg eaaeaaget 540 agteaegtae ataggateea eaegtgge tetgeeega gaeagteea ateegatag 720 gaeegaegatg attageeaa eteecege gaeegeega ageegeetga 960 etgegaaag caeagatgg aceeeega eteegaaa eteeteega gaeageeega 240 eedgeaaga caeagatgg aceeeega gaeageee etgetgaega tegaeagee 780 etgeeaaeg eaegaagtga ateaeeega eedaaee etgegaagaeega egaeaegae 780 etgeeaaeg eaegaagtga ateaeeega eedaaee etgegaagaeega egaeaeae 780 etgeeaaeg eaegaagtga ateaeeega eedaaee etgegaagaeega eedaaeae 980 etgeaaaag ataaeeaga atteeeega eedaaeae eedaaea eedeaeae 980 etgaaaaegg aaaaeeega ateaeeega eedaaeae</pre>	tgcagaaaaa catattt	caa atcctctcat	ctgaaagcac	atcttcggac	ccatacagga	1260
ctcagtogac atogoagga gcatacoggg gaaaaaaat togtttgtoc agttgtgac 1380 agaagattta tgaggtocga ccatotcace aagcacgoog gaogocacat gactacaaag 1440 aaaatteetg getggeaage egaggtggga aaacteaace gaategette egetgaatee 1530 eeeggeagee egetggaag tatgeetgee agtgee 1536 <210> SEQ ID NO 23 <211> TPE DNA <212> TPE TNA <212> OREANISM: Artificial Sequence <220> FEATVRE: <222> OTHER INFORMATION: bescription of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 23 atgaacatte acatgaageg caagaegata aagaacatea atacattega gaacegaatg 60 ttgatgttgg atggeatgee egetgtaegg gtaaaaaceg ageteetga gtetgaacaa 120 ggateeceaa aegteecaaa etaceegga atggaggeag tgeegetet geteaacaat 180 gtgaaggaga geegeetga ggaetetee teegtagag tatteetaga eaagaetga 240 eeegtagate ttteaattaa caaageeaga acateteeta etgeggaag ttetteete 300 gtaagtatga cageaagtge atetagteea agtetaega geaetageag ttetteet 360 agtagaettg etagteee aeegtgae ageegetagt geegeetge eaeaagtt 420 acegtagate ttteaattaa caaageeaga acateteeta etgeggaag teeteeteet 360 agtagaetga eteecgetee agteegga ageegetag eaeaagetg 420 eeegtagate ataggatee egetgete eagtegatg geeggggtgg ceaacaatt 480 eteeaatta tteaeeeegg geeteegte eagteagtg geegggggag eaeaaagett 540 agteaegtae ataggatee egeteegte eagteagtg geeggggag tgeegaetg 720 gaegaeegaag eaceagatgga eceeegege ttgteaeega gaeagteea ateegaaget 720 gaegaeegaag eaceagatgga eceeegege ttgteaeega gaeagteea ateegaagt 720 gaegaeegaag eacagatgga eteeege tgteaeega gaeagteea ateegaagt 720 gaegaeegaag eacagatgga ateeeege tgteaeega gaeagteea ateegaagt 720 gaegaeegaag atatteeega atteeeega gaeaceee gaegeegagt 840 aacaggaag ataateeaga atteeeega eeeeaee eteegaagaeeega gaeeeeae 900 eggagaeega gaegaagtga ateaeeega teeeaaaa gaagaacaa teeeegaega 1020 etataeegga tgaaaaaeegta caaaageee teegaagaeega ageegaaga 1020 etataeegga tgaaaaaeegta caaaageee tegaagaeega tegaeegae 1020 etataeegga tgaaaaaeega caaaagee taeaeegae etagegeega tegaeegae 1020 etataeegga tgaaaaaeega eacaaeaa agteeaee teaaegeega tegaeegae 1020	gagaagcett ttaattg	tag ctgggatggc 1	tgtgataaaa	aattcgcaag	aagtgatgag	1320
agaagattta tgaggteega ecateteace aageaegee gaegeeaat gaetaeaaa 1440 aaaatteetg getggeaage egaggtggg aaaeteaace gaategete egetgaatee (210) SEQ ID NO 23 (212) INORMNTH: 1206 (212) TYPE: DNA (213) ENORMNTH: Artificial Sequence (220) FEATURE: (222) OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide (200) SEQUENCE: 23 atgaacatte acatgaageg caagaegata aagaacatea atacattega gaacegaatg 60 ttgatgttgg atggeatgee egetgtaegg gtaaaaaceg ageteetga gtetgaacaa 120 ggateeceaa acgteecaaa etaeceegat atggaggeag tgeegetet geteaacaat 180 gtgaaggaga ageegeetga ggaetetee teegtagate attteecaga eaegaetga 240 ceegtagate ttteaattaa eaaageeaga acateetea etgeeggaag tteetteete 300 gtaagtatga eageaagtge atetagtee agetetaga geaetageag tteetteete 300 gtaagtatga eageaagtge atetagtee ageetagg geetagaega 420 aeegtaettg etagteee egetggee ageetagg geetaga ateteeaga geaetageag tteetteete 300 gtaagtatga eageaagtge atetagtee agteegata ateeeaga eacaeagete 420 aeegtaegte tteeaattaa eaageeaga acateetee eegetggtag ceaacaattt 480 eteeaatta tteaeeegg geeteegt ageetagg geetagge geageaetag 220 gaagtaegtag eageaagtge atetagtee ageetagg geetggegg ceaacaattt 540 agteagtaeg ateeeggtee eegtegtee eagteegte eegtegtee eageegge 240 cegtagate atagateee egtegtegte eagteegtag acceeage eacaegete 720 gaegaeegaag eacagatgga eeceeegge ttgteeega gaeagteea ateegatagt 720 gaegaeegaag eacagatgga eeceeegge ttgteeega gaeagteea ateegatag 720 gaegaeegaag attageeta egtaacaeeg gaeetegga ageageega egaeeteg 720 gaegaeegae atateegaa atteeetg geetetga eeceegeege ttgteaeega ageegeegaeet 780 etgteaateg eaagaegeg acaegaegee eeceegeege eeceegae eeceegae 240 acaeggaaga gaeegeega acaegaegee eeceegeege eeceegae eeceegeege	ctcagtcgac atcgcag	gac gcataccggg g	gaaaaaaaat	tcgtttgtcc	agtttgtgac	1380
aaaatteetg getggeaage egaggtggga aaacteaace gaategette egetgaatee 1500 ceeegeagee egetggtaag tatgeetgee agtgee 1536 <li><li><li><li>SEQ ID NO 23 <li><li><li><li>ENGTH: 1206 <li><li><li><li>CCANISM: Artificial Sequence <li><li><li><li>CCANISM: Artificial Sequence <li><li><li><li>CCANISM: Artificial Sequence <li><li><li><li>CCANISM: Artificial Sequence <li><li><li><li><li>CCANISM: Artificial Sequence <li><li><li><li><li>CCANISM: Artificial Sequence <li><li><li><li><li><li>CCANISM: Artificial Sequence <li><li><li><li><li><li><li><li><li><li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li>	agaagattta tgaggtc	cga ccatctcacc a	aagcacgcgc	gacgccacat	gactacaaag	1440
<pre>ccccgcgagage cgctggtaag tatgcctgc agtgcc 1536 </pre>	aaaatteetg getggea	agc cgaggtggga a	aaactcaacc	gaatcgcttc	cgctgaatcc	1500
<210> SEQ ID NO 23 <211> LENGTH: 1206 <212> TYPE: DNA <213> ORGNISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 23 atgaacattc acatgaageg caagaegata aagaacatca atacattega gaacegaatg 60 ttgatgttgg atggeatgee egetgtaegg gtaaaaaceg ageteetgga gtetgaacaa 120 ggateeceaa acgteeacaa etaceeggat atggaggeag tgeegetett geteaacaat 180 gtgaagggag ageegeetga ggaetetet teegtagate attreeagae acagaetgag 240 ecceptagate ttreaattaa caaageeaga acateteeta etgeegtaag ttetteee 300 gtaagtatga cageaagtge atetagteea agteetaga geaetageag ttetteee 300 gtaagtatga cageaagtge atetagteea agteetaga geaetageag ttetteee 300 gtaagtatga caegeaagtge atetagteea agteetaga geaetageag ttetteete 360 agtagaettg etagtteeae aaeggegata acaegtetg egetgggtg eeaacaatt 480 eteetaatta tteaeceeg geeteegte agteegtag ateteegag caecaagett 540 agteaegtae ataggateea taataetata gttgtteett tgettgagga tggtagggge 660 catgggaaag eaeagatgga ecceegege ttgteaeega gaeagtetaa atecegatag 720 gaegaeagta attgeetaa egtaacaetg gaetegtga aeagaeegg gagtaeeget 780 etgteaateg etagggeegt acaggaggt eaeeacaeg etgtgteae agteegatg 440 aacaggataga ataateegaa atteeeega deecaage etgtgteae agagaeegg gagtaeeget 780 etgteaateg etagggeegt acaggaggt ecceaaaa ggaggataea tegetggag 960 etgteaateg etagggeegt acaggaggt ecceaaaa ggaggataea tegetggaeg 960 tttgaggget gtaacaaggt etacacaaaa agteeaee teaaggeega tegaegaeg 1020 catactgggg aaaacegt caaatgeee teggagagg geaeaetg taaatgeege 1020 catactgggg aaaacegt caaatgeae tegegagga geaegtgga atttgeegee 1080	cccggcagcc cgctggt	aag tatgeetgee a	agtgcc			1536
<400> SEQUENCE: 23 atgaacattc acatgaageg caagaegata aagaacatea atacattega gaacegaatg 60 ttgatgttgg atggeatgee egetgtaegg gtaaaaaceg ageteetgaa gtetgaacaa 120 ggatececaa acgteeacaa etaeceggat atggaggeag tgeegetett geteacaaat 180 gtgaagggag ageegeetga ggaetetete teegtagate atteeagae acagaetgag 240 ecegtagate ttteaattaa caaageeaga acateeeta etgeeggtaag tteetteee 300 gtaagtatga eageaagtge atetagteea agtetaga geactageag tteetteee 400 acggtaettg etagteeace aaeggtgate acaagtgtt etagegeeag eageagetea 420 acggtaetga eteeeggee aceegtgea ageetagtg gegtgggtg ceaacaatt 480 ecectatatta tteeecegg geeteegte eagteagta ateteecaga eageagete 540 agteeceetg ggaatgeea taataetata gttgteett tgettgaaga tggtagggge 660 catgggaaag eacagatgga ececeggee ttgteecaga gaagteeta ateegatag 720 gaegaeetga ettgeetaa egtaacaetg gaetetgga acgagaeegg gagtaeeget 780 edgagaagt attgeetaa egtaacaetg gaetetgga acgagaeegg gagtaeeget 780 edgagaaga ataateegaa atteeetg ageaegee ettgteaga aggegeage gagtaeege 900 eggagaeegg gaegagtga ateaceega eteeggae eacgagaee eteeggeeg 900 eggagaeeg gaegaagtga ateaceega eteeggae eteeggeegaa aggegeagae 1020 catactggg aaaaaeegta caaatgeee tgggaggat geeegtggaa atttgeeaga 1020 eatactggg aaaaaeegta caaatgeee tgggaggat geeegtggaa atttgeeage 1080 ttgaaggeeg gaegatttag eegetetgat caeetgeeg taaageegt taaategee 1080 etgaegagt tgaeeaegee etaecaaaa agteeaee teaageegat taaategee 1080 etgaegagat tgaeeaegee etaecaaaa agteeee teaageegat taaategee 1080 etgaegaga aaaaeegta caaatgeee tgggaggat geaegtggaa atttgeeaege 1020 etaactgggg aaaaaeegta caaatgeae tgggagge teaageegt taaategee 1080 etgaegagt tgaeeaegee etaecaaaa agteeee teaageeget taaategee 1080 etgaegagat tgaeeaegee etaecaaaa agteeee teaageegat taaategee 1080 etgaegagag aaaaaeegta caaatgeae tgggaggag geaegeega attgeegeegaa attgeegeegaa 200 etgaegaega aaaaeega eeeegaagee taaageegeegaa ageegeegaa ageege	<pre>&lt;210&gt; SEQ ID NO 23 &lt;211&gt; LENGTH: 1206 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Ar &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORM polynucleoti</pre>	tificial Sequend ATION: Descript: de	ce ion of Arti	ificial Sequ	ience: Synth	etic
atgaacattcacatgaagcgcaagacgataaagaacatcaatacattcgagaaccgaatg60ttgatgttggatggcatgcccgctgtacgggtaaaaaccgagctcctggagtctgaacaa120ggatccccaaacgtccacaactacccggatatggaggcagtgccgctcttgctcaacaat180gtgaagggagagccgcctgaggactctctctccgtagatcatttccagacacagactgga240cccgtagatctttcaattaacaaagccagaacatctcctactgcggtaagttctctccc300gtaagtatgacagcaagtgcatctagtccaagttctacgagcactagcagttcttctccc360agtgaacttgctagttcaccaacggtgatcacaagtgtgctagtcaag420acggtactgactcccggtccactcgtggcaagcgctgggcaacaattt480ctccatattattcaccccggcctcgtctagtccgatgacacaagtgg660cgatccctgggaatgtcaataatactatagttgtcctttgctgggg660cgatgcagaacagagtgaccccgcggcttgtcacgaaggaggg660cgagacgatgattgcctaacgtaacactggactctgtgaacgagaggg660catgggaaagcacagatggacccccgcggcttgtcactggaacgg720gacgacgatgattgcctaacgtaacactggactccgc780ctgtcaatcgctagggcgcaccgaacggagggatactcgctggacgaggagaaggacgaagggaatcaccgaagaggggatacatcgctggacacaggaggaacaaggaggaacaccgaagagggagaa960ctggaggggaaaaccgtacaaatgcacc<	<400> SEQUENCE: 23					
ttgatgttgg atggcatgcc cgctgtacgg gtaaaaaccg agctcctga gtctgaacaa 120 ggatccccaa acgtccacaa ctacccggat atggaggcag tgccgctctt gctcaacaat 180 gtgaagggag agccgcctga ggactctct tccgtagatc atttccagac acagactgag 240 cccgtagatc tttcaattaa caaagccaga acatcccta ctgcggtaag ttcttcctcc 300 gtaagtatga cagcaagtgc atctagtcca agttctacga gcactagcag ttcttcatct 360 agtagacttg ctagttcacc aacggtgatc acaagtgtt ctagcgccag cagcagctca 420 acggtactga ctcccggtcc actcgtggca agcgctagtg gcgtgggtgg ccaacaattt 480 ctccatatta ttcacccgt gcctcgtct agtccgatga atctcagag caacaagctt 540 agtcacgtac ataggatcc cgtcgtcgt cagtcagtt ccgtcgtcta cacagtgtg 660 cgatcccctg ggaatgtcaa taatactata gttgttcctt tgcttgagga tggtaggggc 660 catgggaaag cacagatgga cccccgcgc ttgtcaccga gacagtctaa atccgatagt 720 gacgacgatg atttgcctaa cgtaacactg gactctgtga acgagaccgg gagtaccgct 780 ctgtcaatcg ctagggccgt acaggaggtc caccaagcc ctgtgtcacg agtccgaggt 840 aacaggatga ataatcagaa atttccctgt agcaccagc ctgtgtcacg agtccgagg 960 ctgtgaagg gacgaagtga atcacccgac tccagaaaa ggaggataca tcgccgagg 1020 cgaacacgg gaaagtga ctacacaaa agttcacacc tcaaggcga attgcacg 1020 catactgggg aaaaaccgta caaatgcacc tgggaggga gcaccgt taaatgcacc 1080 tttgagggt tgacacgcca ctatcgaag catacgggg taaagcgt taaatgcact 1140 gattgcgaca ggagttttag ccgctctgat cacctgctc ttcaccgag gcgacacatg 1200	atgaacattc acatgaa	gcg caagacgata a	aagaacatca	atacattcga	gaaccgaatg	60
ggatccccaa acgtccacaa ctacccggat atggaggcag tgccgctctt gctcaacaat 180 gtgaagggag agccgcctga ggactctct tccgtagatc atttccagac acagactgag 240 cccgtagatc tttcaattaa caaagccaga acatctccta ctgcggtaag ttcttcccc 300 gtaagtatga cagcaagtgc atctagtcca agttctacga gcactagcag ttcttcatct 360 aqtagacttg ctagttcacc aacggtgatc acaagtgtt ctagcgccag cagcagctca 420 acggtactga ctcccggtcc actcgtggca agcgctagtg gcgtgggtgg ccaacaattt 480 ctccatatta ttcaccccgt gcctccgtct agtccgatga atctcagag caacaagctt 540 agtacccctg ggaatgtcaa taatactata gttgttcctt tgcttgagga tggtagggg 660 caggagaag cacagatgga cccccgcggc ttgtcaccga gacagtctaa atccgatagt 720 gacgacggatg atttgcctaa cgtaacactg gactctgtga acgagaccgg gagtaccgct 780 ctgtcaatcg ctagggccgt acaggaggtc cacccaagcc ctgtgtccac gaggggt 900 cggagacagc gacgaagtga atcacccgac tccagaaaaa ggaggataca tcgacggacg 960 tttgagggct gtaacaaggt ctacccaaaa agttcacacc tcaaggcga attgcagcg 1020 catactgggg aaaaaccgta caaatgcacc tgggaggat gcacgtgga attgcacgc 1020 catactgggg aaaaaccgta caaatgcacc tgggaggat gcacgtgga attgcacgc 1020 catactgggg agagtttag ccgctctgat cacctggcg taaagccgt taaatgcgct 1140 gattgcgaca ggagttttag ccgctctgat cacctgcc ttccaccgag gcgacacatg 1200	ttgatgttgg atggcat	gcc cgctgtacgg o	gtaaaaaccg	agctcctgga	gtctgaacaa	120
gtgaagggag agccgctga ggactetete teegtagate attteeagae acagaetgag240cccgtagate ttteaattaa caaagecaga acateteeta etgeggtaag ttetteetee300gtaagtatga cageaagtge atetagteea agttetaega geaetageag ttetteatet360agtagaettg etagtteace aaeggtgate acaagtgttt etagegeeag cageagetea420acggtaetga eteeeggtee actegtggea agegetagtg gegtgggtgg ecaacaattt480eteeetatta tteacecegt geeteegtet agteegatga ateteeagag eaacaagett540agteacgtae ataggateea eggetgete eagteagte eegtegtee eegtegtegte eegtegtegte eegtegtegte eegtegtegte660cqaggaaag eacagatgga eceeeggee ttgteacega gacagatega tggtagggge660catgggaaag eacagatgga eceeeggee ttgteacega gacagtetaa ateegatagt720gacgaeggatg atttgeetaa egtaacaetg gaetetgtga aegagaeegg gagtaeeget780etgteaateg etagggeegt acaggaggte eaceeagee etteeta agagteecae gaegaagtg840aacaggatga ataateagaa attteeetgt ageateagee ettetteta agagteecae900eggagaeage gaegaagtga ateaceegae teeagaaaa ggaggataea tegetggae960tttgaggget gtaacaaggt etaecaaaa agtteeacee teaaggeegaa attgeeggeegae1020catactgggg aaaaacegta caaatgeee teggaagga geacegtt taaatgegee 10201080tetgaegagt tgaeaegeea etaegaag eataeggeega taaageegtt taaatgeget 11401140gattgegaea gagttttag eegetetga eecetegae eacettgeee teteeseaga1200	ggatccccaa acgtcca	caa ctacccggat a	atggaggcag	tgeegetett	gctcaacaat	180
cccgtagate tttcaattaa caaagecaga acateteeta etgeggtaag ttetteece300gtaagtatga cagcaagtge atetagteea agttetaega geaetageag ttetteatet360agtagaettg etagtteece aaeggtgate acaagtgttt etagegeeag cagcagetea420acggtaetga eteeceggtee actegtggea agegetagtg gegtgggtgg ecaacaattt480eteecatatta tteececegt geeteegtee agteegatga ateteeagag caacaagett540agteacegtae ataggateee egeeggeeggeeggeggag eaaeagetgg600cgateeegg ggaatgteaa taataetata gttgtteett tgettgagga tggtagggge660catgggaaag eaeagatgga eeeeeggeeggeeggaggag gacagtetaa ateegatagt720gaeggaegatg atttgeetaa egtaacaetg gaetetgga acaggageegggaggggggggggggggggggggggg	gtgaagggag agccgcc	tga ggactctctc 1	tccgtagatc	atttccagac	acagactgag	240
gtaagtatgacagcaagtgaatctagtccaagttctacgagcactagcagttcttcatct360agtagacttgctagttcaccaacggtgatcacaagtgtttctagcgccagcagcagctca420acggtactgactcccggtccactcgtggcaaggcgtgggtggccaacaattt480ctccatattattcaccccgtgcctccgtctagtccgatgaatctccagagcaacagctgt540agtcacgtacataggatccccgtcgtcgtccagtcagttcccgtcgtctacacagctgtg600cgatcccctgggaatgtcaataatactatagttgttcctttgcttgaggatggtaggggc660catgggaaagcacagatggacccccgcggcttgtcaccgagacagtcag720gacgacgatgatttgcctaacgtaacactggactctgtgaacgagaccgggagtaccgct780ctgtcaatcgctagggccgtacaggaggtccacccaagccctgtgtcacggat900cggagacagegacgaagtgaatcacccgactccagaaaaggaggatacatcgctgtgacg960tttgagggctgtaacaaggtctacacaaaaagttcacacctcaaggcgcg1020catactggggaaaaccgtacaaatgcacctggagggg10201080tctgacgagttgacacgccactaccgaagcaccgggggtaaagccgtt1140gattgcgacaggagttttagccgctctgatcaccttgctcttcaccgggg1200	cccgtagatc tttcaat	taa caaagccaga a	acatctccta	ctgcggtaag	ttetteteee	300
agtagacttgctagttcaccaacggtgatcacaagtgtttctagcgccagcagcagctca420acggtactgactcccggtccactcgtggcaagcgctagtggcgtgggtggccaacaattt480ctccatattattcaccccgtgcctccgtctagtccgatgaatctccagagcaacaagctt540agtcacgtacataggatccccgtcgtcgtccagtcagttcccgtcgtctacacagctggg600cgatcccctgggaatgtcaataatactatagttgttcctttgcttgaggatggtaggggc660catgggaaagcacagatggacccccgcggcttgtcaccgagacagtcaaatccgatagt720gacgacgatgatttgcctaacgtaacactggactctgtgaacgagaccgggagtaccgct780ctgtcaatcgctagggccgtacaggaggtccacccaagccctgtgtcacggatccgaggt840aacaggatgaataatcagaaatttccctgtagcatcagcccattttctatagagtccact900cggaagacagcgacgaagtgaatcacccgactccagaaaaaggagggatacatcgctgtgac960tttgagggctgtaacaaggtctacacaaaaagttccacctcaagggcga1020catactggggaaaaaccgtacaaatgcacctggagggatgcacgtggaa1140gattgcgacaggagttttagctacctgaacactggcgtaaagccgt1140gattgcgacaggagttttagccgctctgatcaccttgctcttcaccggaggcaccacatg	gtaagtatga cagcaag	tgc atctagtcca a	agttctacga	gcactagcag	ttcttcatct	360
acggtactgactcccggtccactcgtggcaagcgctagtggcgtgggtggccacacattt480ctccatattattcaccccgtgcctccgtctagtccgatgaatctccagagcaacaagctt540agtcacgtacataggatccccgtcgtcgtccagtcagttcccgtcgtctacacagctgtg600cgatcccctgggaatgtcaataatactatagttgttcctttgcttgaggatggtaggggc660catgggaaagcacagatggacccccgcggcttgtcaccgagacagtcaaatccgatagt720gacgacgatgattgcctaacgtaacactggactctgtgaacgagaccgggagtaccgct780ctgtcaatcgctagggccgtacaggaggtccacccaagccctgtgtcacgg40aacaggatgaataatcagaaatttccctgtagcatcagccctgtgtcacg900cggaagacagcgacgaagtgaatcacccgactccagaaaaaggaggatacatcgctgtgac960tttgagggctgtaacaaggtctacacaaaaagttcacacctcaaggcgaa1020catactggggaaaaccgtacaaatgcacctgggaggat1020catactggggaaaaccgtactaccgaagcaccgggggat1080tctgacgagttgacacgccactatcgaaagcatacgggg taaagccgtttaaatgcgct1140gattgcgacaggagttttagccgctctgatcaccttgctcttcaccggaggcaccacatg1200	agtagacttg ctagttc	acc aacggtgatc a	acaagtgttt	ctagcgccag	cagcagctca	420
ctccatatta ttcaccccgt gcctccgtct agtccgatga atctccagag caacaagctt540agtcacgtac ataggatccc cgtcgtcgtc cagtcagttc ccgtcgtcta cacagctgtg600cgatcccctg ggaatgtcaa taatactata gttgttcctt tgcttgagga tggtaggggc660catgggaaag cacagatgga cccccgcggc ttgtcaccga gacagtctaa atccgatagt720gacgacgatg atttgcctaa cgtaacactg gactcgtga acgagaccgg gagtaccgct780ctgtcaatcg ctagggccgt acaggaggtc cacccaagcc ctgtgtcacg agtccgaggt840aacaggatga ataatcagaa atttccctgt agcatcagce cattttctat agagtccact900cggagacagc gacgaagtga atcacccgac tccagaaaaa ggaggataca tcgctgtgac960tttgagggct gtaacaaggt ctacacaaaa agttccacce tcaaggcga attgccacg1020catactgggg aaaaaccgta caaatgcacc tgggaggat gcacgtggaa attgccacg1080tctgacgagt tgacacgcca ctatcgaaag catacgggcg taaagccgt taaatgcgct1140gattgcgaca ggagttttag ccgctctgat caccttgctc ttcaccggag gcgacactg1200	acggtactga ctcccgg	tcc actcgtggca a	agcgctagtg	gcgtgggtgg	ccaacaattt	480
agtcacgtac ataggatece egtegtegte eagteagtte eegtegteta eacagetgtg 600 egateceetg ggaatgteaa taataetata gttgtteett tgettgagga tggtagggge 660 eatgggaaag eacagatgga eeeeegge ttgteacega gacagtetaa ateegatagt 720 gaegaeggtg atttgeetaa egtaacaetg gaetetgtga aegagaeegg gagtaeeget 780 etgteaateg etagggeegt aeaggaggte eaceeaagee etgtgteaeg agteegaggt 840 aacaggatga ataateagaa attteeetg ageateagee etgtgteaeg agteegaggt 900 eggagaeage gaegaagtga ateaeegae teeagaaaaa ggaggataea tegeetgtga 960 tttgaggget gtaacaaggt etaeaeaaa agtteaeaee teaaggeega tegaeggaeg 1020 eataetgggg aaaaaeegta eaaatgeaee tgggaggat geaegtggaa atttgeeege 1080 teegaegagt tgaeaegea etategaaag eataeggeeg taaageegtt taaatgeget 1140 gattgeegaea ggagtttag eegetetgat eaeettgete tteaeeggag gegaeaeatg 1200	ctccatatta ttcaccc	cgt gcctccgtct a	agtccgatga	atctccagag	caacaagctt	540
cgatcccctg ggaatgtcaa taatactata gttgttcctt tgcttgagga tggtaggggc 660 catgggaaag cacagatgga cccccgcggc ttgtcaccga gacagtctaa atccgatagt 720 gacgacgatg atttgcctaa cgtaacactg gactctgtga acgagaccgg gagtaccgct 780 ctgtcaatcg ctagggccgt acaggaggtc cacccaagcc ctgtgtcacg agtccgaggt 840 aacaggatga ataatcagaa atttccctgt agcatcagcc cattttctat agagtccact 900 cggagacagc gacgaagtga atcacccgac tccagaaaaa ggaggataca tcgctgtgac 960 tttgagggct gtaacaaggt ctacacaaaa agttccacce tcaaggcgca tcgacggacg 1020 catactgggg aaaaaccgta caaatgcacc tgggaggat gcacgtggaa atttgcacgc 1080 tctgacgagt tgacacgcca ctatcgaaag catacgggcg taaagccgtt taaatgcgct 1140 gattgcgaca ggagtttag ccgctctgat caccttgctc ttcaccggag gcgacacatg 1200	agtcacgtac ataggat	ccc cgtcgtcgtc (	cagtcagttc	ccgtcgtcta	cacagctgtg	600
catgggaaag cacagatgga cccccgcggc ttgtcaccga gacagtctaa atccgatagt 720 gacgacgatg atttgcctaa cgtaacactg gactctgtga acgagaccgg gagtaccgct 780 ctgtcaatcg ctagggccgt acaggaggtc cacccaagcc ctgtgtcacg agtccgaggt 840 aacaggatga ataatcagaa atttccctgt agcatcagcc cattttctat agagtccact 900 cggagacagc gacgaagtga atcacccgac tccagaaaaa ggaggataca tcgctgtgac 960 tttgagggct gtaacaaggt ctacacaaaa agttcacacc tcaaggcgca tcgacggacg 1020 catactgggg aaaaaccgta caaatgcacc tgggagggat gcacgtggaa atttgcacgc 1080 tctgacgagt tgacacgcca ctatcgaaag catacgggcg taaagccgtt taaatgcgct 1140 gattgcgaca ggagtttag ccgctctgat caccttgctc ttcaccggag gcgacacatg 1200	cgatcccctg ggaatgt	caa taatactata 🤉	gttgttcctt	tgcttgagga	tggtaggggc	660
gacgacgatg atttgcctaa cgtaacactg gactctgtga acgagaccgg gagtaccgct 780 ctgtcaatcg ctagggccgt acaggaggtc cacccaagcc ctgtgtcacg agtccgaggt 840 aacaggatga ataatcagaa atttccctgt agcatcagcc cattttctat agagtccact 900 cggagacagc gacgaagtga atcacccgac tccagaaaaa ggaggataca tcgctgtgac 960 tttgagggct gtaacaaggt ctacacaaaa agttcacacc tcaaggcgca tcgacggacg 1020 catactgggg aaaaaccgta caaatgcacc tgggaggat gcacgtggaa atttgcacgc 1080 tctgacgagt tgacacgcca ctatcgaaag catacgggcg taaagccgtt taaatgcgct 1140 gattgcgaca ggagttttag ccgctctgat caccttgctc ttcaccggag gcgacacatg 1200	catgggaaag cacagat	gga cccccgcggc 1	ttgtcaccga	gacagtctaa	atccgatagt	720
ctgtcaatcg ctagggccgt acaggaggte cacceaagee etgtgteaeg agteegaggt 840 aacaggatga ataateagaa attteeetg ageateagee cattteetat agagteeaet 900 eggagacage gaegaagtga ateaeegae teeagaaaaa ggaggataea tegetgtgae 960 tttgaggget gtaacaaggt etaeaaaaa agtteaeaee teaaggegea tegaeggaeg 1020 cataeegggg aaaaaeegta caaatgeaee tgggagggat geaegtggaa atttgeaege 1080 teegaegagt tgaeaegeea etategaaag eataegggeg taaageegtt taaatgeget 1140 gattgegaea ggagttttag eegetetgat eaeettgete tteaeeggag gegaeaetg 1200	gacgacgatg atttgcc	taa cgtaacactg 🤉	gactctgtga	acgagaccgg	gagtaccgct	780
aacaggatga ataatcagaa attteeetgt ageateagee cattteetat agagteeact 900 eggagacage gaegaagtga ateaecegae teeagaaaaa ggaggataca tegetgtgae 960 tttgaggget gtaacaaggt etaeaeaaaa agtteaeaee teaaggegea tegaeggaeg 1020 eataetgggg aaaaaeegta caaatgeaee tgggagggat geaegtggaa atttgeaege 1080 tetgaegagt tgaeaegeea etategaaag eataegggeg taaageegtt taaatgeget 1140 gattgegaea ggagtttag eegetetgat eaeettgete tteaeeggag gegaeaeatg 1200	ctgtcaatcg ctagggc	cgt acaggaggtc (	cacccaagcc	ctgtgtcacg	agtccgaggt	840
cggagacage gaegaagtga ateaecegae teeagaaaaa ggaggataca tegetgtgae 960 tttgaggget gtaacaaggt etaecaaaaa agtteaeaee teaaggegea tegaeggaeg 1020 cataetgggg aaaaaeegta caaatgeaee tgggagggat geaegtggaa atttgeaege 1080 tetgaegagt tgaeaegeea etategaaag eataegggeg taaageegtt taaatgeget 1140 gattgegaea ggagtttag eegetetgat eaeettgete tteaeeggag gegaeaeatg 1200	aacaggatga ataatca	gaa attteeetgt a	agcatcagcc	cattttctat	agagtccact	900
tttgaggget gtaacaaggt etaeaaaaa agtteaeaee teaaggegea tegaeggaeg 1020 eataetgggg aaaaaeegta caaatgeaee tgggagggat geaegtggaa atttgeaege 1080 tetgaegagt tgaeaegeea etategaaag eataegggeg taaageegtt taaatgeget 1140 gattgegaea ggagttttag eegetetgat eaeettgete tteaeeggag gegaeaeatg 1200	cggagacagc gacgaag	tga atcacccgac I	tccagaaaaa	ggaggataca	tcgctgtgac	960
catactgggg aaaaaccgta caaatgcacc tgggagggat gcacgtggaa atttgcacgc 1080 tctgacgagt tgacacgcca ctatcgaaag catacgggcg taaagccgtt taaatgcgct 1140 gattgcgaca ggagttttag ccgctctgat caccttgctc ttcaccggag gcgacacatg 1200	tttgagggct gtaacaa	ggt ctacacaaaa a	agttcacacc	tcaaggcgca	tcgacggacg	1020
tctgacgagt tgacacgcca ctatcgaaag catacgggcg taaagccgtt taaatgcgct 1140 gattgcgaca ggagttttag ccgctctgat caccttgctc ttcaccggag gcgacacatg 1200	catactgggg aaaaacc	gta caaatgcacc 1	tgggagggat	gcacgtggaa	atttgcacgc	1080
gattgegaca ggagttttag eegetetgat eacettgete tteaeeggag gegaeaeatg 1200	tctgacgagt tgacacg	cca ctatcgaaag d	catacgggcg	taaagccgtt	taaatgcgct	1140
	gattgcgaca ggagttt	tag ccgctctgat (	caccttgctc	ttcaccggag	gcgacacatg	1200

cttgtt	1206						
<210> SEQ ID NO 24 <211> LENGTH: 864 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide							
<400> SEQUENCE: 24							
atggetgegg etgeatatgt ggateatttt geggetgagt geetggtgte aatgtetagt	60						
agageggtgg tacaeggtee cagagaagge ceagaateae geeeagaggg egeegeegte	120						
gctgcaacac cgacgctgcc tcgggtcgag gagcgccgcg acgggaagga cagtgcgtca	180						
cttttcgtag tagcgagaat attggcagat ctgaatcaac aggctccagc acctgcgccc	240						
gctgaacgcc gggagggcgc cgctgccaga aaggccagaa caccatgccg cttgccgcca	300						
cctgcgccag aacccacaag tccaggtgcc gaaggtgcgg cggctgcccc tccttcaccg	360						
geetggtetg aaccagaace agaggeaggt ettgaacetg agegegaace eggeeetgea	420						
ggetetgggg aacetggeet gaggeagegg gtgaggegeg geeggageag ggeegaeetg	480						
gaatcaccgc aaaggaaaca taaatgccat tatgctggtt gcgaaaaggt ttatggaaag	540						
teateceace tgaaageaca eeteegeact cacaegggtg agegacettt tgegtgttee	600						
tggcaagact gcaataaaaa gtttgctaga tctgatgaac ttgcacggca ttatcgaact	660						
cataccggtg aaaagaagtt ctcatgccct atatgtgaga aacggttcat gcgctctgac	720						
cacttgacga aacatgcaag acgacatgct aattttcatc cggggatgtt gcagagacgg	780						
ggagggggaa gtaggactgg aagtetetee gaetatteee gateegaege tteeteacea	840						
acgattagee eegcaageag teee	864						
<210> SEQ ID NO 25 <211> LENGTH: 969 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide							
<400> SEQUENCE: 25							
atgtcagccg cagtcgcatg ccttgattac ttcgcggccg agtgtcttgt ttccatgtca	60						
gcggggggtg tcgttcacag aagaccacca gacccggagg gagcgggagg ggcagctgga	120						
tetgaagteg gegeggetee acetgaatea gegetteeeg geeetggtee teeaggteee	180						
gctagcgtgc cccaactccc acaagtgcct gctccgagtc ctggagcggg cggagcagcc	240						
ccgcatctcc ttgcagcatc agtgtgggcc gatcttcgcg gaagctccgg ggagggctcc	300						
tgggaaaaca geggagagge eeegegaget teaagegget ttteegatee aateeettge	360						
agtgttcaaa ccccatgctc cgagctcgcg cccgcgtccg gagctgcggc agtgtgcgca	420						
cctgaaagct catccgatgc gccggccgtt ccatctgcgc cagctgctcc cggtgcaccc	480						
gcagcatetg geggetttag tggtggaget ettggggegg gteeegeeee tgeggeggat	540						
caageteete geaggegeag tgttaegeee geageaaaae ggeateaatg eeeetteet	600						
ggttgtacaa aagcatacta taagtcatcc catctcaaga gtcaccagag gacgcataca	660						

#### -continued

ggtgagagac cttttagctg tgactggctc gattgcgaca agaaatttac gcggagc	gac 720
gaacttgcgc ggcactaccg cactcacact ggagaaaaga ggttctcttg tcccctg	tgt 780
cccaagcagt tctcacgcag tgatcacttg acaaaacatg ctaggagaca tccaaca	tac 840
catecegaea tgatagagta tegaggtagg egaegeaeae etagaattga teeteeg	ctg 900
actagtgaag tegagteaag tgeeagtgga ageggaeegg gteeegegee eteatt	aca 960
acctgtctt	969
<210> SEQ ID NO 26 <211> LENGTH: 1248 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: S polynucleotide	ynthetic
<400> SEQUENCE: 26	
atggtggacc acttacttcc agtggacgag aacttctcgt cgccaaaatg cccagtt	ggg 60
tatctgggtg ataggetggt tggeeggegg geatateaea tgetgeeete accegte	tct 120
gaagatgaca gegatgeete eageeeetge teetgtteea gteeegaete teaagee	ctc 180
tgctcctgct atggtggagg cctgggcacc gagagccagg acagcatctt ggacttc	cta 240
ttgtcccagg ccacgctggg cagtggcggg ggcagcggca gtagcattgg ggccagc	agt 300
ggccccgtgg cctggggggcc ctggcgaagg gcagcggccc ctgtgaaggg ggagcat	ttc 360
tgettgeeeg agttteettt gggtgateet gatgaegtee caeggeeett eeageet	acc 420
ctggaggaga ttgaagagtt tctggaggag aacatggagc ctggagtcaa ggaggtc	cct 480
gagggcaaca gcaaggactt ggatgcctgc agccagctct cagctgggcc acacaag	agc 540
cacetecate etgggtecag egggagagag egetgttece etceaceagg tggtgee	agt 600
gcaggaggtg cccagggccc aggtgggggc cccacgcctg atggccccat cccagtg	ttg 660
ctgcagatcc agcccgtgcc tgtgaagcag gaatcgggca cagggcctgc ctcccct	ggg 720
caagccccag agaatgtcaa ggttgcccag ctcctggtca acatccaggg gcagacc	ttc 780
gcactcgtgc cccaggtggt accetectee aacttgaace tgeeeteeaa gtttgtg	cgc 840
attgcccctg tgcccattgc cgccaagcct gttggatcgg gacccctggg gcctggc	cct 900
geeggtetee teatgggeea gaagtteeee aagaaceeag eegeagaact cateaaa	atg 960
cacaaatgta ctttccctgg ctgcagcaag atgtacacca aaagcagcca cctcaag	gcc 1020
cacctgcgcc ggcacacggg tgagaagccc ttcgcctgca cctggccagg ctgcggc	tgg 1080
aggttetege getetgaega getgtegegg caeaggeget egeaeteagg tgtgaag	ccg 1140
taccagtgtc ctgtgtgcga gaagaagttc gcgcggagcg accacctctc caagcac	atc 1200
aaggtgcacc gcttcccgcg gagcagccgc tccgtgcgct ccgtgaac	1248
<210> SEO ID NO 27	

<210> SEQ 1D NO 27 <211> LENGTH: 756 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence

<220> FEATURE:

 ${\scriptstyle <223>}$  OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 27

atgtcagccg cggtcgcgtg cgtggattat tttgcagcag atgtgctgat ggcaatttca	60
teeggtgeag tagtteateg eggaagaeea ggteetgagg gtgeggggee tgeggeeggg	120
ttggatgtte gegeegegeg cagggaagee getteteeeg gaacaeetgg eceteeteet	180
ceteegeegg eggeateagg eeegggteet ggtgeagetg eggeteetea eetgttggea	240
gcctccatac tggctgacct gcgagggggg ccaggcgctg cacctggtgg cgcgagtcca	300
gcaagttcca gctccgcggc gtcctccccg agtagtgggc gagctccggg cgcggcacct	360
tetgetgeeg etaaateaca eegatgeeet tteeeagaet gegegaagge gtattataag	420
tccagtcatt tgaaatcaca cttgaggaca cataccggcg agagaccttt tgcgtgcgac	480
tggcagggtt gtgataagaa atttgcgaga agcgacgaac tggcccgcca tcaccgcacc	540
cacacagggg aaaaaagatt ctcatgccca ctctgttcta agcgcttcac gcgaagcgac	600
catcttgcaa agcacgctag gagacaccct gggttccacc ccgacctctt gcgacgacct	660
ggcgcccggt ctactagccc gtctgactca ttgccgtgct ctctcgcagg gtcccctgct	720
ccgagccccg caccgtcccc agctcctgcc gggctt	756
<210> SEQ ID NO 28 <211> LENGTH: 1167 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synth polynucleotide	hetic
<400> SEQUENCE: 28	
atgtacggcc gaccgcaggc tgagatggaa caggaggctg gggagctgag ccggtggcag	60
gcggcgcacc aggctgccca ggataacgag aactcagcgc ccatcttgaa catgtcttca	120
tettetggaa getetggagt geacacetet tggaaceaag geetaeeaag eatteageae	180
tttcctcaca gcgcagagat gctggggtcc cctttggtgt ctgttgaggc gccggggcag	240
aatgtgaatg aagggggggcc acagttcagt atgccactgc ctgagcgtgg tatgagctac	300
tgcccccaag cgacteteac teetteeegg atgatttaet gteagagaat gteteeeet	360
cagcaagaga tgacgatttt cagtgggccc caactaatgc ccgtaggaga gcccaatatt	420
ccaagggtag ccaggccctt cggtgggaat ctaaggatgc cccccaatgg gctgccagtc	480
toggetteea etggaateee aataatgtee cacaetggga acceteeagt geettaeeet	540
ggcctctcga cagtacette tgacgaaaca ttgttgggee cgactgtgee ttecactgag	600
geccaggeag tgeteceete catggeteag atgttgeeee egeaagatge ceatgaeett	660
geccaggeag tgeteceete catggeteag atgttgeeee egeaagatge eeatgaeett gggatgeeee cagetgagte eeagteattg etggttttag gateteagga etetettgte	660 720
geccaggeag tgeteceete catggeteag atgttgeeee egeaagatge eeatgaeett gggatgeeee eagetgagte eeagteattg etggttttag gateteagga etetettgte agteageeag acteteaaga aggeeeattt etaeeagage ageeeggaee tgeteeacag	660 720 780
gcccaggcag tgctccctc catggctcag atgttgcccc cgcaagatg ccatgactt gggatgcccc cagctgagtc ccagtcattg ctggttttag gatctcagga ctctctgtc agtcagccag actctcaaga aggcccatt ctaccagag agcccggac tgctccacag acagtagaga agaactccag gcctcaggaa gggactggta gaaggggct ctcagaggca	660 720 780 840
gcccaggcag tgctcccct catggctcag atgttgccc cgcaagatg ccatgactt gggatgccc cagctgagt ccagtcatt ctggttttag gatctcagga ctcttgtc agtcagccag actetcaaga aggeccatt ctaccagage ageecggace tgetccaeag acagtagaga agaactecag geetcaggaa gggaetggta gaaggggete etcaagagga aggeettaet getgeaaeta egagaaetge ggaaaagett ataceaaaeg etceeaeet	660 720 780 840 900
gcccaggcag tgctcccct catggctcag atgttgccc cgcaagatg ccatgactt gggatgcccc cagctgagt ccagtcattg ctggtttag gatctcagga ctcttgtc agtcagccag actctcaaga aggcccatt ctaccagag agcccggac tgctccacag acagtagag agaactccag gcctcaggaa gggactggta gaaggggct ctcagaggca aggccttact gctgcaacta cgagaactg ggaaaagctt ataccaaacg ctcccacctc gtgagccacc agcgcaagca cacaggtgag aggccatat cttgcaactg ggaaagttg	660 720 780 840 900 960
gcccaggcag tgctcccct catggctag atgttgccc cgcaagatg ccatgactt gggatgccc cagctgagt ccagtcatt ctgctttag gatctcagga ctcttgtc agtcagccag actctcaaga aggcccatt ctaccagag aggcggac tgctcacag acagtagaga agaactccag gcctcaggaa gggactggta gaaggggct ctccagaggca aggccttact gctgcaacta cgagaactg ggaaagct ataccaaacg ctcccacctc gtgagccacc agcgcaagca cacaggtgag aggccatat cttgcaactg ggaaagttgt tcatggtct tcttccgtt tgatgagct agacgacaa tgcgggtaca cacagatat	660 720 780 840 900 960 1020
geccaggeag tgetececte eatggeteag atgttgeeee egeaagatge eeatgaeett gggatgeeee eagetgagte eeagteattg etggttttag gateteagga etetettge agteageeag acteteaga aggeeeatt etaecagage ageeeggaee tgeteeaeg acagtagaga agaacteeag geeteaggaa gggaetggta gaaggggete eteagaggea aggeettaet getgeaaeta egagaaetge ggaaaagett ataecaaaeg eteeeaee gtgageeaee agegeaagea eacaggtgag aggeeatat ettgeaaetg ggaaagttgt teatggtett tetteegtte tgatgagett agaegaeata tgegggtaea eaceagatat egaceatata aatgtgatea gtgeageegg gagtteatga ggtetgaeea teteaageaa	660 720 780 840 900 960 1020 1080
geccaggeag tgetececte eatggeteag atgttgeece egeaagatge eeatgaeett gggatgeeee eagetgagte eeageteatt etgetttag gateteagga eetettge agteageeag acteteaaga aggeeeatt etaecagage ageeeggaee tgeteeaeg acagtagaga agaaeteeag geeteaggaa gggaetggta gaaggggete eteeaggaea aggeettaet getgeaaeta egagaaetge ggaaaagett ataecaaaeg eteeeaeet gtgageeaee agegeaagea eaeaggtgag aggeeatatt ettgeaaetg ggaaagttgt teatggtett tetteegte tgatgageet agaeegaea tgeeggaea eaeagata egaeeatata aatgtgatea gtgeageegg gagtteatga ggtetgaeea teteaageaa caecagaaga eteateggee gggaeeetea gaeeeaeagg eeaaeaaeaa eaatggagag	660 720 780 840 900 960 1020 1080 1140

```
-continued
```

<210> SEQ ID NO 29 <211> LENGTH: 6 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide <400> SEQUENCE: 29 Gly Ser Gly Ser Gly Ser 1 <210> SEQ ID NO 30 <211> LENGTH: 59 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 30 gtctcgtggg ctcggagatg tgtataagag acagagaact atttcctggc tgttacgcg 59 <210> SEO ID NO 31 <211> LENGTH: 58 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE. <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 31 acactettte ectacaegae getetteega tetagaaeta ttteetgget gttaegeg 58 <210> SEQ ID NO 32 <211> LENGTH: 56 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 32 gactggagtt cagacgtgtg ctcttccgat cttgtcttcg ttgggagtga attagc 56 <210> SEQ ID NO 33 <211> LENGTH: 708 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 33 atggtgagca agggcgagga ggataacatg gccatcatca aggagttcat gcgcttcaag 60 gtgcacatgg agggctccgt gaacggccac gagttcgaga tcgagggcga gggcgagggc 120 cgcccctacg agggcaccca gaccgccaag ctgaaggtga ccaagggtgg ccccctgccc 180 ttcgcctggg acateetgte ceetcagtte atgtaegget ecaaggeeta egtgaageae 240 cccgccgaca tccccgacta cttgaagctg tccttccccg agggcttcaa gtgggagcgc 300 gtgatgaact tcgaggacgg cggcgtggtg accgtgaccc aggactcctc cctgcaggac 360

ggcgagttca	tctacaaggt	gaagctgcgc	ggcaccaact	tcccctccga	cggccccgta	420
atgcagaaga	agaccatggg	ctgggaggcc	tcctccgagc	ggatgtaccc	cgaggacggc	480
gccctgaagg	gcgagatcaa	gcagaggctg	aagctgaagg	acggcggcca	ctacgacgct	540
gaggtcaaga	ccacctacaa	ggccaagaag	cccgtgcagc	tgeeeggege	ctacaacgtc	600
aacatcaagt	tggacatcac	ctcccacaac	gaggactaca	ccatcgtgga	acagtacgaa	660
cgcgccgagg	gccgccactc	caccggcggc	atggacgagc	tgtacaag		708
<210> SEQ J <211> LENGJ <212> TYPE <213> ORGAN <220> FEATU <223> OTHEF polyr	D NO 34 H: 708 DNA HISM: Artif: RE: INFORMATIC NUCLEOTIDE	icial Sequer DN: Descript	nce tion of Art:	ificial Seq	lence: Synthet	ic
atggaggtgtt	atactaaaat	agentagge		2202200202	20000222002	60
caacaaccet	tectaceace	qqccqcatat	tttttcacaa	coqctactac	tqctqcaqcq	120
gcggcqqctq	ctgccqccac	gcaatccqcc	caacaqcaac	aacaacaaca	gcagcagcag	180
caacaaqcqc	ctcaacttcq	accegetqea	gacggqcaqc	cctcaqqqqq	agggcacaaq	240
agcgctccga	agcaggttaa	aaggcagagq	agcagtagtc	ccgaactgat	gcgatgtaag	300
aggcgcctca	atttagcgg	ttttggttac	tctttgcccc	agcagcagcc	ggctgccgta	360
gctcgccgaa	atgagcggga	aaggaaccgc	gttaaacttg	tgaatctcgg	tttcgcgaca	420
cttcgagagc	acgtaccaaa	tggggcagct	aacaagaaaa	tgagtaaagt	tgagacactg	480
cggtctgcag	tggagtatat	tagagctctt	caacaattgc	ttgacgagca	cgatgccgta	540
tcagccgcat	ttcaagccgg	ggtgctgtcc	ccaacaatat	ctccgaacta	cagcaatgat	600
cttaatagca	tggcgggaag	tcccgtttcc	tcctactcct	ctgatgaggg	cagctacgac	660
cctctcagtc	ccgaggagca	agagettett	gacttcacta	actggttc		708
<pre>&lt;210&gt; SEQ ID NO 35 &lt;211&gt; LENGTH: 573 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic</pre>						
<400> SEQUE	ENCE: 35					
atgatggaca	acagaggcaa	ctctagtcta	cctgacaaac	ttcctatctt	ccctgattct	60
gcccgcttgc	cacttaccag	gtccttctat	ctggagccca	tggtcacttt	ccacgtgcac	120
ccagaggccc	cggtgtcatc	tccttactct	gaggagctgc	cacggctgcc	ttttcccagc	180
gactctctta	tcctgggaaa	ttacagtgaa	ccctgcccct	tctctttccc	gatgccttat	240
ccaaattaca	gagggtgcga	gtactcctac	gggccagcct	tcacccggaa	aaggaatgag	300
cgggaaaggc	agcgggtgaa	atgtgtcaat	gaaggctacg	cccagctccg	acatcatctg	360
ccagaggagt	atttggagaa	gcgactcagc	aaagtggaaa	ccctcagagc	tgcgatcaag	420
tacattaact	acctgcagtc	tcttctgtac	cctgataaag	ctgagacaaa	gaataaccct	480
ggaaaagttt	cctccatgat	agcaaccacc	agccaccatg	ctgaccctat	gttcagaatt	540

-continued

gtttgeecaa etttettgta caaagttgte	ccc	573
<210> SEQ ID NO 36 <211> LENGTH: 516 <212> TYPE: DNA <213> ORGANISM: Artificial Seque <220> FEATURE: <223> OTHER INFORMATION: Descrip polynucleotide	nce tion of Artificial Sequence: Synthet	ic
<400> SEQUENCE: 36		
atggagacgc gtaaaccggc ggaacggctg	gcettgeeat actegetgeg cacegegeee	60
ctgggcgttc cggggaccct gcccggactc	ccgcggaggg accccctcag ggtcgccctg	120
cgtctggacg ccgcgtgctg ggagtgggcg	cgcagcggct gcgcacgggg atggcagtac	180
ttgcccgtgc cgctggacag cgccttcgag	cccgccttcc tccgcaagcg caacgagcgc	240
gagcggcagc gggtgcgctg cgtgaacgag	ggctatgege geeteegaga ceacetgeee	300
cgggagctgg cagacaagcg cctcagcaaa	. gtggagacgc teegegetge categactae	360
atcaagcacc tgcaggagct gctggagcgc	caggeetggg ggetegaggg egeggeegge	420
geegteeeec agegeaggge ggaatgeaac	agcgacgggg agtccaaggc ctcttcggcg	480
ccttcgccca gcagcgagcc cgaggagggg	ggcagc	516
<210> SEQ ID NO 37 <211> LENGTH: 833 <212> TYPE: DNA		
<pre></pre>	nce tion of Artificial Sequence: Synthet	ic
<pre>                       <!--</td--><td>nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct</td><td>ic 60</td></pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct	ic 60
<pre>       </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca	ic 60 120
<pre> ORGANISM: Artificial Seque &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Descrip polynucleotide &lt;400&gt; SEQUENCE: 37 atgccgatgg gggcagcaga aagagggtgct ggttcagaaa aggcggcaaa gagagggcca tctgatgtca cgtgcccgac tggtggtgat</pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg accttttgga	ic 60 120 180
<pre> ORGANISM: AFTIIICIAI Seque &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Descrip polynucleotide &lt;400&gt; SEQUENCE: 37 atgccgatgg gggcagcaga aagagggggc ggttcagaaa aggcggcaaa gagaggggcca tctgatgtca cgtgcccgac tggtggtgat ggtggtttag ctttagggcc tgcgcccaga</pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg accttttgga ggaacaatga ataataattt ctgcagggcc	ic 60 120 180 240
<pre>   </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg accttttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg	ic 60 120 180 240 300
<pre>  </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg accttttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg . ccccttggaa atgtaccttt cctcctatac	ic 60 120 180 240 300 360
<pre>       </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg accttttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg ccccttggaa atgtaccttt cctcctatac gcatatgctg gtgttttccc atatgtgcct	ic 60 120 180 240 300 360 420
<pre>  </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg accttttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg ccccttggaa atgtaccttt cctcctatac gcatatgctg gtgttttccc atatgtgcct ccttttgagc cggctttat ccaaaagagg	ic 60 120 180 240 300 360 420 480
<pre>       </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg accttttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg ccccttggaa atgtaccttt cctcctatac gcatatgctg gtgttttccc atatgtgcct ccttttgagc cggcttttat ccaaaagagg gtgaatgaag gatacgccag attgagaggc	ic 60 120 180 240 300 360 420 480 540
<pre>   </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg acctttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg ccccttggaa atgtaccttt cctcctatac gcatatgctg gtgttttccc atatgtgcct ccttttgagc cggctttat ccaaaagagg gtgaatgaag gatacgccag attgagaggc ttatccaaaag ttgaaacctt gagggcggca	ic 60 120 180 240 300 360 420 480 540 600
<pre>  </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg acctttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg ccccttggaa atgtaccttt cctcctatac gcatatgctg gtgttttccc atatgtgcct ccttttgagc cggctttat ccaaaagagg gtgaatgaag gatacgccag attgagaggc ttatcaaaag ttgaaaccct gagggcggca cctttcatcag cacctgatgg atcgacacca	ic 60 120 180 240 300 360 420 480 540 600
<pre>       </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg acctttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg ccccttggaa atgtaccttt cctcctatac gcatatgctg gtgttttccc atatgtgcct ccttttgagc cggcttttat ccaaaagagg gtgaatgaag gatacgccag attgagaggc ttatcaaaag ttgaaaccct gagggcggca cctttcatcag cacctgatgg atcgacacca ccatgccctg caccgcctgc tacaccaagg	ic 60 120 180 240 300 360 420 480 540 600 660 720
<pre>   </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg acctttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg ccccttggaa atgtaccttt cctcctatac gcatatgctg gtgttttccc atatgtgcct ccttttgagc cggctttat ccaaaagagg gtgaatgaag gatacgccag attgagaggc ttatcaaaag ttgaaaccct gagggcggca cctttcatcag cacctgatgg atcgacacca ccatgccctg caccgcctgc tacaccaagg gcaccttctt cccttgtcc tgaatcttct	ic 60 120 180 240 300 360 420 480 540 600 660 720 780
<pre>       </pre>	nce tion of Artificial Sequence: Synthet gggccccaat catctgcagc accatgggct tcaaaaagct ggtacccaag agctgctgca ggagctgacc caaaacctgg acctttgga ggaacaatga ataataattt ctgcagggcc tcatgtatgc aattaggtgt aatgccaccg ccccttggaa atgtaccttt cctcctatac gcatatgctg gtgttttccc atatgtgcct ccttttgagc cggctttat ccaaaagagg gtgaatgaag gatacgccag attgagaggc ttatcaaaag ttgaaaccct gagggcggca cctttcatcag cacctgatgg atcgacacca ccatgccctg caccgcctgc tacaccaagg gcaccttctt cccttgtccc tgaatcttct ttagaaagtg aagaatcctg gca	ic 60 120 180 240 300 360 420 480 540 600 660 720 780 833

<210> SEQ ID NO 38 <211> LENGTH: 1482 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence

<220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide						
<400> SEQU	ENCE: 38					
atgggagacg	acagaccgtt	tgtgtgcaat	gccccgggct	gtggacagag	atttacaaac	60
gaggaccacc	tggcagttca	taaacacaag	catgagatga	cattgaaatt	tggcccagcc	120
cgaactgact	cagtcatcat	tgcagatcaa	acgcctactc	caactagatt	cctgaagaac	180
tgtgaggagg	tgggactctt	caatgaacta	gctagctcct	ttgaacatga	attcaagaaa	240
gctgcagatg	aggatgagaa	aaaggcaaga	agcaggactg	ttgccaaaaa	actggtggct	300
gctgctgggc	cccttgacat	gtetetgeet	tccacaccag	acatcaaaat	caaagaagaa	360
gagccagtgg	aggtagactc	atccccacct	gatagccctg	cctctagtcc	ctgttcccca	420
ccactgaagg	agaaggaggt	taccccaaag	cctgttctga	tctctacccc	cacacccacc	480
attgtacgtc	ctggctccct	gcctctccac	ttgggctatg	atccacttca	tccaaccctt	540
ccctccccaa	cctctgtcat	cacacaggct	ccaccatcca	acaggcaaat	ggggtctccc	600
actggctccc	tccctcttgt	catgcatctt	gctaatggac	agaccatgcc	tgtgttgcca	660
gggcctccag	tacagatgcc	gtctgttata	tcgctggcca	gacctgtgtc	catggtgccc	720
aacattcctg	gtatccctgg	cccaccagtt	aacagtagtg	gctccatttc	tccctctggc	780
caccctatac	catcagaagc	caagatgaga	ctgaaagcca	ccctaactca	ccaagtctcc	840
tcaatcaatg	gtggttgtgg	aatggtggtg	ggtactgcca	gcaccatggt	gacagcccgc	900
ccagagcaga	gccagattct	catccagcac	cctgatgccc	catcccctgc	ccagccacag	960
gtctcaccag	ctcagcccac	ccctagtact	gggggggcgac	ggcggcgcac	agtagatgaa	1020
gatccagatg	agcgacggca	gcgctttctg	gagcgcaacc	gggctgcagc	ctcccgctgc	1080
cgccaaaagc	gaaagctgtg	ggtgtcctcc	ctagagaaga	aggccgaaga	actcacttct	1140
cagaacattc	agctgagtaa	tgaagtcaca	ttactacgca	atgaggtggc	ccagttgaaa	1200
cagctactgt	tagctcataa	agactgccca	gtcactgcac	tacagaaaaa	gactcaaggc	1260
tatttagaaa	gccccaagga	aagctcagag	ccaacgggtt	ctccagcccc	tgtgattcag	1320
cacagctcag	caacagcccc	tagcaatggc	ctcagtgttc	gctctgcagc	tgaagctgtg	1380
gccacctcgg	tcctcactca	gatggccagc	caaaggacag	aactgagcat	gccgatacaa	1440
tcgcatgtaa	tcatgacccc	acagtcccag	tctgcgggca	ga		1482
<210> SEQ ID NO 39 <211> LENGTH: 939 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide						
<400> SEQU	ENCE: 39					
atgtacgtga	gctacctcct	ggacaaggac	gtgagcatgt	accctagctc	cgtgcgccac	60
tctggcggcc	tcaacctggc	gccgcagaac	ttcgtcagcc	ccccgcagta	cccggactac	120
ggcggttacc	acgtggcggc	cgcagctgca	gcggcagcga	acttggacag	cgcgcagtcc	180
ccgggggccat	cctggccggc	agcgtatggc	gccccactcc	gggaggactg	gaatggctac	240
gcgcccggag	gegeegegge	cgccgccaac	gccgtggctc	acggcctcaa	cggtggctcc	300
96

ccggccgcag ccatgggcta cagcagcccc gcagactacc atccgcacca ccacccgcat 360 420 caccaccege accaccegge egecgegeet teetgegett etgggetget geaaaegete aaccccggcc ctcctgggcc cgccgccacc gctgccgccg agcagctgtc tcccggcggc 480 540 cageggegga acctgtgega gtggatgegg aageeggege ageagteett eggeageeaa gtgaaaacca ggacgaaaga caaatatcga gtggtgtaca cggaccacca gcggctggag 600 ctggagaagg agtttcacta cagtcgctac atcaccatcc ggaggaaagc cgagctagcc 660 gccacgctgg ggctctctga gaggcaggtt aaaatctggt ttcagaaccg cagagcaaag 720 gagaggaaaa tcaacaagaa gaagttgcag cagcaacagc agcagcagcc accacagccg 780 cctccgccgc caccacagec tccccagect cagecaggtc ctctgagaag tgtcccagag 840 cccttgagtc cggtgtcttc cctgcaagcc tcagtgtctg gctctgtccc tggggttctg 900 gggccaactg ggggggtgct aaaccccacc gtcacccag 939 <210> SEQ ID NO 40 <211> LENGTH: 897 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 40 atgatggcgt atatgaaccc ggggccccac tattctgtca acgccttggc cctaagtggc 60 cccagtgtgg atctgatgca ccaggctgtg ccctacccaa gcgcccccag gaagcagcgg 120 cgggagcgca ccaccttcac ccggagccaa ctggaggagc tggaggcact gtttgccaag 180 acccagtacc cagacgtcta tgcccgtgag gaggtggctc tgaagatcaa tctgcctgag 240 tccagggttc aggtttggtt caagaaccgg agggctaaat gcaggcagca gcgacagcag 300 cagaaacagc agcagcagcc cccaggggggc caggccaagg cccggcctgc caagaggaag 360 gegggeacgt ceceaagace etceacagat gtgtgtecag accetetggg cateteagat 420 tectacagte cecetetgee eggeceetea ggeteeceaa ceaeggeagt ggeeaetgtg 480 tccatctgga gcccagcctc agagtcccct ttgcctgagg cgcagcgggc tgggctggtg 540 gcotcagggo ogtototgao otcogococo tatgocatga octaogococo ggootcogot 600 ttetgetett eccecteege etatgggtet eegageteet attteagegg eetagaeeee 660 tacetttete ceatggtgee ceagetaggg ggeeeggete ttageeeeet etetggeeee 720 tccgtgggac cttccctggc ccagtccccc acctccctat caggccagag ctatggcgcc 780 tacageceeg tggatagett ggaatteaag gaeeeeaegg geaeetggaa atteaeetae 840 aatcccatgg accctctgga ctacaaggat cagagtgcct ggaagtttca gatcttg 897 <210> SEQ ID NO 41 <211> LENGTH: 1437 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 41

tgtgcctacg	gaacgccaca	cctggctaag	acagagatga	ccgcgtcctc	ctccagcgac	120
tatggacaga	cttccaagat	gagcccacgc	gtccctcagc	aggattggct	gtctcaaccc	180
ccagccaggg	tcaccatcaa	aatggaatgt	aaccctagcc	aggtgaatgg	ctcaaggaac	240
tctcctgatg	aatgcagtgt	ggccaaaggc	gggaagatgg	tgggcagccc	agacaccgtt	300
gggatgaact	acggcagcta	catggaggag	aagcacatgc	cacccccaaa	catgaccacg	360
aacgagcgca	gagttatcgt	gccagcagat	cctacgctat	ggagtacaga	ccatgtgcgg	420
cagtggctgg	agtgggcggt	gaaagaatat	ggccttccag	acgtcaacat	cttgttattc	480
cagaacatcg	atgggaagga	actgtgcaag	atgaccaagg	acgacttcca	gaggeteace	540
cccagctaca	atgccgacat	ccttctctca	catctccact	acctcagaga	gactcctctt	600
ccacatttga	cttcagatga	tgttgataaa	gccttacaaa	actctccacg	gttaatgcat	660
gctagaaaca	caggggggtgc	agcttttatt	ttcccaaata	cttcagtata	tcctgaagct	720
acgcaaagaa	ttacaactag	gccagattta	ccatatgagc	cccccaggag	atcagcctgg	780
accggtcacg	gccaccccac	gccccagtcg	aaagctgctc	aaccatctcc	ttccacagtg	840
cccaaaactg	aagaccagcg	tcctcagtta	gatccttatc	agattettgg	accaacaagt	900
agccgccttg	caaatccagg	cagtggccag	atccagcttt	ggcagttcct	cctggagctc	960
ctgtcggaca	gctccaactc	cagctgcatc	acctgggaag	gcaccaacgg	ggagttcaag	1020
atgacggatc	ccgacgaggt	ggcccggcgc	tggggagagc	ggaagagcaa	acccaacatg	1080
aactacgata	agctcagccg	cgccctccgt	tactactatg	acaagaacat	catgaccaag	1140
gtccatggga	agcgctacgc	ctacaagttc	gacttccacg	ggatcgccca	ggccctccag	1200
ccccaccccc	cggagtcatc	tctgtacaag	tacccctcag	acctcccgta	catgggctcc	1260
tatcacgccc	acccacagaa	gatgaacttt	gtggcgcccc	accctccagc	cctccccgtg	1320
acatcttcca	gttttttgc	tgccccaaac	ccatactgga	attcaccaac	tgggggtata	1380
taccccaaca	ctaggctccc	caccagccat	atgccttctc	atctgggcac	ttactac	1437
<210> SEQ 1 <211> LENG <212> TYPE <213> ORGAI <220> FEAT <223> OTHEI polyn	ID NO 42 TH: 1326 DNA VISM: Artif: RE: INFORMATIC NUCLEOTIDE NUCE: 42	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	aence: Synthe	etic
atatassa	aagatggagg	aattaattaa	agetattact	aattaataaa	asaassaat	60
tocagoooad	cetecetee	ggagagggtg	aaccaccacca	acactantaa	ctcttcacac	120
acceptacce	actacactto	aaccatcast	aacataaca	acquartte	atagaaaaat	120
atatassatt	atactact	aataaraat	ogtagest	tanganast	atotacters	100
GLOLACCOLL	elgetectat	eergggaggt	ayuyggcctg	Leaggaaact	yrargatgac	240
tgetecagea	ccattgttga	agatecceag	accaagtgtg	aatacatgct	caactcgatg	300
cccaagagac	tgtgtttagt	gtgtggtgac	atcgcttctg	ggtaccacta	tggggtagca	360
tcatgtgaag	cctgcaaggc	attcttcaag	aggacaattc	aaggcaatat	agaatacagc	420
tgccctgcca	cgaatgaatg	tgaaatcaca	aagcgcagac	gtaaatcctg	ccaggettge	480
cgcttcatga	agtgtttaaa	agtgggcatg	ctgaaagaag	gggtgcgtct	tgacagagta	540

-continued	
cgtggaggtc ggcagaagta caagcgcagg atagatgcgg agaacagccc atacctgaac	600
cctcagctgg ttcagccagc caaaaagcca ttgctctggt ctgatcctgc agataacaag	660
attgtctcac atttgttggt ggctgaaccg gagaagatct atgccatgcc	720
gtccccgaca gtgacatcaa agccctcact acactgtgtg acttggccga ccgagagttg	780
gtggttatca ttggatgggc gaagcatatt ccaggettet ccaegetgte eetggeggae	840
cagatgagcc ttctgcagag tgcttggatg gaaattttga tccttggtgt cgtataccgg	900
tetetttegt ttgaggatga acttgtetat geagaegatt atataatgga egaagaeeag	960
tccaaattag caggcettet tgatetaaat aatgetatee tgeagetggt aaagaaatae	1020
aagagcatga agctggaaaa agaagaattt gtcaccctca aagctatagc tcttgctaat	1080
tcagactcca tgcacataga agatgttgaa gccgttcaga agcttcagga tgtcttacat	1140
gaagegetge aggattatga agetggeeag cacatggaag accetegteg agetggeaag	1200
atgetgatga caetgecaet eetgaggeag acetetaeea aggeegtgea geattetae	1260
aacatcaaac tagaaggcaa agtcccaatg cacaaacttt ttttggaaat gttggaggcc	1320
aaggtc	1326
<210> SEQ ID NO 43 <211> LENGTH: 1110 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synth polynucleotide <400> SEQUENCE: 43	netic
atggatettt ggaaetggga tgaagettee eeteaagaag tteeeeegg aaataaaete	60
gcggggcttg gaagactccc tcgccttccg caacgcgtct ggggcggatg ccctggtgga	120
gcctcagcgg acccaaaccc tttgtctcca gcggagggggg caaagttggg tttctgcttc	180
ccggatettg etttgeaagg egataeteea aeggegaegg eagagaeetg ttggaaagge	240
accagtaget ecctggeeag ettteegeag etegattggg ggteageeet tetecateee	300
gaagtteeet gggggggggga accegaetee caageeette eetggagtgg tgattggaea	360
gatatggcat gcacageetg ggacagttgg teeggggegt cacagacatt gggaceagee	420
ccacttggac cggggcctat ccccgcagca ggaagcgaag gagctgctgg tcagaactgt	480
gtgcccgtgg ctggtgaggc taccagttgg tccagggccc aggcagcagg cagtaacacc	540
agetgggatt geteagtggg geetgaeggg gataettatt ggggetetgg tettggtgga	600
gaaccgagaa cggactgtac gataagttgg ggcggtccag ctgggcctga ttgtactacg	660
tcatggaatc ctggcttgca cgccggcggc acgacaagcc ttaagagata tcaaagttca	720
geeettacag tttgetcaga acetteeceg caaagtgace gagegteact ggegegatgt	780
cctaaaacta atcatcgagg gccgatccag ttgtggcagt ttttgcttga actccttcac	840
gatggcgcga ggagcagttg catcagatgg accggtaaca gcagggagtt ccaattgtgt	900
gaccccaagg aagtggctcg actgtggggt gagcgcaaac ggaagcctgg tatqaattac	960
gaaaagttga gtaggggttt gcgatattac tataqqcqcq acatcqttcq aaaqtccqqt	1020
qqtcqaaaqt acacatacaq attcqqcqqt cqcqtaccat ctcttqcata ccctgattgc	1080
	1110

-continued

<210> SEQ ID NO 44 <211> LENGTH: 1356 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 44 atggacggga ctattaagga ggctctgtcg gtggtgagcg acgaccagtc cctctttgac 60 tcagcgtacg gagcggcagc ccatctcccc aaggccgaca tgactgcctc ggggagtcct 120 180 qactacqqqc aqccccacaa qatcaacccc ctcccaccac aqcaqqaqtq qatcaatcaq 240 ccaqtqaqqq tcaacqtcaa qcqqqaqtat qaccacatqa atqqatccaq qqaqtctccq gtggactgca gcgttagcaa atgcagcaag ctggtgggcg gaggcgagtc caaccccatg 300 aactacaaca gctatatgga cgagaagaat ggcccccctc ctcccaacat gaccaccaac 360 qaqaqqaqaq tcatcqtccc cqcaqacccc acactqtqqa cacaqqaqca tqtqaqqcaa 420 tqqctqqaqt qqqccataaa qqaqtacaqc ttqatqqaqa tcqacacatc ctttttccaq 480 aacatggatg gcaaggaact gtgtaaaatg aacaaggagg actteeteeg egecaceace 540 ctctacaaca cggaagtgct gttgtcacac ctcagttacc tcagggaaag ttcactgctg 600 geetataata caaceteeca caeegaecaa teeteaegat tgagtgteaa agaagaeeet 660 tettatgaet eagteagaag aggagettgg ggeaataaca tgaattetgg eeteaacaaa 720 agtcctcccc ttggaggggc acaaacgatc agtaagaata cagagcaacg gccccagcca 780 gatecgtate agatectggg ecegaecage agtegeetag ecaaecetgg aagegggeag 840 atccagctgt ggcaatteet eetggagetg eteteegaea gegeeaaege eagetgtate 900 acctgggagg ggaccaacgg ggagttcaaa atgacggacc ccgatgaggt ggccaggcgc 960 tggggcgagc ggaaaagcaa gcccaacatg aattacgaca agctgagccg ggccctccgt 1020 tattactatg ataaaaacat tatgaccaaa gtgcacggca aaagatatgc ttacaaattt 1080 gacttecaeg geattgeeca ggetetgeag ceaeateega eegagtegte eatgtaeaag 1140 taccettetg acateteeta catgeettee taccatgeee accageagaa ggtgaaettt 1200 gtccctcccc atccatcctc catgcctgtc acttcctcca gcttctttgg agccgcatca 1260 caatactgga cctcccccac gggggggaatc taccccaacc ccaacgtccc ccgccatcct 1320 aacacccacg tgccttcaca cttaggcagc tactac 1356 <210> SEQ ID NO 45 <211> LENGTH: 1416 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 45 atgttgggca ccgtgaagat ggaggggcat gagacaagcg actggaattc ctactacgcg 60 qatacccaaq aaqcqtattc ttcaqttccc qtaaqcaata tqaactccqq attqqqqqqc 120 atgaatagta tgaacacgta tatgacaatg aatacgatga ccaccagcgg caacatgaca 180 ccggcctcct ttaatatgtc atatgcgaac cctggtcttg gcgctggcct ctcaccaggt 240

gcggtcgctg	gaatgcccgg	ggggagcgcc	ggagcgatga	actccatgac	cgctgcgggc	300
gtgacggcca	tgggtacggc	cctgtcaccc	agtggaatgg	gagctatggg	ggcccagcaa	360
gccgcctcaa	tgaatggatt	ggggccctat	gccgcggcga	tgaatccctg	catgtcccct	420
atggcttatg	cccccagcaa	tttgggtcgc	agtagagccg	gcggtggtgg	cgatgccaaa	480
accttcaagc	gaagttatcc	tcatgcgaag	cctccttatt	catatatatc	cttgattacg	540
atggcgatac	agcaggcccc	gtctaagatg	ctgactctga	gtgagatata	ccagtggatc	600
atggaccttt	ttccttacta	ccggcaaaac	caacagagat	ggcaaaactc	aatacgccat	660
agcettteet	tcaatgattg	ctttgtcaaa	gtcgctcgga	gccctgacaa	gcccggtaaa	720
gggtcctatt	ggacccttca	tccagatagc	ggcaatatgt	tcgagaatgg	ttgttatctt	780
agacggcaga	aacgattcaa	atgtgagaaa	cagccaggtg	ccggcggtgg	tggcggcagc	840
ggttcaggcg	gaagtggtgc	caagggtggg	cctgagtcta	gaaaagaccc	cagcggagca	900
agcaatccaa	gcgcggactc	tcccctgcac	cgcggtgttc	atggtaagac	aggtcagctt	960
gaggggggggg	ctgctccagg	cccggctgcg	tcaccgcaaa	cactggacca	tagtggagct	1020
acagcgaccg	gaggtgcttc	agaactcaag	acgcctgcgt	cctccactgc	gcctccgatc	1080
tccagtggtc	ccggtgcact	tgcctctgtt	cctgcatctc	atccagcaca	cggactcgcg	1140
ccgcacgagt	cccagctcca	tttgaaaggg	gacccacact	acagctttaa	ccacccattc	1200
tctattaaca	atttgatgtc	atcctcagaa	cagcagcata	aactcgactt	caaagcctat	1260
gaacaggccc	tgcagtattc	tccatatggc	tctacacttc	ctgcttctct	tccattgggg	1320
tctgcaagtg	tgacaacgcg	ctccccaatc	gagccaagtg	ccctcgagcc	tgcttattat	1380
caaggagtat	attcccgacc	agttttgaat	acaagt			1416
<210> SEQ : <211> LENG' <212> TYPE <213> ORGAI <220> FEATU <223> OTHEI polyu	ID NO 46 IH: 1374 : DNA NISM: Artif: JRE: R INFORMATI( nucleotide	icial Sequer DN: Descript	nce tion of Arti	ificial Sequ	uence: Synth	etic
<400> SEQUI	ENCE: 46					
atgctgggag	cggtgaagat	ggaagggcac	gagccgtccg	actggagcag	ctactatgca	60
gagcccgagg	gctactcctc	cgtgagcaac	atgaacgccg	gcctggggat	gaacggcatg	120
aacacgtaca	tgagcatgtc	ggcggccgcc	atgggcagcg	gctcgggcaa	catgagcgcg	180
ggctccatga	acatgtcgtc	gtacgtgggc	gctggcatga	gcccgtccct	ggcggggatg	240
tcccccggcg	cgggcgccat	ggcgggcatg	ggcggctcgg	ccggggggggc	tggcgtggcg	300
ggcatggggc	cgcacttgag	tcccagcctg	agcccgctcg	ggggggcaggc	ggccgggggcc	360
atgggcggcc	tggcccccta	cgccaacatg	aactccatga	gccccatgta	cgggcaggcg	420
ggcctgagcc	gcgcccgcga	ccccaagacc	tacaggcgca	gctacacgca	cgcaaagccg	480
ccctactcgt	acatctcgct	catcaccatg	gccatccagc	agageeecaa	caagatgctg	540
acgctgagcg	agatctacca	gtggatcatg	gacctcttcc	ccttctaccg	gcagaaccag	600
cagcgctggc	agaactccat	ccgccactcg	ctctccttca	acgactgttt	cctgaaggtg	660
ccccgctcgc	ccgacaagcc	cggcaagggc	tccttctgga	ccctgcaccc	tgactcgggc	720

COD	÷	п.	n	11	Δ	<u>a</u>	
COIL	-	-	τ.	u	$\sim$	a	

	-continued	
aacatgttcg agaacggctg ctacctgcgc cg	gccagaagc gcttcaagtg cgagaagcag	780
ctggcgctga aggaggccgc aggcgccgcc gg	gcagcggca agaaggcggc cgccggggcc	840
caggceteae aggeteaaet eggggaggee ge	ccgggccgg cctccgagac tccggcgggc	900
accgagtege etcactegag egecteeeeg tg	gccaggagc acaagcgagg gggcctggga	960
gagetgaagg ggaegeegge tgeggegetg ag	geceeccag ageeggegee eteteeeggg	1020
cagcagcagc aggccgcggc ccacctgctg gg	gecegeeee accaeceggg eetgeegeet	1080
gaggcccacc tgaagccgga acaccactac go	cetteaace accegttete cateaacaac	1140
ctcatgtcct cggagcagca gcaccaccac ag	gecaceace accaceagee ceacaaaatg	1200
gacctcaagg cctacgaaca ggtgatgcac ta	accccggct acggttcccc catgcctggc	1260
agettggeea tgggeeeggt caegaacaaa ac	cgggcctgg acgcctcgcc cctggccgca	1320
gataceteet actaceaggg ggtgtactee eg	ggcccatta tgaactcctc tttg	1374
<210> SEQ ID NO 47 <211> LENGTH: 1050 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description polynucleotide	e on of Artificial Sequence: Synth	etic
<400> SEQUENCE: 47		
atgctgggct cagtgaagat ggaggcccat ga	acctggccg agtggagcta ctacccggag	60
gcgggcgagg tetactegee ggtgaeeeea gt	geecacca tggeeceet caacteetae	120
atgaccetga atcetetaag etetecetat ee	ceeetgggg ggeteeetge eteeeeactg	180
ccctcaggac ccctggcacc cccagcacct go	cageeeece tggggeeeac ttteeeagge	240
ctgggtgtca gcggtggcag cagcagctcc gg	gtacgggg ccccgggtcc tgggctggtg	300
cacgggaagg agatgccgaa ggggtatcgg cg	ggcccctgg cacacgccaa gccaccgtat	360
tectatatet caeteateae catggeeate ca	agcaggcgc cgggcaagat gctgaccttg	420
agtgaaatet accagtggat catggaeete tt	cccttact accgggagaa tcagcagcgc	480
tggcagaact ccattegeca etegetgtet tt	ccaacgact gcttcgtcaa ggtggcgcgt	540
tccccagaca agcctggcaa gggctcctac tg	gggccctac accccagctc agggaacatg	600
tttgagaatg gctgctacct gcgccgccag aa	aacgcttca agctggagga gaaggtgaaa	660
aaaggggggca gcgggggtgc caccaccacc ag	ggaacggga cagggtctgc tgcctcgacc	720
accacccccg cggccacagt cacctccccg cc	cccageeee egeeteeage eeetgageet	780
gaggcccagg gcggggaaga tgtgggggct ct	tggactgtg getcaccege tteetceaca	840
ccctatttca ctggcctgga gctcccaggg ga	agetgaage tggaegegee etacaaette	900
aaccaccctt tctccatcaa caacctaatg to	cagaacaga caccagcacc tcccaaactg	960
gacgtggggt ttgggggcta cggggctgaa gg	ytggggagc ctggagtcta ctaccagggc	1020
ctctattccc gctctttgct taatgcatcc		1050
<210> SEQ ID NO 48 <211> LENGTH: 342		

 ${\scriptstyle <223>}$  OTHER INFORMATION: Description of Artificial Sequence: Synthetic

polynucleotide	
<400> SEQUENCE: 48	
atgatgcaag aatctgggac tgagacaaaa agtaacggtt cagccatcca gaatgggtcg	60
ggcggcagca accacttact agagtgcggc ggtcttcggg aggggggggtc caacggagag	120
acgeeggeeg tggacategg ggeagetgae etegeeeaeg eeeageagea geageaaeag	180
tggcatetea taaaccatea geeetetagg agteeeagea gttggettaa gagaetaatt	240
tcaagccctt gggagttgga agtcctgcag gtccccttgt ggggagcagt tgctgagacg	300
aagatgagtg gacctgtgtg tcagcctaac ccttccccat tt	342
<210> SEQ ID NO 49 <211> LENGTH: 1005 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	tic
<400> SEQUENCE: 49	
atggagttee etggeetggg gteeetgggg aceteagage eeeteeeea gtttgtggat	60
cctgctctgg tgtcctccac accagaatca ggggttttct tcccctctgg gcctgagggc	120
ttggatgcag cagetteete caetgeeeeg ageacageea eegetgeage tgeggeaetg	180
gcctactaca gggacgctga ggcctacaga cactccccag tctttcaggt gtacccattg	240
ctcaactgta tggaggggat cccagggggc tcaccatatg ccggctgggc ctacggcaag	300
acggggctct accetgeete aaetgtgtgt eecaeeegeg aggaetetee teeceaggee	360
gtggaagatc tggatggaaa aggcagcacc agcttcctgg agactttgaa gacagagcgg	420
ctgagcccag acctectgae eetgggaeet geaetgeett eateacteee tgteeceaat	480
agtgettatg ggggeeetga etttteeagt acettettt eteceaeegg gageeeeete	540
aattcagcag cotattooto toocaagott ogtggaacto toocootgoo toootgtgag	600
gccagggagt gtgtgaactg cggagcaaca gccactccac tgtggcggag ggacaggaca	660
ggccactacc tatgcaacgc ctgcggcctc tatcacaaga tgaatgggca gaacaggccc	720
ctcatccggc ccaagaagcg cctgattgtc agtaaacggg caggtactca gtgcaccaac	780
tgccagacga ccaccacgac actgtggcgg agaaatgcca gtggggatcc cgtgtgcaat	840
geetgeggee tetactacaa getacaceae cageactaet gtggtggete egeteagete	900
atgagggcac agagcatggc ctccagagga ggggtggtgt ccttctcctc ttgtagccag	960
aattetggae aacceaagte tetgggeeee aggeaceee tgget	1005
<210> SEQ ID NO 50 <211> LENGTH: 1440 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	tic
<400> SEQUENCE: 50	
atggaggtgg cgccggagca gccgcgctgg atggcgcacc cggccgtgct gaatgcgcag	60
caccccgact cacaccaccc gggcctggcg cacaactaca tggaacccgc gcagctgctg	120

- 0	ront	ing	ed
- (	JOILC	TITC	eu

cctccagacg aggtggacgt cttcttcaat cacctcgact cgcagggcaa cccctactat	180	
gecaaeeeeg eteaegegeg ggegegegte teetaeagee eegegeaege eegeetgaee	240	
ggaggccaga tgtgccgccc acacttgttg cacagcccgg gtttgccctg gctggacggg	300	
ggcaaagcag ccctctctgc cgctgcggcc caccaccaca acccctggac cgtgagcccc	360	
ttetecaaga egecaetgea ecceteaget getggaggee etggaggeee actetetgtg	420	
tacccagggg ctgggggtgg gagcggggga ggcagcggga gctcagtggc ctccctcacc	480	
cctacagcaa cccactctgg ctcccacctt ttcggcttcc cacccacgcc acccaaagaa	540	
gtgteteetg accetageae caeggggget gegteteeag eeteatette egeggggggt	600	
agtgcagccc gaggagagga caaggacggc gtcaagtacc aggtgtcact gacggagagc	660	
atgaagatgg aaagtggcag teeeetgege eeaggeetag etaetatggg eaceeageet	720	
gctacacacc accccatccc cacctacccc tcctatgtgc cggcggctgc ccacgactac	780	
agcageggae tettecacee eggaggette etgggggggae eggeeteeag etteaceeet	840	
aagcagcgca gcaaggctcg ttcctgttca gaaggccggg agtgtgtcaa ctgtggggcc	900	
acageeacee etetetggeg gegggaegge aceggeeaet acetgtgeaa tgeetgtgge	960	
ctctaccaca agatgaatgg gcagaaccga ccactcatca agcccaagcg aagactgtcg	1020	
gccgccagaa gagccggcac ctgttgtgca aattgtcaga cgacaaccac caccttatgg	1080	
cgccgaaacg ccaacgggga ccctgtctgc aacgcctgtg gcctctacta caagctgcac	1140	
aatgttaaca ggccactgac catgaagaag gaagggatcc agactcggaa ccggaagatg	1200	
tccaacaagt ccaagaagag caagaaaggg gcggagtgct tcgaggagct gtcaaagtgc	1260	
atgcaggaga agtcatecee ettcagtgea getgeeetgg etggacacat ggeaeetgtg	1320	
ggccacctcc cgcccttcag ccactccgga cacatcctgc ccactccgac gcccatccac	1380	
ccctcctcca gcctctcctt cggccacccc cacccgtcca gcatggtgac cgccatgggc	1440	
<210> SEQ ID NO 51 <211> LENGTH: 1326 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthes polynucleotide	tic	
<400> SEQUENCE: 51		
atgtaccaga geetggetat ggetgetaat eatggaeete eeetggage etatgaagee	60	
ggaggacetg gegettttat geatggaget ggegeegett etteteeegt gtatgtgeet	120	
acacctagag tgcccagcag cgtgctgggc ctttcttatc ttcagggagg aggagcagga	180	
tetgettetg geggagette aggeggatet tetggaggeg etgetteagg tgetggaeet	240	
ggaactcaac agggatctcc tggatggtca caggcaggag ctgatggagc cgcttatacc	300	
cctcctcctg tgagccccag gtttagcttt cctggcacaa caggctcttt agctgccgct	360	
gctgctgcag ccgcagctag agaagcagct gcatattcta gtggcggagg agctgctgga	420	
gccggcttag ctggaagaga gcagtacgga agagccggat ttgccggaag ctatagcagc	480	
cettaceetg eetatatgge egatgttgge geatettggg eageegeege ageagettet	540	
gcaggacett ttgaeteace tgtgetteae tetetgeetg geagagetaa teetgeegee	600	
agacateeea aeetggacat gttegaegae tteagegagg geagagaatg egtgaaetge	660	

#### -continued

ggagccatga gcacccccct ttggagaaga gacggcaccg gccactacct ttgcaatgcc 720 tgtggcctgt accacaagat gaacggcatc aacagacccc tgatcaagcc ccagagaaga 780 ctgagcgcta gcagaagagt gggcctgtcc tgcgccaatt gccagaccac aaccaccaca 840 900 ctgtggagga gaaatgccga gggcgagcct gtgtgtaacg cctgtggact gtacatgaag ctgcacggcg tgcccagacc tctggccatg agaaaggagg gcatccagac cagaaagaga 960 aageecaaga acetgaacaa gageaagaee eeegetgete ettetggaag egagageetg 1020 cctccagcct ctggagccag cagcaatagc tctaacgcca ccacatette ttetgaggag 1080 atgaggeeca teaaaacega geeaggeetg ageageeact aeggeeaeag etetagegtg 1140 agecagaett ttagegtgte tgecatgtea ggecaeggae etageattea ceetgtgetg 1200 agegeeetga agttgageee acagggetat getteteetg tgteteagag ceeteagaee 1260 tocaqcaaqc aqqacteetq qaattetetq qtqctqqeeq acaqceacqq eqatateate 1320 1326 accqcc <210> SEO ID NO 52 <211> LENGTH: 1785 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 52 atggccctga ccgacggcgg atggtgtctc cctaaaagat tcggcgccgc tggcgctgat 60 gcttctgaca gcagagcctt ccccgctagg gaacccagca caccacctag ccccatcagc 120 agetcaaget ctagetgtag cagaggegga gagagaggae etggaggege ttetaactge 180 ggcacacete agetggatae agaageegee geeggaeeae eageeagate tetttaett 240 agcagetaeg ceageeacee tittggeget ceteatggae eetetgetee tggtgtggee 300 ggacctggcg gaaacctgag ctcttgggag gaccttctgc tgtttaccga cctggaccag 360 420 gctgccaccg ctagcaagct tctgtggagc agcaggggggg ctaagctgag cccttttgcc cctgagcagc ccgaggagat gtaccagacc ctggctgctt taagctctca gggacctgcc 480 gettatgacg gageceetgg tggatttgtt caeteagegg cageageege agetgetgea 540 gccgctgcca gctcacctgt gtatgtgcct accacaagag tgggcagcat gttacctgga 600 cttccttacc atctgcaggg cagcggaagc ggccctgcta accatgccgg aggagctgga 660 720 geteacceeq gatggeetea ggettetgea gatteteete ettatggate tggaggagga gcagctggag ggggagctgc aggaccaggt ggagccggaa gcgcagcagc acatgtgtct 780 gccagatttc cctatagccc tagccctcct atggccaatg gcgctgctag agaacccgga 840 ggatatgetg eggeaggete tggeggeget ggeggagttt etggaggtgg atetteaetg 900 960 geogetatgg gaggaagaga geoteagtae tettetetga gegeogetag accaetgaae ggcacctatc atcaccacca ccatcaccat catcatcacc ccagccctta ctccccttat 1020 gtgggagccc cccttacacc cgcttggcct gccggccctt tcgagacacc tgtgctgcac 1080 ageetteagt ctagagetgg egeacettta ceagtgeeta gaggeeeete tgeegaettg 1140 ctggaggatc tgagcgagag cagagagtgc gtgaactgtg gcagcatcca gacacccctg 1200

-continued	
tggagaagag acggcaccgg ccactacctg tgcaacgctt gcggcctgta cagcaagatg	1260
aatgggetga geagaeeeet gateaageee cagaagaggg tgeeeageag cagaeggetg	1320
ggactgagct gcgccaactg tcataccaca acaaccacac tgtggcggag aaacgccgag	1380
ggcgagcccg tgtgtaacgc ctgcggcctt tacatgaagc tgcacggcgt gcccagacct	1440
ctggccatga agaaggaggg aatccagacc agaaagagaa agcccaagaa catcaacaag	1500
agcaagacct gcagcggcaa cagcaacaac agcatcccca tgacccccac cagcacatct	1560
agcaacagcg acgactgtag caagaacaca tcacctacca cccagcccac agctagcgga	1620
geeggegeee eegtgatgae aggegeegga gagteeacaa ateeegagaa tagegaactg	1680
aagtactctg gacaggacgg actgtatatc ggcgtgagcc tggcttctcc cgccgaggtg	1740
accagetetg teagaeetga etettggtgt geeetegeee tggee	1785
<210> SEQ ID NO 53 <211> LENGTH: 3318 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synth polynucleotide	etic
<400> SEQUENCE: 53	
atgttcaact cgatgacccc accaccaatc agtagctatg gcgagccctg ctgtctccgg	60
cccctcccca gtcagggggc ccccagtgtg gggacagaag gactgtctgg cccgcccttc	120
tgccaccaag ctaacctcat gtccggcccc cacagttatg ggccagccag agagaccaac	180
agetgeaceg agggeeeact etttettet eeeeggagtg eagteaagtt gaeeaagaag	240
cgggcactgt ccatctcacc tctgtcggat gccagcctgg acctgcagac ggttatccgc	300
acctcaccca geteeetegt agettteate aactegegat geacatetee aggaggetee	360
tacggtcatc tctccattgg caccatgagc ccatctctgg gattcccagc ccagatgaat	420
caccaaaaag ggccctcgcc ttcctttggg gtccagcctt gtggtcccca tgactctgcc	480
cggggtggga tgatcccaca tectcagtee eggggaeeet teccaaettg ceagetgaag	540
tetgagetgg acatgetggt tggeaagtge egggaggaae eettggaagg tgatatgtee	600
agececaaet ceaeaggeat acaggateee etgttgggga tgetggatgg gegggaggae	660
ctcgagagag aggagaagcg tgagcctgaa tctgtgtatg aaactgactg ccgttgggat	720
ggctgcagcc aggaatttga ctcccaagag cagctggtgc accacatcaa cagcgagcac	780
atccacgggg agcggaagga gttcgtgtgc cactgggggg gctgctccag ggagctgagg	840
cccttcaaag cccagtacat gctggtggtt cacatgcgca gacacactgg cgagaagcca	900
cacaagtgca cgtttgaagg gtgccggaag tcatactcac gcctcgaaaa cctgaagacg	960
cacctgcggt cacacacggg tgagaagcca tacatgtgtg agcacgaggg ctgcagtaaa	1020
gccttcagca atgccagtga ccgagccaag caccagaatc ggacccattc caatgagaag	1080
ccgtatgtat gtaagctccc tggctgcacc aaacgctata cagatcctag ctcgctgcga	1140
aaacatgtca agacagtgca tggtcctgac gcccatgtga ccaaacggca ccgtggggat	1200
ggccccctgc ctcgggcacc atccatttct acagtggagc ccaagaggga gcgggaagga	1260
ggtcccatca gggaggaaag cagactgact gtgccagagg gtgccatgaa gccacagcca	1320
ageeetgggg eecagteate etgeageagt gaeeaeteee eggeagggag tgeageeaat	1380

#### -continued

acagacagtg	gtgtggaaat	gactggcaat	gcaggggggca	gcactgaaga	cctctccagc	1440
ttggacgagg	gaccttgcat	tgctggcact	ggtctgtcca	ctcttcgccg	ccttgagaac	1500
ctcaggctgg	accagctaca	tcaactccgg	ccaataggga	cccgggggtct	caaactgccc	1560
agcttgtccc	acaccggtac	cactgtgtcc	cgccgcgtgg	gccccccagt	ctctcttgaa	1620
cgccgcagca	gcagctccag	cagcatcagc	tctgcctata	ctgtcagccg	ccgctcctcc	1680
ctggcetete	ctttcccccc	tggeteecea	ccagagaatg	gagcatcctc	cctgcctggc	1740
cttatgcctg	cccagcacta	cctgcttcgg	gcaagatatg	cttcagccag	aggggggtggt	1800
acttcgccca	ctgcagcatc	cagcctggat	cggataggtg	gtcttcccat	gcctccttgg	1860
agaagccgag	ccgagtatcc	aggatacaac	cccaatgcag	gggtcacccg	gagggccagt	1920
gacccagccc	aggctgctga	ccgtcctgct	ccagctagag	tccagaggtt	caagagcctg	1980
ggctgtgtcc	ataccccacc	cactgtggca	ggggggaggac	agaactttga	tccttacctc	2040
ccaacetetg	tctactcacc	acageceecc	agcatcactg	agaatgctgc	catggatgct	2100
agagggctac	aggaagagcc	agaagttggg	acctccatgg	tgggcagtgg	tctgaacccc	2160
tatatggact	tcccacctac	tgatactctg	ggatatgggg	gacctgaagg	ggcagcagct	2220
gagccttatg	gagcgagggg	tccaggctct	ctgcctcttg	ggcctggtcc	acccaccaac	2280
tatggcccca	acccctgtcc	ccagcaggcc	tcatatcctg	accccaccca	agaaacatgg	2340
ggtgagttcc	cttcccactc	tgggctgtac	ccaggcccca	aggctctagg	tggaacctac	2400
agccagtgtc	ctcgacttga	acattatgga	caagtgcaag	tcaagccaga	acaggggtgc	2460
ccagtggggt	ctgactccac	aggactggca	ccctgcctca	atgcccaccc	cagtgagggg	2520
cccccacatc	cacagcetet	cttttcccat	tacccccagc	cctctcctcc	ccaatatctc	2580
cagtcaggcc	cctataccca	gccaccccct	gattatcttc	cttcagaacc	caggcettge	2640
ctggactttg	attcccccac	ccattccaca	gggcagctca	aggeteaget	tgtgtgtaat	2700
tatgttcaat	ctcaacagga	gctactgtgg	gagggtgggg	gcagggaaga	tgeeceegee	2760
caggaacctt	cctaccagag	tcccaagttt	ctggggggtt	cccaggttag	cccaagccgt	2820
gctaaagctc	cagtgaacac	atatggacct	ggctttggac	ccaacttgcc	caatcacaag	2880
tcaggttcct	atcccacccc	ttcaccatgc	catgaaaatt	ttgtagtggg	ggcaaatagg	2940
gcttcacata	gggcagcagc	accacctcga	cttctgcccc	cattgcccac	ttgctatggg	3000
cctctcaaag	tgggaggcac	aaaccccagc	tgtggtcatc	ctgaggtggg	caggctagga	3060
gggggtcctg	ccttgtaccc	tcctcccgaa	ggacaggtat	gtaaccccct	ggactctctt	3120
gatcttgaca	acactcagct	ggactttgtg	gctattctgg	atgagcccca	ggggctgagt	3180
cctcctcctt	cccatgatca	gcgggggcagc	tctggacata	ccccacctcc	ctctgggccc	3240
cccaacatgg	ctgtgggcaa	catgagtgtc	ttactgagat	ccctacctgg	ggaaacagaa	3300
ttcctcaact	ctagtgcc					3318

<210> SEQ ID NO 54 <211> LENGTH: 540 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE:

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

<400> SEQUI	ENCE: 54					
atgagtctgg	taggtggttt	tccccaccac	ccggtggtgc	accacgaggg	ctacccgttt	60
gccgccgccg	ccgccgccag	ccgctgcagc	catgaggaga	acccctactt	ccatggctgg	120
ctcatcggcc	accccgagat	gtcgcccccc	gactacagca	tggccctgtc	ctacagcccc	180
gagtatgcca	gcggcaccgc	caaccgcaag	gagcggcgca	ggactcagag	catcaacagc	240
gccttcgccg	aactgcgcga	gtgcatcccc	aacgtacccg	ccgacaccaa	actctccaaa	300
atcaagaccc	tgcgcctggc	caccagctac	atcgcctacc	tcatggacct	gctggccaag	360
gacgaccaga	atggcgaggc	ggaggccttc	aaggcagaga	tcaagaagac	cgacgtgaaa	420
gaggagaaga	ggaagaagga	gctgaacgaa	atcttgaaaa	gcacagtgag	cagcaacgac	480
aagaaaacca	aaggccggac	gggctggccg	cagcacgtct	gggccctgga	gctcaagcag	540
<210> SEQ : <211> LENG <212> TYPE <213> ORGAI <220> FEATU <223> OTHEN polyn	ID NO 55 TH: 1896 : DNA NISM: Artif: JRE: R INFORMATI( hucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	lence: Synth	etic
<400> SEQUI	ENCE: 55					
atggtttcta	aactgagcca	gctgcagacg	gagctcctgg	cggccctgct	ggagtcaggg	60
ctgagcaaag	aggcactgct	ccaggcactg	ggtgagccgg	ggccctacct	cctggctgga	120
gaaggeeeee	tggacaaggg	ggagtcctgc	ggcggcggtc	gaggggagct	ggctgagctg	180
cccaatgggc	tgggggagac	tcggggctcc	gaggacgaga	cggacgacga	tggggaagac	240
ttcacgccac	ccatcctcaa	agagctggag	aacctcagcc	ctgaggaggc	ggcccaccag	300
aaagccgtgg	tggagaccct	tctgcaggag	gacccgtggc	gtgtggcgaa	gatggtcaag	360
tcctacctgc	agcagcacaa	catcccacag	cgggaggtgg	tcgataccac	tggcctcaac	420
cagtcccacc	tgtcccaaca	cctcaacaag	ggcactccca	tgaagacgca	gaagcgggcc	480
gccctgtaca	cctggtacgt	ccgcaagcag	cgagaggtgg	cgcagcagtt	cacccatgca	540
gggcagggag	ggctgattga	agagcccaca	ggtgatgagc	taccaaccaa	gaagggggggg	600
aggaaccgtt	tcaagtgggg	cccagcatcc	cagcagatcc	tgttccaggc	ctatgagagg	660
cagaagaacc	ctagcaagga	ggagcgagag	acgctagtgg	aggagtgcaa	tagggcggaa	720
tgcatccaga	gaggggtgtc	cccatcacag	gcacaggggc	tgggctccaa	cctcgtcacg	780
gaggtgcgtg	tctacaactg	gtttgccaac	cggcgcaaag	aagaagcctt	ccggcacaag	840
ctggccatgg	acacgtacag	cgggcccccc	ccagggccag	gcccgggacc	tgegetgeee	900
gctcacagct	cccctggcct	gcctccacct	gccctctccc	ccagtaaggt	ccacggtgtg	960
cgctatggac	agcctgcgac	cagtgagact	gcagaagtac	cctcaagcag	cggcggtccc	1020
ttagtgacag	tgtctacacc	cctccaccaa	gtgtccccca	cgggcctgga	gcccagccac	1080
agcctgctga	gtacagaagc	caagetggte	tcagcagctg	ggggccccct	cccccctgtc	1140
agcaccctga	cagcactgca	cagcttggag	cagacatccc	caggcctcaa	ccagcagccc	1200
cagaacctca	tcatggcctc	acttcctggg	gtcatgacca	tcgggcctgg	tgagcctgcc	1260
tccctgggtc	ctacgttcac	caacacaggt	gcctccaccc	tggtcatcgg	cctggcctcc	1320
acgcaggcac	agagtgtgcc	ggtcatcaac	agcatgggca	gcagcctgac	caccctgcag	1380

cccgtccagt	tctcccagcc	gctgcacccc	tcctaccagc	agccgctcat	gccacctgtg	1440
cagagccatg	tgacccagag	ccccttcatg	gccaccatgg	ctcagctgca	gagcccccac	1500
gccctctaca	gccacaagcc	cgaggtggcc	cagtacaccc	acacaggcct	gctcccgcag	1560
actatgctca	tcaccgacac	caccaacctg	agcgccctgg	ccagcctcac	gcccaccaag	1620
caggtettea	cctcagacac	tgaggcctcc	agtgagtccg	ggetteacae	gccggcatct	1680
caggccacca	ccctccacgt	ccccagccag	gaccctgccg	gcatccagca	cctgcagccg	1740
gcccaccggc	tcagcgccag	ccccacagtg	tcctccagca	gcctggtgct	gtaccagagc	1800
tcagactcca	gcaatggcca	gagccacctg	ctgccatcca	accacagcgt	catcgagacc	1860
ttcatctcca	cccagatggc	ctcttcctcc	cagttg			1896
<pre>&lt;210&gt; SEQ 1 &lt;211&gt; LENG &lt;212&gt; TYPE &lt;213&gt; ORGA1 &lt;220&gt; FEATU &lt;223&gt; OTHEH polyr &lt;400&gt; SEQUE</pre>	ID NO 56 TH: 1671 DNA NISM: Artif: RE: INFORMATIC Nucleotide ENCE: 56	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	lence: Synthe	tic
atggttagca	aactgacatc	cctccagcag	gaacttcttt	ctgccctcct	ctccagtggg	60
gtaaccaaag	aggtactggt	ccaggctttg	gaggagttgc	tcccctcacc	gaattttggt	120
gtaaagttgg	agactctccc	cctctcccct	ggttctggag	cagagccgga	tactaaaccg	180
gtatttcata	cgcttacaaa	cggacacgca	aagggtcggc	tttcaggtga	cgaagggtct	240
gaggacggcg	atgattatga	caccccgccc	atcctcaaag	aactgcaggc	ccttaataca	300
gaggaagcgg	cggagcagcg	agctgaagtt	gacagaatgc	tctcagaaga	tccgtggaga	360
gctgcgaaaa	tgattaaggg	atatatgcag	caacataaca	ttccccagag	agaggtagtt	420
gatgttaccg	gccttaacca	gagccacctg	tctcagcatc	tcaataaggg	tactcctatg	480
aaaacacaga	agcgagcggc	cctttacaca	tggtacgtgc	ggaagcaacg	agaaattctc	540
cgacagttca	atcagacagt	acaatcttca	gggaacatga	cggataaaag	ctcacaggat	600
cagctcttgt	ttctcttccc	cgagttcagc	caacagtccc	acggtccagg	tcaatctgat	660
gatgcttgca	gtgaacctac	aaacaaaaaa	atgaggagga	acaggtttaa	atggggaccg	720
gcctctcagc	agatactgta	ccaagcgtac	gatcggcaga	aaaacccaag	caaagaggag	780
cgcgaggcat	tggtcgagga	gtgtaatcgg	gccgagtgct	tgcaacgggg	tgtaagtcct	840
agcaaagccc	atggtctcgg	ctcaaacttg	gtcacggagg	tgagggtata	taattggttt	900
gccaacaggc	ggaaggagga	agcattccgg	caaaagctgg	cgatggatgc	ctactcaagc	960
aaccagacac	atagcctcaa	ccctctgttg	tcacacgggt	cccctcatca	ccaaccttct	1020
tcctctccac	ccaacaaact	ttctggtgtc	cgatattccc	agcaggggaa	caacgagata	1080
acatcttcct	ctactataag	tcatcacgga	aattctgcaa	tggtaacgtc	acagagtgtg	1140
ttgcaacagg	tatcacccgc	gtctcttgat	ccaggccaca	atctgttgag	ccctgacgga	1200
aagatgatct	ctgtttctgg	tggcggactc	ccgccggtct	ccacacttac	caacatacat	1260
agteteagte	atcataatcc	tcagcagagc	caaaacctga	ttatgactcc	tcttagcgga	1320
gtgatggcta	ttgcgcaatc	tttgaacacc	tcacaagcac	aatctgtacc	cgtcataaac	1380

aont		$n_{11}$	<u>~ d</u>
COIL	L	шu	eu

		-continued		
agcgtagcgg gctcattggc	ggcgctccaa ccagtgcag	t teteccagea getecattea	1440	
ccccatcaac agcctctgat	gcagcagagc cctggtagt	c acatggetea acageegtte	1500	
atggcagctg tcactcagct	ccagaactcc catatgtat	g cccacaagca agaaccacca	1560	
caatacagtc acacatcaag	attccccagt gctatggtt	g ttactgacac atcctctatc	1620	
tcaactctga cgaacatgtc	cagtagtaaa caatgtcct	c tgcaagcatg g	1671	
<210> SEQ ID NO 57 <211> LENGTH: 1251 <212> TYPE: DNA <213> ORGANISM: Artif <220> FEATURE: <223> OTHER INFORMATI polynucleotide	icial Sequence ON: Description of Ar	tificial Sequence: Synth	hetic	
<400> SEQUENCE: 57				
atgcgactct ccaaaaccct	cgtcgacatg gacatggcc	g actacagtgc tgcactggac	60	
ccagcctaca ccaccctgga	atttgagaat gtgcaggtg	t tgacgatggg caatgacacg	120	
tccccatcag aaggcaccaa	cctcaacgcg cccaacagc	c tgggtgtcag cgccctgtgt	180	
gccatctgcg gggaccgggc	cacgggcaaa cactacggt	g cctcgagctg tgacggctgc	240	
aagggettet teeggaggag	cgtgcggaag aaccacatg	t actcctgcag atttagccgg	300	
cagtgcgtgg tggacaaaga	caagaggaac cagtgccgc	t actgcaggct caagaaatgc	360	
ttccgggctg gcatgaagaa	ggaagccgtc cagaatgag	c gggaccggat cagcactcga	420	
aggtcaagct atgaggacag	cagcetgece tecateaat	g cgctcctgca ggcggaggtc	480	
ctgtcccgac agatcacctc	ccccgtctcc gggatcaac	g gcgacattcg ggcgaagaag	540	
attgccagca tcgcagatgt	gtgtgagtcc atgaaggag	c agctgctggt tctcgttgag	600	
tgggccaagt acatcccagc	tttctgcgag ctccccctg	g acgaccaggt ggccctgctc	660	
agageeeatg etggegagea	. cctgctgctc ggagccacc	a agagatccat ggtgttcaag	720	
gacgtgctgc tcctaggcaa	. tgactacatt gtccctcgg	c actgcccgga gctggcggag	780	
atgageeggg tgteeataeg	; catcettgae gagetggtg	c tgecetteea ggagetgeag	840	
atcgatgaca atgagtatgc	ctacctcaaa gccatcatc	t totttgacoo agatgocaag	900	
gggctgagcg atccagggaa	. gatcaagcgg ctgcgttcc	c aggtgcaggt gagcttggag	960	
gactacatca acgaccgcca	gtatgactcg cgtggccgc	t ttggagaget getgetgetg	1020	
ctgcccacct tgcagagcat	cacctggcag atgatcgag	c agatccagtt catcaagctc	1080	
ttcggcatgg ccaagattga	. caacctgttg caggagatg	c tgctgggagg tccgtgccaa	1140	
gcccaggagg ggcggggttg	gagtggggac teeccagga	g acaggeetea cacagtgage	1200	
tcacccctca gctccttggc	ttccccactg tgccgcttt	g ggcaagttgc t	1251	
<210> SEQ ID NO 58 <211> LENGTH: 1005 <212> TYPE: DNA <213> ORGANISM: Artif	icial Sequence			

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

<400> SEQUENCE: 58

COD	÷	п.	n	11		$\sim$
COIL	-	_	<b>T T</b>	u	$\sim$	u

ggcacttgca gtgcccgagc ctatccatca gaccacagaa ttacaacatt ccaaagctgt	120
gcggtgtcag ccaacagttg cggcggagac gaccgcttcc tggtcggaag aggggttcaa	180
attggatcac ctcaccatca ccatcaccac caccatcacc acccccaacc ggcgacttac	240
caaaccageg geaatttggg egtgagetat agecatteet eatgtggaee tteetatggg	300
tctcagaatt tctccgcccc ttatagccca tacgccctga accaagaggc cgatgtatca	360
ggaggetate eccagtgege gecageggtt tacteaggta atetttetag eccgatggte	420
cagcaccacc atcaccatca aggttatgcc ggcggtgcag tcggatcccc acaatacata	480
caccatagtt acggccaaga gcaccaatcc ctggccctcg ctacatataa caactcactg	540
teteogette atgetteeca ecaagaaget tgteggagte eegeeteaga aaetteetet	600
ccageteaga ettttgattg gatgaaggte aageggaate egeetaaaae gggeaaagta	660
ggtgaatatg gctatttggg acagcctaat gctgtccgca ccaatttcac aacaaaacag	720
cttactgaac tcgagaagga atttcatttt aataagtatt tgactcgagc gagacgagtc	780
gaaatcgccg ctagtcttca acttaacgag acccaggtta agatatggtt ccagaacaga	840
agaatgaaac aaaaaaagcg ggagaaggaa ggactcotoo otatatcaco agocacacoo	900
ccaggtaacg acgagaaggc ggaggaatet teagagaaga gtteeagete eeettgtgtt	960
cetteteetg gtageteaac cagegatace etcaegaega gteae	1005
<211> LENGTH: 282 <212> TYPE: DNA	
<pre>&lt;21&gt;&gt; OKGANISM: ArtIIICIAI Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide</pre>	tic
<pre>&lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59</pre>	tic
<pre>&lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgtgtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt</pre>	etic 60
<pre>&lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgggtcaag gcaattccaa aggggaaaac gcagccaact ggctcacggc aaagagtggt cgggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg</pre>	etic 60 120
<pre>&lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgtgtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt cggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt accttactcg agagcggcgc ctagagatta gccgcagcgt ccacctcacg</pre>	etic 60 120 180
<pre>&lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgggtcaag gcaattccaa aggggaaaac gcagccaact ggctcacggc aaagagtggt cggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt accttactcg agagcggcgc ctagagatta gccgcagcgt ccacctcacg gacagacaag tgaaaatctg gttccagaac cgcaggatga aactgaagaa aatgaatcga</pre>	etic 60 120 180 240
<pre>&lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgtgtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt cggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt accttactcg agagcggcgc ctagagatta gccgcagcgt ccacctcacg gacagacaag tgaaaatctg gtttcagaac cgcaggatga aactgaagaa aatgaatcga gaaaaccgga tccgggagct cacagccaac tttaattttt cc</pre>	etic 60 120 180 240 282
<pre>&lt;210&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgtgtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt cggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt accttactcg agagcggcgc ctagagatta gccgcagcgt ccacctcacg gacagacaag tgaaaatctg gttcagaac cgcaggatga aactgaagaa aatgaatcga gaaaaccgga tccgggagct cacagccaac tttaatttt cc </pre>	etic 60 120 180 240 282
<pre><cli><li><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><p< td=""><td>etic 60 120 180 240 282</td></p<></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></li></cli></pre>	etic 60 120 180 240 282
<pre><cli>&gt; OrGANISW: AFTITICIAL Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgtgtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt cggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt accttactcg agagcggcgc ctagagatta gccgcagcgt ccacctcacg gacagacaag tgaaaatctg gtttcagaac cgcaggatga aactgaagaa aatgaatcga gaaaaccgga tccgggagct cacagccaac tttaatttt cc &lt;210&gt; SEQ ID NO 60 &lt;211&gt; LENGTH: 942 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 60 atggattttg atgagcgtgg tccctgctcc tctaacatgt attgccaag ttgtacttac</cli></pre>	etic 60 120 180 240 282 etic
<pre><cli><li><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><pre><cli><p< td=""><td>etic 60 120 180 240 282 etic 60 120</td></p<></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></cli></pre></li></cli></pre>	etic 60 120 180 240 282 etic 60 120
<pre><cli>&gt; OrGANISW: AFTIIICIAL Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atggtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt cggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt accttactcg agagcggcgc ctagagatta gccgcagcgt ccacctcacg gacagacaag tgaaaatctg gtttcagaac cgcaggatga aactgaagaa aatgaatcga gaaaaccgga tccgggagct cacagccaac tttaatttt cc &lt;210&gt; SEQ ID NO 60 &lt;211&gt; LENGTH: 942 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 60 atggatttg atgagcgtgg tccctgctcc tctaacatgt attgccaag ttgtacttac tacgtctcgg gtccagatt ctccagcctc ccttctttc tgccccagac cccgtctcg cgcccaatga catactccta ctcctccaac ctgccccagg tccaacccgt gcgcgaagtg</cli></pre>	etic 60 120 180 240 282 etic 60 120 180
<pre>&lt;210&gt; OKUANISM: ATUITICIAL Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgtgtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt cggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt accttactcg agagcggcgc ctagagatta gccgcagcgt ccacctcacg gacagacaag tgaaaatctg gtttcagaac cgcaggatga aactgaagaa aatgaatcga gaaaaccgga tccgggagct cacagccaac tttaatttt cc &lt;210&gt; SEQ ID NO 60 &lt;211&gt; LENGTH: 942 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 60 atggatttg atgagcgtgg tccctgctcc tctaacatgt attgccaag ttgtacttac tacgtccgg gtccagatt ctccagccc ccttctttc tgccccagac cccgtctcg cgcccaatga catactccta ctcctccaac ctgccccag tccaacccgt gcgcgaagtg accttcagag agtacgccat tgagcccgc actaaatggc acccccgcgg caatctggcc</pre>	etic 60 120 180 240 282 etic 60 120 180 240 240
<pre>&lt;210&gt; GRUANISM: Aftilicial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgtgtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt cggaagaagc gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt accttactcg agagcggege ctagagatta gccgcagegt ccacetcacg gacagacaag tgaaaatctg gttccagaac cgcaggatga aactgaagaa aatgaatcga gaaaaccgga tccgggaget cacagccaac tttaatttt cc &lt;210&gt; SEQ ID NO 60 &lt;211&gt; LENGTH: 942 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 60 atggatttg atgagegtgg tccctgetce tctaacatgt attgccaag ttgacttac tacgtctegg gtccagatt ctccagect ccttettte tgccccagac cccgtetteg cgcccaatga catactccta ctectccaac ctgccccag tccaaccegt gcgcgaagtg accttcagag agtacgccat tgagccegc actaaatgge acccccgegg caatetggce cactgctact ccgcggagga gctcgtgcac agagactgce tgcaggegc caqcqcqcc</pre>	etic 60 120 180 240 282 etic 60 120 180 240 240 300
<pre>&lt;210&gt; ONGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;220&gt; COTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide &lt;400&gt; SEQUENCE: 59 atgtgtcaag gcaattccaa aggtgaaaac gcagccaact ggctcacggc aaagagtggt cggaagaage gctgccccta cacgaagcac cagacactgg agctggagaa ggagttctg ttcaatatgt acettactcg agageggege ctagagatta gcegcagegt ccacetcacg gacagacaag tgaaaatctg gttcagaac cgcaggatga aactgaagaa aatgaatcga gaaaacegga teegggaget cacagecaac tttaatttt ee </pre>	etic 60 120 180 240 282 etic 60 120 180 240 300 360

				-contir	nued	
gcagtctcgt	ccaatttcta	tagcaccgtg	ggcaggaacg	gcgtcctgcc	acaggettte	420
gaccagtttt	tcgagacagc	ctacggcacc	ccggaaaacc	tegeeteete	cgactacccc	480
ggggacaaga	gcgccgagaa	ggggcccccg	gcggccacgg	cgacctccgc	ggcggcggcg	540
gcggctgcaa	cgggcgcgcc	ggcaacttca	agttcggaca	gcggcggcgg	cggcggctgc	600
cgggagatgg	cggcggcagc	agaggagaaa	gagcggcggc	ggcgccccga	gagcagcagc	660
agccccgagt	cgtcttccgg	ccacactgag	gacaaggccg	gcggctccag	tggccaacgc	720
acccgcaaaa	agegetgeee	ctataccaag	taccagatcc	gagagctgga	acgggagttc	780
ttcttcagcg	tctacattaa	caaagagaag	cgcctgcaac	tgtcccgcat	gctcaacctc	840
actgatcgtc	aagtcaaaat	ctggtttcag	aacaggagaa	tgaaggaaaa	aaaaattaac	900
agagaccgtt	tacagtacta	ctcagcaaat	ccactcctct	tg		942
<210> SEQ 1 <211> LENG <212> TYPE <213> ORGAN <220> FEATU <223> OTHEN polyr	ID NO 61 FH: 672 : DNA VISM: Artif: JRE: R INFORMATIC nucleotide	icial Sequer DN: Descript	nce tion of Arti	ificial Sequ	lence: Synthet	ic
<400> SEQUE	ENCE: 61					
atgagttcct	atttcgtgaa	ctccaccttc	cccgtcactc	tggccagcgg	gcaggagtcc	60
tteetgggee	agctaccgct	ctattcgtcg	ggctatgcgg	acccgctgag	acattacccc	120
gcgccctacg	ggccagggcc	gggccaggac	aagggctttg	ccacttcctc	ctattacccg	180
ccggcgggcg	gtggctacgg	ccgagcggcg	ccctgcgact	acgggccggc	gccggccttc	240
taccgcgaga	aagagtcggc	ctgcgcactc	tccggcgccg	acgagcagcc	cccgttccac	300
cccgagccgc	ggaagtcgga	ctgcgcgcag	gacaagagcg	tgttcggcga	gacagaagag	360
cagaagtgct	ccactccggt	ctacccgtgg	atgcagcgga	tgaattcgtg	caacagttcc	420
teetttggge	ccagcggccg	gcgaggccgc	cagacataca	cacgttacca	gacgctggag	480
ctggagaagg	agtttcacta	caatcgctac	ctgacgcggc	ggcggcgcat	cgagatcgcg	540
cacgccctgt	gcctgacgga	gaggcagatc	aagatatggt	tccagaaccg	acgcatgaag	600
tggaaaaagg	agagcaaact	gctcagcgcg	tctcagctca	gtgccgagga	ggaggaagaa	660
aaacaggccg	ag					672
<210> SEQ 1 <211> LENG <212> TYPE <213> ORGAN <220> FEAT <223> OTHEN polyn	ID NO 62 TH: 1410 : DNA NISM: Artif: RE: R INFORMATIC nucleotide	icial Sequer DN: Descript	nce tion of Arti	ificial Sequ	lence: Synthet	ic
<400> SEQUE	ENCE: 62					
atggctgtca	gcgacgcgct	gctcccatct	ttctccacgt	tcgcgtctgg	cccggcggga	60
agggagaaga	cactgcgtca	agcaggtgcc	ccgaataacc	gctggcggga	ggagetetee	120
cacatgaagc	gacttccccc	agtgetteee	ggccgcccct	atgacctggc	ggcggcgacc	180
gtggccacag	acctggagag	cggcggagcc	ggtgcggctt	gcggcggtag	caacctggcg	240
cccctacctc	ggagagagac	cgaggagttc	aacgatctcc	tggacctgga	ctttattctc	300

tocaattogo tgacotato tooggagta gigocogoca ooggagotga geogogota 360 gottottt optogrago oggaaaga ooggagotga geogogotga caeggogga 480 ggottott atggatogga ooggagataga ooggagotga geogogotga caeggogga 480 ggottott atggatogga geogotto oottogaagg toottotaa ootggaga 480 ggottott atggatogga geogotto oottogaagg toottottaa ootggaga 480 ggottotta atggatogga geogotto oottogaagg toottottaa ootggaga 600 coggatgataa thoogooga gaagootga ooggaggag toottottoa ootggagga geogotaga geogotga geogotga ageagetag geogotgat geogotgat geogotga geogoca gegocotag agacagaag ooggaggatog geogotgat geogotgat caetagogg 720 geogogod ootgagada ootgacotga geogotgat tottogegac caetagog 720 geogoca gegocotag gataagaag geogotgat oottogegac caetagog 720 cagotgacoo cottagaat geogotgat geogotgaa aagattogo caetagoa 100 agutoacotg ootgacogt toottooggo toottoe acoogggoo caetacoo 100 agutoacotg coetacagat geogotgaa aagagegaa geogogo caetacoa 100 agutoacot toegagoda agugagaco aagoagaag atggegaga 180 caecoogtt coetacoogt caegagata (seogoogt caegagaa 120 agutoacot toegagoda attogoogt caegagaa geogoga 180 caecoogtt coetacoogt caegagata (seogogg caeagagaa 120 agutocaet taeggeaa attogoogt caegagaa geogoga 180 caecoogtt geogoaaa aatggegac gaacatago geogaa 180 caecoogtt geogoaaa aatggegac gaacatag gegaatto 110 caecoogtt seosagaa geogocst caegegac seogogga ceoofoogta 100 geatogot figatigga attogoogt caegagaac seogogga ceoofootgea 40 seosoogt seosaaag gegacetag geogoaga ceoofootgea 40 geatageig ceogegagg geocegac geogagage ceoofootgea 40 geatageig ceogagage geocegac ceoofootgea 40 geatageig ceogagage geocegac geogagage ceolootge geogagage 240 gettatboa aggagacet uttocaego ceoggaga geogaga geogagage 240 gettatboa aggacgatt titaagego togoogac geogagaa geogagag 240 geatageig geogegagag geogoega ceoogaga geogagag 240 geatageig geogegacega geogoega geogoga geogagage 140 aagaagage geogegacegag geogoga geogoga geogagag geogagag 240 geacoogg chaegagacet secacaaa geogogaga geogagaga geogagage 240 geacoogg chaegagacet geogocaa geogoegaga geogagaga geogagage 240 geacoogg chaegage	-continued	
gototott ogtotogo grogsanga gototogot gogogogo anogogoga         400           ttoacetae ogtotogog ogganga oggogoga gotogogog anogogoga         400           gotottet stogotogog degenete oftogotog ototogogo anogogoga         600           ogspitaa trogcopi oggogotto gygogoga toetogogo agatogog         700           getogotaga ogtocotog oggogotto oggogotta oggogoga oftogogoga         700           getogotaga ogtocota oggogotta oggogotta oggogotta oggogoga         700           getogotaga ogtocota oggottaga oggocotaga oggogoga         900           ogtogotaga ogtocota oggottaga oggottaga oggottaga oggogoga         900           ogtocotat oftogoga toetocotag oggottaga oggottaga oggogoga         900           ogtocotat oggogatae oggottaga oggottaga oggottaga oggottaga         900           ogtocotat oftogoga toetocotaga oggottaga oggottaga oggottaga oggottaga         900           cottocota oggogatae oggottaga oggottaga oggottaga oggottaga         900           cottocota oggottaga oggottaga oggottaga oggottaga oggottaga         900           cottocota oftogogatae otogogatae         900           cottocota oftogogatae         900	tccaattcgc tgacccatcc tccggagtca gtggccgcca ccgtgtcctc gtcagcgtca	360
teacatate epiteregge eggaacate eeggeegig egegegig eeggegig eeggegig e teacatate epiteregge eggaacate eeggeegig etcetegeig erecettea eeggegig e teacatate epitereggi eggeegite giggeegite giggeegite eggeegite eggeegite eggeegite effetegeig eggeefite eggeegite eg	geeteetett egtegtegee gtegageage ggeeetgeea gegegeeete eacetgeage	420
gootootot atggoagga gtoogotoo otoogagg totoottoa otaganga 540 accaangao tagaocoto gangagtto gtagoogas tototgoago agaatagaa 600 caggitaa toogooga gaacoota orgatgatag gengoota genocato genocato genocato genocato a gtagtagaag cagootaa orgataca orgatgatag tagoocoto genocatogag 700 cagoogaa egtaccocaa gataacga gangagta totogoo genocatogag 700 cagoogaa egtaccocaa gataacga gangagta totogoo accaggagg 900 cagoogaa egtaccocaa gataacga gangagta genocato accaggagg 900 cagoogaa egtaccocaa gataacga gangagea agtactoo cotaggagga 900 caccoogat cotagoagt totocaccog accaggag castatacca 1000 tototace genagaa genocaca agagtage agtoogoo accagagga 010 caccoogat cotagoagt egoogaac accagaga gagaagaa deggagoo 1100 agtococa gangaata coogooga caatagoo agagaaaca taccataga 1200 agtococat caaagaa attogooga toagtagaa tagoogaa tagoagaa 1200 agtococat caaagaa attogooga toagtagaa gagaagaa ctaccataga 1200 accacoggt eggagaga attogooga tagatagaa tagoogaa 1200 accacoggt eggaagaa attogooga toagtagaa 1200 accacoggt eggaagaa attogooga toagtagaa tagoogaa 1200 accacoggt eggaagaa attogooga caacaggt gaaacota 1200 accacoggt eggaagaa attogooga caacaggt gaaacota 1200 accacoggt eggaagaa attogooga caacaggt gaaacota 1200 accacogaa aggocata 1200 accacogaa gegaatgaa attogooga caacaggt gaaacota 1200 accacogaa aggocagaa attogooga caacaggt gaaacota 1200 accacogaa gegaagaa attogooga caacaggt gaagaaco 1200 accacogaa aggocagaa gagaagaa 1200 accacogaa gegaagaa gagaagaa 1200 accacogaa gegaagaa gaagaagaa 1200 accacogaa gegaagaa gaagaagaa 1200 accacogaa gegaagaa gagaagaa 1200 accacogaa gegaagaa accaco gagaagaa 240 gtagaabaa gagaagaa gagaagaa 1200 accacogaa gegaagaaga gagaagaa gagaagaa 1200 accacogaa gegaagaa gagaagaa gagaaga	ttcacctatc cgatccgggc cgggaacgac ccgggcgtgg cgccgggcgg cacgggcgga	480
atcacegog tyggcccct gyggggtte tyggccgoe corresponder of the second technologie of the second techno	ggeeteetet atggeaggga gteegeteee eeteegaegg eteeetteaa eetggeggae	540
coggtgtata ttroegrege gesgerege gesgerege gesgerege gesgerege         600           gtgtgtaag getgettag gegeerege gesgerege gesgerege getgetee         720           gtagacaag geagectag eggeageae coggtgtgt gesgerete ttegtgee catagesggg         700           getggeeree gesgegea eggeageae (eggetgee aggettee coggtggge gesgerege)         900           cogeregegea eggeageae tegeereegg gesgetee coggtgge gesgeree cottegege         900           cogeregegea eggeageae tegeereegg gesgeree cottege gesteree cottegege         900           codecoggt cottageag gesgeree aggeageae gegeageae tegetgege         900           codecoggt cottageae tegeereegge gesgesge gesgeagea tegeggee aggeageae         1200           codecoggt cottageae tegeege tegesgeae tegeereegae tecestagea         1200           codecoggt aggeegea eggeegeae tegesgeae tegesgeae tecestagea         1200           codecoggt aggeegea eggeegea eggeggea eggeageae tegesgeae         1200           codecogge aggeegea eggeggea tegegegea etgesgegea         1200           codecogge aggeegea eggeggeae tegegegea etgesgeae tegesgeae         1200           codecogge aggeegea eggeggaae         1200           codecoge tegeegeae aggeggeae         1200           codecoge tegeegea aggeggee eggeageae eggeageae         1200           codecoge tegeegeae eggeageae         1200           codecoge eggeageae         1200           codecoge eggeageae	atcaacgacg tgagcccctc gggcggcttc gtggccgagc tcctgcggcc agaattggac	600
gtotqaaag       gtotqaaag       gtotqaaag       gagaqacad       cogcogeget       ftoqaaag       gagaqacad       cogcogeget       dgtoqaaag       gagaqacad       dgtoqaaag       gagacadad       gagaqacad       dgtoqaaag       gadacadg       gagacadad       dgtoqaaag       gadacadg       dgtoqaadg       dgtoqaadg       dgtoqaadg       dgtoqaadg       dgtoqaadg       dgtoqaadg       dgtoqaadg       ddtoqaag       ddtoqaag       ddtoqaagd       ddtoqaagdag       ddtoqaagd       ddtoqaagdag       ddtoqaagd       ddtoqaagd       ddtoqaagd       ddtoqaagdag       ddtoqaagd       ddtoqaagdag       ddtoqaagd       ddtoqaagdag       ddtoqaagd       ddtoqaagdag       ddtoqaagd       ddtoqaagdag       ddtoqaagd       ddtoqaagdag       ddtoqaagd       ddtoqaagd       ddtoqaagd       ddtoqaagd       ddtoqaagd       ddtoqaagdacad       ddtoqagd       dd	ccggtgtaca ttccgccgca gcagccgcag ccgccaggtg gcgggctgat gggcaagttc	660
gitagaaaa gaagaaa gaagaaa agagaagaa aagaataa aagaataa aagaataa aagaataa aagaataa aagaabaa aagaataa aagaaxaa aag	gtgctgaagg cgtcgctgag cgcccctggc agcgagtacg gcagcccgtc ggtcatcagc	720
ecgecgegea eggecea gateaageag gageggett titteggee eatgettee etgggegg 940 etggaece titteagea tggeeceag geggetea agattee etggggeg 940 tgteaeceag geggaete eccegaeceg ggttegag aggeggee aatteeea 1020 tectteeg eegaeagat geagegea gteeggee teeataeea agagetaag 1080 ecaeceggt etgeaege aggeggee aggeegge teeataeea agagetaag 1080 eggaaagga eegeeceae etggeget aggeggee gegeaaae etaeaaag 1200 agtteeet etaageaa etgeggaee agteeggege teeataeea agagetaag 1200 agtteeet etaageaa etgeggaee aetgeegge gegeaaae etaeeaaag 1200 agtteeet taageaa etgeggaee aattegegge gegeaaae etaeeaaag 1200 ecaecegget gtggatgaa ategeegge eggeaaae tgaeeagga 1200 eaceegget aggeeget eegtgeea aatgegee gaeatte eacetaga 130 eaceegget aggeeget eegtgeea aatgegee gaeatte eagetgege taeeagege aegeeget eegtgeea aatgegee gageatte eaggegae 1380 eacetegeet taeeatgaa gaggaett 1410 e210 - SEO ID NO 63 e211 - DENOTH: 1236 e222 - FUTHE: e220 - FENTHE: e220 -	gtcagcaaag gcagccctga cggcagccac ccggtggtgg tggcgcccta caacggcggg	780
getggacce         etteragear tiggecace         eggetterear         geog           cagetterear         geogartae         coegarceig         geogartae         fee           tigtecaced         geogartae         coegarceig         tettecacear         fee           tigtecaced         geogartae         geogartae         geogartae         fee           coecocegat         coegarcagat         geogarcagat	ccgccgcgca cgtgccccaa gatcaagcag gaggcggtct cttcgtgcac ccacttgggc	840
cageteecea geaggatea ecegaeeet gettetgagg aagtgetgag cageaggae       960         tgteaeeetg ecetaeegt teeteeegg tteeateee acegggge caattaeea       1000         caeeeggtt ectgeaeget teeteeegg teeaagtae       1000         cggaaagga ecgeeeae geeeggget geggaaaae etaeeaaga       1000         agtteeeat teaaggaee eageeaag teegggget geggaaaae etaeeaaga       1000         agtteeeat teaaggaee eageeaggt geggeaaae etaeeaaga       1200         agtteeeat teaaggaee ecgeegget geggeaaae etaeeaaga       1200         agtteeeat teaaggaee acegeegge teagatgae tgaeeagge etaeggegee       1300         caeaeggge acegeegt ecagteea ateegeagge gegeattte eagetgegge       1300         caeaeggge acegeegt ecagteea aaatgegaee gageattte cagteggee       1300         caeaeggge acegeegt ecagteeaa aaatgegaee gageattte cagteggee       1300         caeaeggge acegeegt ecagteeaa aaatgegaee gegaegteea etaeeaga       1300         caeaeggge acegeegt ecagteeaa aaatgegaee gageattee systegeeggeege       1300         caeaegge acegeegt ecagteeae ecestgeege gegaeggee etgeteeee Synthetee       1300         c310> SEQUENCE: 63       130         atggaagge geggeggeg gegeeggee geeggaege ecestgee eggaaggeega       1300         gegaeteege ecageagg gegeeege ecestgee ecestgeege eggaaggeega ecestgeege eggaaggaega       1400         gegaeteege ecageagg geggeege ecestgeege eggaaggeega ecestgeege eggaaggaega       160         g	getggaceee eteteageaa tggeeaeegg eeggetgeae aegaetteee eetggggegg	900
tgtaacectg coctgooget tootoooge ttooatooo accogggee caatacea 1020 toottootge cogatogga gegeogea gecoogea gecoogege tooatacea agageteag 100 eggaaaagga cogocaace caetgtgat taegeogget gegeaaaa etaaceaag 120 agttoocate tooaggeae ootgegaace caeacaggtg agaacetta coactgtga 1260 tgggaegget gtggatgga attogeoge toogatgaac tgacoagge atcacegtaa 1320 caacoggge acegeoegt coagtoocaa aaatgegae gageattte caggtogga 1380 caecteget tacacatga gaggeattt 1410 -210> SEQ ID N0 63 -211> LENNTH: 1236 -220> SEQ ID N0 63 -211> LENNTH: 1236 -220> SEQUENCE: 63 atggaagge gegggaget gggecegee cogggagege cogggageg cogggageg acegeoegt cagatgaac tgacage -200> SEQUENCE: 63 atggaagge gegggaget gggecegee cogggagege gggaggega cotgetega 120 caacacet ggeceget caectecaag gegeege gegagagege cogggagege acegeoegt cagatgaac tgoogage cogggagege 120 cacacet ggeoege cogggaget ggeoegee cogggagege gggaggega cotgetega 120 cagacatee tggaceget gggeegee gggagage ggaagate geoagge cogggagege acegeoegt 120 caagacatee tggaceget caececaag geceagge geoaggee ggegaggeg a cotgetega 120 cacacet geoegeg gggaget ggeoegee geoegege geoegege geoegege geoegegeg 120 gecatege geoggageg ggeoegee geoegega agtegee geoegegeg 120 gecatege geoggageg ggeoegee geoegege geoegege geoegege geoegege 120 gecatege geoggaget ggeoegee geoegege geoegege geoegege 120 gecatege geogegege ggeogee geoegege geoegege geoegege 120 gecatege geogegege ggeogee geoegege geoegege geoegege 120 gecatege geogegege ggeogee geoegege geoegege geoegege 120 gecatege geogegea ggeogeae geoegege agtegee geoegege 120 geogetecaege ggeogeae geoegege geoegege agaagee 340 geoegeege tegtgtega ggeogeae geoegege agaagee gagageeg 340 geoegeege tegtgtega ggeogeae geoegege agaege gagageegegee 340 geoegeege tegtgtega ggeogeae geoegege agaegege gagageegegeege	cageteecca geaggaetae eeegaeeetg ggtettgagg aagtgetgag eageagggae	960
teetteetge eegateagat geageegea gteeegeege teeattace agageteag 1080 eeaceeggtt eetgetage agaggagee aggeegaag ggggagaga ategtggee 1140 eggaaaagga eegeecee acettgtgat taegeegget geggeaaae etaceaaag 1200 agtteeeate teaaggeage eegeege tagatgaae tgaceagge agaaeette eeagtgaa 1320 eacaeegggg acegeeegt eeagtgeae aaatgegaee gageattte eaggeegga 1380 eacetegeet taeeatgaa gaggeattt 1140 e210> SEQ ID N0 63 e211> SEQ ID N0	tgtcaccctg ccctgccgct tcctcccggc ttccatcccc acccggggcc caattaccca	1020
ccacceggt cctgcatgc agaggagce agecaaga ggggagae ategtggee 1140 cggaaaagg ccgccacca cactgtgat tacgegggt geggaaaac ctaccaaaa 1200 agtteccate teaaggeaac degegaee cacaeggg agaacetta ceaetgge 1260 tgggaeggg geggege geggegee teaggtgea aaatgegae ggeeggee ctacegtaa 1320 cacaegggg acegeeegtt coagtgeea aaatgegaee ggeeggeg acegeegge 1380 cacetegeet tacaeatgaa gageattt 1410 callo SEQ ID NO 63 <111 - LENGTH: 1236 <111 - LENGTH: 1236 <112 - VTFF: NNA <113 - ROADING: Artificial Sequence <122 - VTHER: NNO Aggaegge geggggggg ggaeetgege eggaggteg eggaggee eggegege tegetgetge 1400 - SEQUENCE: 63 atggaggeg geggggggg ggaeetgege eggaggee gegaggee eggegggg gaeggaggee ctaagteg gegeggagg ggaeetgege eggagget getegegeg aggggagge 240 gttactgea aggaegaett tteaagege teegtggee aegagtete eetgegeg gaeg geaetgeeg e ceaegeagg ggeeegeg geeeggaget geeeggege eggeggage 240 gttactgea aggaegaet ggeegeeg geeeggaget geeeggage eggeege eggeagge 240 gttactgea aggeeget ggeegege geeeggaegt geeeggag geeeggee etgeeg gaeg geeageege teegtgtegea geegegeeg geeeggaege geeeggeeg eggeaggee 240 gttactgea aggeegeet geeegeeg geeeggeeg aegagtete eetgegeeg gee geeggeege eeegeegg ggeegeege geeeggeeg geeeggeeg geeeggeeg efgeegge gae geeggeege eeegeegg ggeegeege geeeggeeg geeeggeeg geeeggeeg efgeeggeeg	teetteetge eegateagat geageegeaa gteeegeege teeattaeea agageteatg	1080
cggaaaagga cogccacca cattgtgat tacgcggot goggaaaa ctacacaaag 1200 agttcccat tacaggaca octgogaac cacacaggtg agaaactta cacatgag 1260 tgggacggt gtggatggaa attogocogo tagatgaa tgacaggac atacogtaa 1320 cacacgggge accgccett cagtgcaa aaatgogace gagattte caggtogga 1380 cactogoet tacacatgaa gaggattt 11410 cacacgggge accgccett cacatgaa gaggattt 111 cacacagggge gaggagat gaggacgae cacacgaga gagatgee synthetic c210> SEQ ID NO 63 c211> LEBIOTH: 1236 c212> TTPE DIA c213> OFANINSH: Artificial Sequence c223> OFEATURE: c223> OFEATURE: c223> OFEATURE: c223> OFEATURE: c223> Gaggagge gaggegggg ggacegge cggagggeg cggagggg cactgege cggaggage 120 cacacgeg goggggag tgggccogge cggagagte grggaggeg acctgege tggtgtgaa 120 cacacgeg gogggggg ggacetgee cgagagate cgetgtgee tggetgtgaa 120 geatacee tggacget catectaag geteggace gacatgge cagaggage 340 gttacega aggacgatt ttraagege tegggagge gacacgege gagaaget cggtagagg 340 gttacega goggegga gggeggae ggacage gacacgega aggacgae categaagg 340 gttacega goggega gggeggae gacacace gacacggae tggtgagae gacaggee 480 acggecaage goggeggae ggeeggae gacacace gacaaggae tggagaegge 540 gacaagee gogeggega gacacace gacaaggae tggagaege gagatgee 540 tacaagee gogeggea ggeeggea ggeeggae tggagaege gagaaggee 340 ctggacaage geeggegae ggeeggea ggeeggae tggagaagee gagaaggee 340 acggecaage goegegae ggeeggea ggeeggae tggagaege gagaaggee 340 ctggacaage goegegae ggeeggea gacacae gacaaggae tggagaegee 340 acggeegae tegteggaa gacacae gaaacgee agaegeeg gagaaggee 340 ctggacaage geegegae ggeeggea gacaagee tggagaagee gagaaggee 340 ctggacaage geegegae ggeeggea gttggegae agaegeeg gagaaggeeg 540 cagaagaeg ceegeegge geegegae geeggegae agaegee gagaaggee 340 ctggacaage geegegae geeggeegae gttggegae agaegee gagaaggee 370 gaeggeetea ageeggeeg geeggeeg agaegeeg agaegeegee 370 gaeggeetea ageeggeeg geeggeege agaegeeg agaegeegeege 720 geeggeetea ageeggaea gaegeegee agaagaegee ageegeegee 720	ccacccggtt cctgcatgcc agaggagccc aagccaaaga ggggaagacg atcgtggccc	1140
agtteecate teaaggaca eetgegaaee cacacaggt agaaacetta eeaetgaa 1260 tgggaegget giggatgga attegeeege teagatgaae tgaecagge atteegataa 1320 cacacagggge acegeeegt eeggeegae teagatgaae tgaecagge attee caggtegga 1380 cacetegeet tacacatgaa gaggattt 1140 <110 SEQ ID NO 63 <111 SEQUENTE: 1236 <112 SEQUENTE: 1236 <123 SORAISMIN Attificial Sequence <123 SORAISMIN Attificial Sequence <123 SORAISMIN Attificial Sequence <123 SORAISMIN Attificial Sequence <123 SORAISMIN Attificial Sequence <120 SEQUENCE: 63 atggaaggege gegggagget gggeeegge eggagagte gegeaggeg tettagee tggetggagge 120 eagacatee tggaeegte tatecaga getetggee gegaaggee geeaggag tettagee aggaggage 120 gettactgea aggacgatt ttecagee tegggaee geegaggee geeaggege teggtggagge 120 gtttactgea aggacgatt ttecagee tegggaee aggeggee geeaggag teggaegge 120 geategee tegtgeea geeggeage geeaggee geeaggee tegtgeea geeaggee tegtgeea geeaggee tegtgeea geeaggee tegtgeea geeaggee tegtgeea geeaggee tegtgeea geeagege tegtgeea geeagege tegtgeea geeagege tegtgeea geeagege tegtgeea geeacagee tegtgeea geeagegee tegtgeea geeagege tegtgeea geeageeae tegtgeea geeagegee tegtgeeae geeagegee tegtgeeae geeagegee tegtgeeae geeagegee tegtgeeae geeagegee tegtgeeae geeageeae geeacegeae tegaaggeee tegtgeeae geeageeae geeacegeae tegaageeee tegaaggeee fegeeaggeeae geeagegee tegaagaeee geeeaegae tegeagaeee tegaagaeee fegeeaggee fegeeaggeee fegeeaggee fegeeaggeee fegeeaggeee fegeeaggeee fegeeaggeeee fegeeaggeeee fegeeaggeeee fegeeaggeeeee fegeeaggeeeee fegeeaggeeeeeeeeee	cggaaaagga ccgccaccca cacttgtgat tacgcgggct gcggcaaaac ctacacaaag	1200
tgggacggt gtggatgga attcgcccgt tcagatgac tgaccagga ctaccgtaa 1320 cacacgggg accgccgtt tcagtgcaa aaatgcgac gagatttt caggtggag 1410 cactcgcct tacacatga gaggattt 1410 c210 > SEQ ID NO 63 c211 > LENCTH: 1236 c212 > TTFE: DNA c213 > ORGNISM: Artificial Sequence c220 > FEATURE: c223 > OTHER INFORMITION: Description of Artificial Sequence: Synthetic polynucleotide cagacated tggacgg gagectgge eggaggteg eggaggeg acctgegteg 120 cagacated tggacget eggegagge gagectgge eggaggate eggegggagga 240 gtttactgea aggacgatt tttcaagge ttegggace gegaggage 240 gtttactgea aggacgatt tttcaagge ttegggace gegaggate eggeggagg acctgege gag gacagecge teggtgea gegegagge gaccecgge gacagge acgtege gag gttactgea eggeagge gacgacge gecaggge acgtege aggaggteg effe ttigeetgg tegttgea geggagge gaccecge gacagge tegggagga 240 gtttactgea aggacgatt tttcaagge ttegggace aggacgate effe teggacateg eggeggag gacgacge gacagge aggagge aggaggag 240 gtttactgea aggacgatt tttcaagge tegggagae aggacge gacgagge 340 gttgactege tegttgea geggagge gacacge gacagge aggagge aggaggagg 340 tttgeetgg tegttgea geggagge gacacge gacagge aggagge 340 gacagecgg tegtgtgea geggage gacacge gacagge aggagge 340 deggecaag gegege gegegage gacagge aggagge 340 deggecaag gegege gegegage gacacge gacagge aggagge 340 deggecaag gegege tegtgtgea gegegage gacacge aggaggag 340 deggecaag gegege tegtgtgea gegegage gacacge teggagag 340 deggecaag gegege tegtgtgea gegegege gecaggag aggagge 340 deggecaag gegegege gegegege gegegege agacgege aggaggeg 340 deggecaag gegegege gegegege gegegege aggacgeg agaaggege 340 deggecaag gegegege gegegege gegegege agacgege agaaggege 340 deggeege teggegea gegegege gegegege agaaggege 340 deggeege teggtgea gegegege gegegege agaaggege 340 deggeege teggegea gegegege gegegege 340 deggeege teggegea gegegege gegegege 340 deggeege 140 deggeegege teggegega gegegegege 340 deggeegege 140 deggeegegegegegegegegegegegegegegegegeg	agttcccatc tcaaggcaca cctgcgaacc cacacaggtg agaaacctta ccactgtgac	1260
cacacgggg accgcccgt cacgtgcca aaatgcgac ggactttt caggtggg 1380 cacctoget tacacatga gggeattt 1410 <pre></pre>	tgggacgget gtggatggaa attegeeege teagatgaae tgaceaggea etacegtaaa	1320
cacctcgcct tacactga gaggattt 1410 <li>Callo SEQ ID NO 63 <li>Callo LENGTH: 1236 <li>Callo SEQUENTH: 1236 <li>Callo FEATURE: <li>Callo FEATURE: <li>Callo FEATURE: <li>Callo SEQUENT: 63 adggagggg gaggaggt gggccgggc cgggagtcg cgggagtcg cggtgggg c cggggaggg 1 ggggggg g cggggggg g cggggggg c ggggggg</li></li></li></li></li></li></li>	cacacggggc accgcccgtt ccagtgccaa aaatgcgacc gagcattttc caggtcggac	1380
210> SEQ ID NO 63 (211> LENGTH: 1236 (212> TYPE: DNA (213> ORGANISM: Artificial Sequence (220> FEATURE: (223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide (400> SEQUENCE: 63 atggaggege geggggggge ggacetgege egggagtege eggetggege tggetgtgae 120 cageacatee tggacegett cateeteaag getetggaee gecaetggea eageagtgt 180 eteaagtgea gegaetgeea eageecaetg geegagget egteteageeg 300 ggeateege caeegeagt ggtegegeege gecaeggaet tegtgtaeea ectgeaetge 360 tttgeetgeg tegtgtgeaa geeggeagtg gecaeggege acgagteta ecteatege 360 tttgeetgeg tegtgtgeaa geeggeaeta gaaaceeea ageageaga ggeegagee 480 acggeecaage ggeegegea gaecateea gaeaeegea ageageage 540 ttaeaaeeee egeecaage ggeegeeae gecaaggae tegtgagaege 540 taeaaaeee egeecaage ggeegeeae gecaaggae tegtgagaegee 540 taeaaeeee egeecaagee ggeegeeae ggeegagee gaeagagae 660 etggacatee geetggtgea ggettggte eagaaeegea gaeageage 660 etggacatee geetggtgea ggttggte eagageege ggeecaagga gaagagetg 660 aagaagggaeg eeggegeeae ggeegegge egaeagge agaetteege 720 ggeegeteea agteggeeae ggeegegge aggaeagee 780 teetteeeeq atgaeeette ettageggge aggaeagee 780 teetteeeeq atgaeeette ettageggage aggaeagee 670 teetteeeeq atgaeeette ettageeggae aggaeageet etagegeege 780	cacetegeet taeacatgaa gaggeatttt	1410
atggaggegegeggggagetgggeceggeeegggagteggegggagteggeggaeteggeegggagteggeggaeteggeeggagggeggeggeeteggeegggaggggeggeeteggeeggaeteggeeggeateggeat	<210> SEQ ID NO 63 <211> LENGTH: 1236 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synth polynucleotide <400> SEQUENCE: 63	hetic
gcactgctgg cgcggaggg ggacctgcg cgagagate cgctgtgcg tggctgtga120cagcacate tggaccget catectcaag getetggae gccatggae cagcaagtgt180ctcaagtgea gegactgee caegeeactg geegageget getteageeg agggagage240gtttaetgea aggacgaett ttteaageeg teegggaeea agtgegeeeg gtgeeagetg300ggcatecege caegeaggt ggtgegeege geecaggaet tegtgtaeea eetgeactge360tttgeetgeg tegtgtgeaa geeggaetge geecaggae ageagteta eetegagae420gacageegge tegtgtgeaa ggeggaetae gaaacegeea ageageagaa ggeegaggee480acggeecaage ggeegeeae gaceateace geeaageag tggagaege tgagaggee600etggaeatge gegtggtgea ggtttggte cagaacegee ggeecaagga gaegaggee600etggaeatge gegtgggeea geetgggg eagtattee geaacatgaa gegeteecege720ggeeggeteea agteggaeaa ggaeagett caggaggge aggaeagea egetgaggee780tectteeceeg atgageette ettggegaa atgageeegg eceaatggee eagatgeee getgaggee840	atggaggege geggggaget gggeeeggee egggagtegg egggaggega eetgetgeta	60
cagcacatcetggaccgcttcatectcaaggetetggacegecactggaccagcacagtgt180ctcaagtgcagegactgecacacgecactggecgacgecggettaagegagggagagg240gtttactgcaaggacgactttttcaagegettegggaccaagtgegecgegecactge300ggcatcecegeccacgcaggtggtgegecgegeccacgggggagtgegecge360tttgeetgegtegtgtgeaageggeageggaegagttetaeetaetggagg420gacageeggetegtgtgeaaggeggactaegaaacegeaageageggge480acggeeageggeegegeaegacateegeegeggge660ctggacatgegegtggtgeaggettggttecagaacegge660aagaaggaegceggeeggaegeedeggggeaggaeagegt720ggeeggeeeaagtegggaeaggaeageggaaggaeagege780teetteeegatgageetteettaggaeaggegaeagegg640	gcactgctgg cgcggagggc ggacctgcgc cgagagatcc cgctgtgcgc tggctgtgac	120
ctcaagtgca gcgactgcca cacgccactg gccgagcgct gcttcagccg aggggagagc240gtttactgca aggacgactt tttcaagcge ttcgggacca agtgcgccge gtgccagctg300ggcatcccge ccacgcaggt ggtgcgcge gcccaggact tcgtgtacca cctgcactge360tttgcctgcg tcgtgtgcaa gcggcagctg gccacgggcg acgagttcta cctcatggag420gacagccgge tcgtgtgcaa ggcggactae gaaacegcea agcagcgaga ggccgaggec480acggccaage ggccgcgcae gaccatcace gccaagcage tggagacget gaagagcget540tacaacacet cgcccaagce ggcggcgcae gtgcgcgage agctctcgte cgagacggge600ctggacatge gcgtggtgca ggtttggtte cagaacegce gggccaagga gacgactccge720ggcggctcca agtcggacaa ggacagctt caggaggge aggacagcag cgctgaggte780tccttccccg atgacctte cttggcggaa atgggcccgg ccaatggcet ctacgcgacg840	cagcacatee tggacegett cateetcaag getetggace gecaetggea cagcaagtgt	180
gtttactgca aggacgactt tttcaagege ttegggacca agtgegeege gtgecagetg300ggeateeege ceaeggagt ggtgegeege geeeaggaet tegtgtacea eetgeaetge360tttgeetgeg tegtgtgeaa geggeagetg geeaegggeg aegagtteta eeteatggag420gacageegge tegtgtgeaa ggeggaetae gaaacegeea ageagegaga ggeegaggee480aeggeeaage ggeegegeae gaceateee geeaageage tggagaeget gaagageget540tacaacaeet egeeeaagee ggeggegeeae gtgegegaga ageteteegte egagaeggege600etggaeatge gegtggtgea ggtttggtte eagaacegee gggeeaagga geeeteege720ggeegeteea agteggaeaa ggaeagett eaggaggge aggaeageea egetgagget780teetteeeeg atgageette ettggeeggaa atgggeeegg ecaatggeet etaeeggaag840	ctcaagtgca gegactgeca caegecaetg geegageget getteageeg aggggagage	240
ggcatcccgc ccacgcaggt ggtgcgccgc gcccaggact tcgtgtacca cctgcactgc360tttgcctgcg tcgtgtgcaa gcggcagctg gccacgggcg acgagttcta cctcatggag420gacagccggc tcgtgtgcaa ggcggactac gaaaccgcca agcagcgag ggccgaggcc480acggccaagc ggccgcgcac gaccatcacc gccaagcagc tggagacgct gaagagcgct540tacaacacct cgcccaagce ggcggcgcca gtgcgcgag aggttcta cgaaaccgcc gggccaagga gacgaggct660aagaaggacg ccggccggca gcgctgggg cagtattcc gcaacatga gcgctcccgc720ggcggctcca agtcggacaa ggacagcgt caggaggg aggacagcg cgctgaggt cgctgaggt840	gtttactgca aggacgactt tttcaagcgc ttcgggacca agtgcgccgc gtgccagctg	300
tttgcctgcg tcgtgtgcaa gcggcagctg gccacgggcg acgagttcta cctcatggag420gacagccggc tcgtgtgcaa ggcggactac gaaaccgcca agcagcgag ggccgaggcc480acggccaage ggccgcgcac gaccatcace gccaagcage tggagacgct gaagaggcgt540tacaacacet cgeccaagee ggcggcgcae gtgcgcgage agetetegte cgagacgggc600ctggacatge gcgtggtgca ggtttggtte cagaaccgce gggccaagga ggcgtggg660aagaaggacg ccggccggca gcgctggggg cagtattee gcaacatgaa gcgetecege720ggcggctcca agteggacaa ggacagegt caggaggge aggacagega cgetgaggt840	ggcatcccgc ccacgcaggt ggtgcgccgc gcccaggact tcgtgtacca cctgcactgc	360
gacageegge tegtgtgeaa ggeggactae gaaacegeea ageageggag ggeegaggee480aeggeeaage ggeegegeae gaceateaee geeaageage tggagaeget gaagageget540tacaacaeet egeecaagee ggeegegeeae gtgeegegage agetetegte egagaeggege600etggaeatge gegtggtgea ggtttggtte eagaacegee gggeeaagga gaagaggetg660aagaaggaeg eeggeeggeea geegetggggg eagtattee geaacatgaa geegeteeege720ggeeggeteea agteeggaea ggaeagegtt eaggaggge aggaeagega egetgaggte780teetteeeeg atgageette ettggeeggaa atgggeeegg eeaatggeet etaeeggaag840	tttgcctgcg tcgtgtgcaa gcggcagctg gccacgggcg acgagttcta cctcatggag	420
acggccaage ggecgegeae gaccateaee gecaageage tggagaeget gaagageget540tacaacaeet egeceaagee ggegegeeae gtgegegage agetetegte egagaeggge600ctggacatge gegtggtgea ggtttggtte cagaacegee gggccaagga gaagaggetg660aagaaggaeg eeggeeggea gegetggggg eagtattee geaacatgaa gegeteeege720ggeeggeteea agteggaeaa ggacagegt eaggaegge aggaeagega egetgaggte780teetteeeeg atgageette ettggeeggaa atgggeeegg eeaatggee teaeggaegg840	gacagccggc tcgtgtgcaa ggcggactac gaaaccgcca agcagcgaga ggccgaggcc	480
tacaacacct cgcccaagce ggcgcgccae gtgcgcgage agetetegte cgagaeggge600ctggaeatge gegtggtgea ggtttggtte cagaaeegge gggceaagga gaagaggetg660aagaaggaeg eeggeeggea gegetggggg eagtattee geaaeatgaa gegeteegge720ggeeggeteea agteggaeaa ggaeagegtt eaggagggge aggaeagega egetgaggte780teetteeeeg atgageette ettggeeggaa atgggeeegg eeaatggee teaeggaage840	acggccaagc ggccgcgcac gaccatcacc gccaagcagc tggagacgct gaagagcgct	540
ctggacatgc gcgtggtgca ggtttggttc cagaaccgcc gggccaagga gaagaggctg 660 aagaaggacg ccggccggca gcgctggggg cagtatttcc gcaacatgaa gcgctcccgc 720 ggcggctcca agtcggacaa ggacagcgtt caggaggggc aggacagcga cgctgaggtc 780 tccttccccq atgagccttc cttgqcqqaa atgqqcccqq ccaatqqcct ctacqqqaqc 840	tacaacacct cgcccaagcc ggcgcgccac gtgcgcgagc agctctcgtc cgagacgggc	600
aagaaggacg ccggccggca gcgctggggg cagtatttee gcaacatgaa gegeteege 720 ggeggeteea agteggacaa ggacagegtt caggaggggge aggacagega egetgaggte 780 teetteeeeg atgageette ettggeggaa atgggeeegg ecaatggeet etaegggage 840	ctggacatgc gcgtggtgca ggtttggttc cagaaccgcc gggccaagga gaagaggctg	660
ggeggeteea agteggaeaa ggaeagegtt caggagggge aggaeagega egetgaggte 780	aagaaggacg ccggccggca gcgctggggg cagtatttcc gcaacatgaa gcgctcccgc	720
teetteeeeq atgageette ettggeggaa atgggeeegg ceaatggeet etaegggage 840	ggeggeteea agteggacaa ggacagegtt caggagggge aggacagega egetgaggte	780
	teetteeeeg atgageette ettggeggaa atgggeeegg ceaatggeet etaegggage	840

ttgggggaac ccacccaggc cttgggccgg ccctcgggag ccctgggcaa cttctccctg	900
gagcatggag geetggeagg eecagageag taeegagage tgegteeegg eageeeetae	960
ggtgteeece cateceeege egeceegeag ageeteeetg geeeeeagee eeteetetee	1020
ageetggtgt acceagaeae eagettggge ettgtgeeet egggageeee eggegggeee	1080
ccacccatga gggtgctggc agggaacgga cccagttctg acctatccac ggggagcagc	1140
gggggttacc ccgacttccc tgccagcccc gcctcctggc tggatgaggt agaccacgct	1200
cagtteteag geeteatggg eccagettte ttgtae	1236
<210> SEQ ID NO 64 <211> LENGTH: 300 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthet polynucleotide	tic
<400> SEQUENCE: 64	
atggaaggaa tcatgaaccc ctacacggct ctgcccaccc cacagcagct cctggccatc	60
gagcagagtg tctacagctc agatcccttc cgacagggtc tcaccccacc ccagatgcct	120
ggagaccaca tgcaccetta tggtgeegag eccettttee atgaeetgga tagegaegae	180
acctccctca gtaacctggg tgactgtttc ctagcaacct cagaagctgg gcctctgcag	240
tccagagtgg gaaaccccat tgaccatctg tactccatgc agaattctta cttcacatct	300
<210> SEQ ID NO 65 <211> LENGTH: 1419 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthet polynucleotide	tic
<400> SEQUENCE: 65	
atggggagaa aaaagattca gattacgagg attatggatg aacgtaacag acaggtgaca	60
tttacaaaga ggaaatttgg gttgatgaag aaggettatg agetgagegt getgtgtgae	120
tgtgagattg cgctgatcat cttcaacagc accaacaagc tgttccagta tgccagcacc	180
gacatggaca aagtgcttet caagtacaeg gagtacaaeg ageegeatga gageeggaea	240
aactcagaca tcgtggagac gttgagaaag aagggcctta atggctgtga cagcccagac	300
cccgatgcgg acgattccgt aggtcacagc cctgagtctg aggacaagta caggaaaatt	360
aacgaagata ttgatctaat gatcagcagg caaagattgt gtgctgttcc acctcccaac	420
ttcgagatge cagteteeat eccagtgtee agecacaaca gtttggtgta cageaaceet	480
gtcageteae tgggaaaeee caacetattg ceaetggete accettetet geagaggaat	540
agtatgtete etggtgtaac acategaeet eeaagtgeag gtaacaeagg tggtetgatg	600
ggtggagacc tcacgtctgg tgcaggcacc agtgcaggga acgggtatgg caatccccga	660
aactcaccag gtctgctggt ctcacctggt aacttgaaca agaatatgca agcaaaatct	720
cctcccccaa tgaatttagg aatgaataac cgtaaaccag atctccgagt tcttattcca	780
ccaggcagca agaatacgat gccatcagtg tctgaggatg tcgacctgct tttgaatcaa	840
aggataaata acteecagte ggeteagtea ttggetaeee cagtggttte egtageaaet	900

cctactttac caggacaagg aatgggagga tatccatcag ccatttcaac aacatatggt	960
accgagtact ctctgagtag tgcagacctg tcatctctgt ctgggtttaa caccgccagc	1020
gctcttcacc ttggttcagt aactggctgg caacagcaac acctacataa catgccacca	1080
tetgeeetea gteagttggg agettgeact ageacteatt tateteagag tteaaatete	1140
teeetgeett etaeteaaag eeteaacate aagteagaae etgtttetee teetagagae	1200
cgtaccacca ccccttcgag atacccacaa cacacgcgcc acgaggcggg gagatctcct	1260
gttgacaget tgageagetg tageagtteg taegaeggga gegaeegaga ggateaeegg	1320
aacgaattee acteececat tggacteace agacettege eggaegaaag ggaaagteee	1380
tcagtcaagc gcatgcgact ttctgaagga tgggcaaca	1419
<210> SEQ ID NO 66 <211> LENGTH: 807 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	etic
<400> SEQUENCE: 66	
atggeeeage eeetgtgeee geegetetee gagteetgga tgetetetge ggeetgggge	60
ccaactogge ggeogeogee etcogacaag gactgoggee geteeetegt etcgteecea	120
gactcatggg gcagcacccc agccgacagc cccgtggcga gcccgcgcg gccaggcacc	180
ctccgggacc cccgcgcccc ctccgtaggt aggcgcggcg cgcgcagcag ccgcctgggc	240
agegggeaga ggeagagege eagtgagegg gagaaaetge geatgegeae getggeeege	300
geeetgeacg agetgegeeg etttetaeeg eegteegtgg egeeegeggg eeagageetg	360
accaagatog agaogotgog ootggotato ogotatatog gooacotgto ggoogtgota	420
ggeeteageg aggagagtet eeagegeegg tgeeggeage geggtgaege ggggteeeet	480
cggggctgcc cgctgtgccc cgacgactgc cccgcgcaga tgcagacacg gacgcaggct	540
gaggggcagg ggcaggggcg cgggctgggc ctggtatccg ccgtccgcgc cggggcgtcc	600
tggggateee egeetgeetg eeeeggagee egagetgeae eegageegeg egaeeegeet	660
gcgctgttcg ccgaggcggc gtgcccggaa gggcaggcga tggagccaag cccaccgtcc	720
ccgctccttc cgggcgacgt gctggctctg ttggagacct ggatgcccct ctcgcctctg	780
gagtggctgc ctgaggagcc caagttg	807
<210> SEQ ID NO 67 <211> LENGTH: 1239 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	etic
<400> SEQUENCE: 67	
atgctggaaa tgctagaata taatcactat caggtgcaga cccacctcga aaaccccacc	60
aagtaccaca tacagcaagc ccaacggcag caggtaaagc agtacctttc taccacttta	120
gcaaataaac atgccaacca agtcctgagc ttgccatgtc caaaccagcc tggcgatcat	180
gtcatgccac cggtgccggg gagcagcgca cccaacagcc ccatggctat gcttacgctt	240

aactccaact	gtgaaaaaga	gggattttat	aagtttgaag	agcaaaacag	ggcagagagc	300
gagtgcccag	gcatgaacac	acattcacga	gcgtcctgta	tgcagatgga	tgatgtaatc	360
gatgacatca	ttagcctaga	atcaagttat	aatgaggaaa	tcttgggctt	gatggatcct	420
gctttgcaaa	tggcaaatac	gttgcctgtc	tcgggaaact	tgattgatct	ttatggaaac	480
caaggtctgc	ccccaccagg	cctcaccatc	agcaactcct	gtccagccaa	ccttcccaac	540
ataaaaaggg	agctcacaga	gtctgaagca	agagcactgg	ccaaagagag	gcagaaaaag	600
gacaatcaca	acctgattga	acgaagaaga	agatttaaca	taaatgaccg	cattaaagaa	660
ctaggtactt	tgattcccaa	gtcaaatgat	ccagacatgc	gctggaacaa	gggaaccatc	720
ttaaaagcat	ccgtggacta	tatccgaaag	ttgcaacgag	aacagcaacg	cgcaaaagaa	780
cttgaaaacc	gacagaagaa	actggagcac	gccaaccggc	atttgttgct	cagaatacag	840
gaacttgaaa	tgcaggctcg	agctcatgga	ctttccctta	ttccatccac	gggtetetge	900
tctccagatt	tggtgaatcg	gatcatcaag	caagaacccg	ttcttgagaa	ctgcagccaa	960
gaceteette	agcatcatgc	agacctaacc	tgtacaacaa	ctctcgatct	cacggatggc	1020
accatcacct	tcaacaacaa	cctcggaact	gggactgagg	ccaaccaagc	ctatagtgtc	1080
cccacaaaaa	tgggatccaa	actggaagac	atcctgatgg	acgacaccct	ttctcccgtc	1140
ggtgtcactg	atccactcct	ttcctcagtg	tcccccggag	cttccaaaac	aagcagccgg	1200
aggagcagta	tgagcatgga	agagacggag	cacacttgt			1239
<210> SEQ : <211> LENG <212> TYPE <213> ORGAI <220> FEAT <223> OTHEI polyn <400> SEQUI	ID NO 68 TH: 1317 : DNA NISM: Artif: JRE: R INFORMATI( nucleotide ENCE: 68	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	ience: Synthe	etic
atgeccetca	acgttagctt	caccaacagg	aactatgacc	tcgactacga	ctcggtgcag	60
ccgtatttct	actgcgacga	ggaggagaac	ttctaccagc	agcagcagca	gagcgagctg	120
cagcccccgg	cgcccagcga	ggatatctgg	aagaaattcg	agctgctgcc	caccccgccc	180
ctgtccccta	gccgccgctc	cgggctctgc	tcgccctcct	acgttgcggt	cacacccttc	240
tcccttcggg	gagacaacga	cggcggtggc	gggagcttct	ccacggccga	ccagctggag	300
atggtgaccg	agctgctggg	aggagacatg	gtgaaccaga	gtttcatctg	cgacccggac	360
gacgagacct	tcatcaaaaa	catcatcatc	caggactgta	tgtggagcgg	cttctcggcc	420
gccgccaagc	tcgtctcaga	gaagctggcc	tcctaccagg	ctgcgcgcaa	agacagcggc	480
agcccgaacc	ccgcccgcgg	ccacagcgtc	tgctccacct	ccagcttgta	cctgcaggat	540
ctgagcgccg	ccgcctcaga	gtgcatcgac	ccctcggtgg	tcttccccta	ccctctcaac	600
gacagcagct						
	cgcccaagtc	ctgcgcctcg	caagactcca	gcgccttctc	tccgtcctcg	660
gattetetge	cgcccaagtc tctcctcgac	ctgcgcctcg ggagtcctcc	caagactcca ccgcagggca	gcgccttctc gccccgagcc	tccgtcctcg cctggtgctc	660 720
gattetetge catgaggaga	cgcccaagtc tctcctcgac caccgcccac	ctgcgcctcg ggagtcctcc caccagcagc	caagactcca ccgcagggca gactctgagg	gcgccttctc gccccgagcc aggaacaaga	tccgtcctcg cctggtgctc agatgaggaa	660 720 780
gattctctgc catgaggaga gaaatcgatg	cgcccaagtc tctcctcgac caccgcccac ttgtttctgt	ctgcgcctcg ggagtcctcc caccagcagc ggaaaagagg	caagactcca ccgcagggca gactctgagg caggctcctg	gcgccttctc gccccgagcc aggaacaaga gcaaaaggtc	tccgtcctcg cctggtgctc agatgaggaa agagtctgga	660 720 780 840

cacgtctcca	cacatcagca	caactacgca	gcgcctccct	ccactcggaa	ggactatcct	960	
gctgccaaga	gggtcaagtt	ggacagtgtc	agagtcctga	gacagatcag	caacaaccga	1020	
aaatgcacca	gccccaggtc	ctcggacacc	gaggagaatg	tcaagaggcg	aacacacaac	1080	
gtcttggagc	gccagaggag	gaacgagcta	aaacggagct	ttttgccct	gcgtgaccag	1140	
atcccggagt	tggaaaacaa	tgaaaaggcc	cccaaggtag	ttatccttaa	aaaagccaca	1200	
gcatacatcc	tgtccgtcca	agcagaggag	caaaagctca	tttctgaaga	ggacttgttg	1260	
cggaaacgac	gagaacagtt	gaaacacaaa	cttgaacagc	tacggaactc	ttgtgcg	1317	
<210> SEQ I <211> LENGT <212> TYPE: <213> ORGAN <220> FEATU <223> OTHER polyn	D NO 69 H: 618 DNA ISM: Artifi RE: INFORMATIC ucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	aence: Synthet	tic	
<400> SEQUE	NCE: 69						
atggactacg	actcgtacca	gcactatttc	tacgactatg	actgcgggga	ggatttctac	60	
cgctccacgg	cgcccagcga	ggacatctgg	aagaaattcg	agctggtgcc	atcgcccccc	120	
acgtcgccgc	cctggggctt	gggtcccggc	gcaggggacc	cggcccccgg	gattggtccc	180	
ccggagccgt	ggcccggagg	gtgcaccgga	gacgaagcgg	aatcccgggg	ccactcgaaa	240	
ggctggggca	ggaactacgc	ctccatcata	cgccgtgact	gcatgtggag	cggcttctcg	300	
gcccgggaac	ggctggagag	agctgtgagc	gaccggctcg	ctcctggcgc	gccccggggg	360	
aacccgccca	aggcgtccgc	cgccccggac	tgcactccca	gcctcgaagc	cggcaacccg	420	
gcgcccgccg	ccccctgtcc	gctgggcgaa	cccaagaccc	aggcctgctc	cgggtccgag	480	
agcccaagcg	actcgggtaa	ggaceteece	gagccatcca	agaggggggcc	accccatggg	540	
tggccaaagc	tctgcccctg	cctgaggtca	ggcattggct	cttctcaagc	tcttgggcca	600	
tctccgcctc	tetttgge					618	
<pre>&lt;210&gt; SEQ I &lt;211&gt; LENGT &lt;212&gt; TYPE: &lt;213&gt; ORGAN &lt;220&gt; FEATU &lt;223&gt; OTHER polyn &lt;400&gt; SEQUE</pre>	D NO 70 H: 1392 DNA IISM: Artifi RE: INFORMATIC ucleotide NCE: 70	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	ence: Synthet	tic	
atgeegagtt	gttccacgtc	tacgatgcca	ggaatgatat	gcaagaaccc	cgacttggag	60	
tttgactctt	tgcaaccatg	cttttatccg	gatgaagacg	acttttattt	cggcggcccg	120	
gacagcaccc	ctcctggaga	ggacatctgg	aaaaaattcg	aacttttgcc	tacaccccca	180	
ctcagtccct	ctcgaggatt	tgcggaacac	agcagtgaac	cgccgtcttg	ggtgacagag	240	
atgeteeteg	agaacgaatt	gtggggaagc	cctgcggagg	aagacgcttt	cgggctcggt	300	
ggactcggag	gtctcacgcc	gaacccagtc	atactgcagg	attgcatgtg	gtctggattc	360	
tcagctcggg	agaagctgga	acgggcagtt	tctgagaaac	tccaacatgg	ccggggccct	420	
ccaacagcgg	gttctaccgc	acagtcccct	ggtgctggag	ccgctagtcc	cgcgggggaga	480	
ggccatgggg	gcgcggcagg	agcgggtagg	geeggegetg	cgttgcctgc	tgagcttgcg	540	
				2			

-continu	ed
CONCING	~~

caccccgccg	ctgaatgtgt	agatcccgcg	gtagtgtttc	cgttccccgt	taataagcga	600	
gaaccggcac	cggtgccagc	cgctcctgcg	tctgcacccg	cggcaggtcc	tgctgtcgcc	660	
tcaggagcag	gtattgccgc	tcctgcaggg	gcaccaggag	tageceetee	aaggcccggc	720	
ggtaggcaaa	cctccggcgg	cgaccacaaa	gcactctcaa	cgagcggaga	ggatacactg	780	
tccgatagtg	atgacgagga	cgacgaagag	gaggacgagg	aggaggagat	agatgttgtc	840	
acggtcgaga	agcgaaggag	ttcttcaaat	acaaaagcgg	taacgacatt	cacgataaca	900	
gtaagaccta	agaacgcagc	cctcggtcca	gggcgggccc	agtccagtga	gcttatactt	960	
aagcgctgcc	tgccgattca	ccagcagcat	aactacgcgg	cccctagtcc	ctacgttgag	1020	
agcgaggatg	ccccccaca	aaaaaaata	aagtctgaag	cgtccccccg	ccccctgaaa	1080	
tccgtaatcc	ccccaaaggc	gaagtcactc	agtcccagga	attcagattc	cgaggactcc	1140	
gaacggcggc	ggaatcataa	catacttgag	agacaacgac	gcaatgacct	gaggtettet	1200	
tttttgaccc	tccgagatca	cgtccccgag	ctggttaaga	atgagaaagc	tgcgaaggta	1260	
gtcatactga	aaaaggccac	cgagtatgtc	catagtttgc	aagctgagga	gcaccagctt	1320	
ctccttgaaa	aggagaaact	tcaggcacga	caacagcaat	tgctgaaaaa	gattgagcat	1380	
gcacgcactt	gt					1392	
<pre>&lt;213&gt; 0KGA &lt;220&gt; FEAT &lt;223&gt; 0THE poly &lt;400&gt; SEQU</pre>	URE: R INFORMATION nucleotide ENCE: 71	DN: Descrip	tion of Art:	ificial Sequ	lence: Synth	netic	
atggagctac	tgtcgccacc	gctccgcgac	gtagacctga	cggcccccga	cggctctctc	60	
tgctcctttg	ccacaacgga	cgacttctat	gacgacccgt	gtttcgactc	cccggacctg	120	
cgcttcttcg	aagacctgga	cccgcgcctg	atgcacgtgg	gcgcgctcct	gaaacccgaa	180	
gagcactcgc	acttccccgc	ggcggtgcac	ccggccccgg	gcgcacgtga	ggacgagcat	240	
gtgcgcgcgc	ccagcgggca	ccaccaggcg	ggccgctgcc	tactgtgggc	ctgcaaggcg	300	
tgcaagcgca	agaccaccaa	cgccgaccgc	cgcaaggccg	ccaccatgcg	cgagcggcgc	360	
cgcctgagca	aagtaaatga	ggcctttgag	acactcaagc	gctgcacgtc	gagcaatcca	420	
aaccagcggt	tgcccaaggt	ggagatcctg	cgcaacgcca	tccgctatat	cgagggcctg	480	
caggetetge	tgcgcgacca	ggacgccgcg	ccccctggcg	ccgcagccgc	cttctatgcg	540	
ccgggcccgc	tgcccccggg	ccgcggcggc	gagcactaca	gcggcgactc	cgacgcgtcc	600	
agcccgcgct	ccaactgctc	cgacggcatg	atggactaca	geggeeeeee	gagcggcgcc	660	
cggcggcgga	actgctacga	aggcgcctac	tacaacgagg	cgcccagcga	acccaggccc	720	
gggaagagtg	cggcggtgtc	gagcctagac	tgcctgtcca	gcatcgtgga	gcgcatctcc	780	
accgagagcc	ctgcggcgcc	cgccctcctg	ctggcggacg	tgccttctga	gtcgcctccg	840	
cgcaggcaag	aggctgccgc	ccccagcgag	ggagagagca	gcggcgaccc	cacccagtca	900	
ccggacgccg	ccccgcagtg	ccctgcgggt	gcgaacccca	acccgatata	ccaggtgctc	960	

<210> SEQ ID NO 72 <211> LENGTH: 672

```
-continued
```

<212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 72 atggagetgt atgagacate eccetaette taccaggaae eccegetteta tgatggggaa 60 aactacctgc ctgtccacct ccagggcttc gaaccaccag gctacgagcg gacggagctc 120 accetgagee eegaggeeee agggeeeett gaggacaagg ggetggggae eeegageae 180 tgtccaggcc agtgcctgcc gtgggcgtgt aaggtgtgta agaggaagtc ggtgtccgtg 240 300 qaccqqcqqc qqqcqqccac actqaqqqaq aaqcqcaqqc tcaaqaaqqt qaatqaqqcc ttcgaggccc tgaagagaag caccctgctc aaccccaacc agcggctgcc caaggtggag 360 atcctgcgca gtgccatcca gtacatcgag cgcctccagg ccctgctcag ctccctcaac 420 480 caggaggagg gtgacctccg ctaccggggc gggggcgggc cccagccagg ggtgcccagc qaatqcaqct ctcacaqcqc ctcctqcaqt ccaqaqtqqq qcaqtqcact qqaqttcaqc 540 gccaacccag gggatcatct gctcacggct gaccctacag atgcccacaa cctgcactcc 600 ctcacctcca tcgtggacag catcacagtg gaagatgtgt ctgtggcctt cccagatgaa 660 672 accatocca ac <210> SEO ID NO 73 <211> LENGTH: 1068 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 73 atgaccaaat cgtacagcga gagtgggctg atgggcgagc ctcagcccca aggtcctcca 60 agetggaeag acgagtgtet cagtteteag gaegaggage acgaggeaga caagaaggag 120 180 gacgacctcg aagccatgaa cgcagaggag gactcactga ggaacggggg agaggaggag 240 gacgaagatg aggacctgga agaggaggaa gaagaggaag aggaggatga cgatcaaaag cccaagagac gcggccccaa aaagaagaag atgactaagg ctcgcctgga gcgttttaaa 300 ttgagacgca tgaaggctaa cgcccgggag cggaaccgca tgcacggact gaacgcggcg 360 ctagacaacc tgcgcaaggt ggtgccttgc tattctaaga cgcagaagct gtccaaaatc 420 gagactetge gettggeeaa gaactacate tgggetetgt eggagateet gegeteagge 480 aaaageecag acetggtete ettegtteag acgetttgea agggettate ceaaceeace 540 600 accaacctgg ttgcgggctg cctgcaactc aatcctcgga cttttctgcc tgagcagaac caggacatge ceceecacet geogaeggee agegetteet teeetgtaca eccetactee 660 taccagtege etgggetgee cagteegeet taeggtacea tggacagete ceatgtette 720 cacgttaagc ctccgccgca cgcctacagc gcagcgctgg agcccttctt tgaaagccct 780 ctgactgatt gcaccagccc ttcctttgat ggacccctca gcccgccgct cagcatcaat 840 ggcaacttct ctttcaaaca cgaaccgtcc gccgagtttg agaaaaatta tgcctttacc 900 atgcactate etgeagegae actggeaggg geecaaagee aeggateaat etteteagge 960 accgctgccc ctcgctgcga gatccccata gacaatatta tgtccttcga tagccattca 1020

catcatgagc gagtcatgag tgccca	gctc aatgccatat ttcatgat	1068
<210> SEQ ID NO 74 <211> LENGTH: 711 <212> TYPE: DNA <213> ORGANISM: Artificial S <220> FEATURE: <223> OTHER INFORMATION: Des polynucleotide	equence cription of Artificial Sequence	: Synthetic
<400> SEQUENCE: 74		
atgecageee geettgagae etgeat	ctcc gacctcgact gcgccagcag cagc	ggcagt 60
gacctateeg getteeteae egaega	uggaa gactgtgcca gactccaaca ggca	geetee 120
gettegggge egecegegee ggeeeg	cagg ggcgcgccca atatctcccg ggcg	tctgag 180
gttccagggg cacaggacga cgagca	lggag aggeggegge geegeggeeg gaeg	cgggtc 240
cgctccgagg cgctgctgca ctcgct	gcgc aggagccggc gcgtcaaggc caac	gatcgc 300
gagegeaace geatgeacaa ettgaa	legeg geeetggaeg caetgegeag egtg	ctgccc 360
tcgttccccg acgacaccaa gctcac	caaa atcgagacgc tgcgcttcgc ctac	aactac 420
atctgggctc tggccgagac actgcg	cctg gcggatcaag ggctgcccgg aggc	ggtgcc 480
cgggagcgcc tcctgccgcc gcagtg	cgtc ccctgcctgc ccggtccccc aago	eccgcc 540
agcgacgcgg agtcctgggg ctcagg	tgee geegeegeet eeeegetete tgae	ecccagt 600
ageeeageeg eeteegaaga etteac	etac cgccccggcg accctgtttt ctcc	ttccca 660
ageetgeeca aagaettget eeacae	aacg ccctgtttca ttccttacca c	711
<210> SEQ ID NO 75 <211> LENGTH: 642 <212> TYPE: DNA <213> ORGANISM: Artificial S <220> FEATURE: <223> OTHER INFORMATION: Des polynucleotide	equence cription of Artificial Sequence	: Synthetic
<400> SEQUENCE: 75		
atgacaccac aaccatctgg tgctcc	caca gtccaggtga cgcgagagac tgaa	agatca 60
tteecaegeg egteegagga tgaggt	gaca tgtccaacta gcgcaccccc ctct	cctacc 120
cggacccgcg ggaattgtgc tgaggc	cgaa gagggaggat gcagaggagc acca	aggaaa 180
cttcgagccc gacggggtgg aagaag	ccgc cccaagtctg agctcgccct tagc	aagcag 240
cgccgcagtc ggaggaaaaa ggcaaa	logac ogggaaagga ataggatgca taat	cttaat 300
tetgetetgg acgetetgeg aggegt	actt cctactttcc cggatgacgc gaaa	ttgacc 360
aagatagaga ctctccggtt tgcaca	taat tacatctggg ctcttacaca aaca	ctgaga 420
attgccgatc acagtcttta cgctct	tgag ccacccgccc cgcactgtgg cgag	ctgggt 480
ageeeeggeg geteteetgg agaetg	gggg tetttgtatt eteetgteag eeaa	gcggga 540
tetttgagte eggetgeeag tetega	agaa agacccggac tccttggagc gact	ttttca 600
gcatgtctgt cccctggctc attggc	tttc tcagactttt tg	642

<210> SEQ ID NO 76 <211> LENGTH: 741 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence

<220> FEATU <223> OTHEN polyn	JRE: R INFORMATIO nucleotide	DN: Descript	ion of Arti	ificial Sequ	lence: Synth	etic	
<400> SEQUI	ENCE: 76						
atggccctgc	ctcccagccc	gctggccatg	gaatatgtca	atgactttga	cttgatgaag	60	
tttgaggtaa	agcgggaacc	ctctgagggc	cgacctggcc	cacctacagc	ctcactggga	120	
tccacacctt	acagctcagt	gcctccttca	cccaccttca	gtgaaccagg	catggtaggg	180	
gcaaccgagg	gtacacgacc	aggtttggag	gagctgtact	ggcttgctac	cctgcagcag	240	
cagcttgggg	ctggggaggc	attgggactg	agtcctgaag	aggccatgga	gctactgcaa	300	
ggtcagggcc	cagtccctgt	tgatggaccc	catggttact	acccagggag	cccagaggag	360	
acaggagccc	agcacgttca	gttggcagag	cggttttccg	acgcggcgct	tgtctcgatg	420	
tctgtgcgag	aactaaaccg	gcagctgcgg	ggatgcggga	gagacgaggc	tctacgactg	480	
aagcagaggc	gtcgaacgct	gaagaaccgt	ggctatgcgc	aagcatgtcg	ttccaagagg	540	
ctgcaacaga	ggcgaggtct	tgaggccgag	cgcgcccgtc	ttgcagccca	gctagatgcg	600	
ctacgagctg	aagtagcacg	tttggcaaga	gagcgagatc	tctacaaggc	tcgctgtgac	660	
cggctaacct	cgagtggccc	cgggtccggg	gatccctccc	accttttcct	ctgcccaact	720	
ttcttgtaca	aagttgtccc	с				741	
<210> SEQ 3 <211> LENG <212> TYPE <213> ORGAN <220> FEATU	ID NO 77 TH: 1395 : DNA VISM: Artif: JRE:	icial Sequer	nce				
<400> SEQUI	R INFORMATIC nucleotide ENCE: 77	DN: Descript	ion of Arti	ificial Sequ	lence: Synth	etic	
<pre>&lt;400&gt; SEQUI atgaacgcgc</pre>	R INFORMATIO nucleotide ENCE: 77 agctgaccat	DN: Descript ggaagcgatc	ion of Arti ggcgagctgc	ificial Sequ acggggtgag	ence: Synth ccatgagccg	etic 60	
<400> SEQUI atgaacgcgc gtgcccgccc	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct	DN: Descript ggaagcgatc gctgggcggc	ion of Art: ggcgagctgc agcccccacg	ificial Sequ acggggtgag cgcgcagctc	ence: Synth ccatgagccg cgtggcgcac	60 120	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc	<pre>&gt;N: Descript ggaagcgatc gctgggcggc cgcgcacccg</pre>	ggcgagctgc agcccccacg cgctccatgg	ificial Sequ acggggtgag cgcgcagctc gcatggcgtc	ence: Synth ccatgagccg cgtggcgcac cctgctggac	60 120 180	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc ggcggcagcc ggcggcagcg</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga	<pre>N: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac</pre>	ggcgagctgc agcccccacg cgctccatgg caccaccggg	ificial Sequ acggggtgag cgcgcagctc gcatggcgtc cccctgagca	ence: Synth ccatgagccg cgtggcgcac cctgctggac cagcctggcc	60 120 180 240	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc ggcggcagcg ggccccctgc</pre>	INFORMATIC nucleotide SNCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga atcccaccat	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc	acggggtgag cgcgcagctc gcatggcgtc cccctgagca	ence: Synth ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc	60 120 180 240 300	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc ggcggcagcg ggccccctgc accacctaca</pre>	agctgaccat ctgccgacct acctgcccgacct gcggcggaga atcccaccat	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc	acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat	ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc agtctcggac	60 120 180 240 300 360	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc ggcggcagcg ggccccctgc accacctaca aagttccccc</pre>	agctgaccat acctgccgacat acctgccgacat acctgcccacat accctgcccca accctgcccca accctcgccca	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag ccaccaccat	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc caccaccacc	acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca	ence: Synth ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc agtctcggac ccaccagcgc	60 120 180 240 300 360 420	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc ggcggcagcg ggcccctgc accacctaca aagttccccc ctggcggcagca</pre>	agctgaccat ctgccgacct acctgccgacct acctgccgacat ccacctgcccc gcggcggaga atcccaccat accaccttgac accatcacca	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag ccaccaccat tagcttcacg	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc caccaccacc	acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca	ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc agtctcggac ccaccagcgc gctggcctcc	60 120 180 240 300 360 420 480	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc ggcggcagcc ggcccctgc accacctaca aagttccccc ctggcgggca atgaataacc</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga atcccaccat ccaccttgac accatcacca acgtgagcgg tctatacccc	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag ccaccaccat tagcttcacg ctaccacaag	ggcgagctgc agcccccacg cgctccatgg tgcgagactc ccgctgcctc caccaccacc ctcatgcggg gacgtggccg	acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca atgagcgcgg gcatgggcca	ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc agtctcggac ccaccagcgc gctggcctcc gagcctctcg	etic 60 120 180 240 300 360 420 480 540	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc ggcggcagcg ggccccctgc accacctaca aagttccccc ctggcgggca atgaataacc cccctctcca</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga atcccaccat ccaccttgac accatcacca acgtgagcgg tctatacccc gctccggtct	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag ccaccaccat tagcttcacg ctaccacaag gggcagcatc	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc caccaccacc ctcatgcggg gacgtggccg cacaactccc	acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca atgagcgcgg gcatgggcca agcaagggct	ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc agtctcggac ccaccagcgc gctggcctcc gagcctctcg cccccactat	60 120 180 240 300 360 420 480 540 600	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc ggcggcagcc ggcccctgc accacctaca aagttccccc ctggcgggca atgaataacc cccctctcca gcccacccgg</pre>	INFORMATIC oucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga atcccaccat ccaccttgac accatcacca acgtgagcgg tctatacccc gctccggtct gggccgccat	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag ccaccaccat tagcttcacg gggcagcatc gcccaccgac	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc caccaccacc gacgtgcgg gacgtggccg cacaactccc aagatgctca	acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca atgagcgcgg gcatgggcca agcaagggct	ence: Synth ccatgagccg cgtggcgcac cctgctggac gagcatgccc agtctcggac gctggcctcc gagcctctcg cccccactat cttcgaagcc	etic 60 120 180 240 300 360 420 480 540 600 660	
<pre>&lt;2233 Offinity polyn &lt;400&gt; SEQUI atgaacgcgc gtgcccgccc ggcggcagcc ggcggcagcg ggccccctgc accacctaca aagttcccccc ctggcggggca atgaataacc cccctctcca gcccacccgg caccacccgg</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga atcccaccat ccaccttgac accatcacca acgtgagcgg tctatacccc gctccggtct gggccgccat ccatgctcgg	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag ccaccaccat tagcttcacg gggcagcatc gcccaccgac ccgccaccggg	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc caccaccacc ctcatgcggg gacgtggccg cacaactccc aagatgctca gagcagcacc	acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac accgcacca atgagcgcgg gcatgggcca agcaagggct cccccaacgg tcacgccaca	ence: Synth ccatgagccg cgtggcgcac cctgctggac cagcctggcc agtctcggac ccaccagcgc gctggcctcc gagcctctcg cccccactat cttcgaagcc ctcggccggc	etic 60 120 180 240 300 360 420 480 540 600 660 720	
<pre>&lt;2233 offilial polyn &lt;400&gt; SEQUI atgaacgcgc gtgcccgccc ggcggcagcc ggcggcagcg ggccccctgc accacctaca aagttccccc ctggcgggca atgaataacc cccctctcca gcccacccgg atggtgcca</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgcccccc gcggcggaga atcccaccat ccaccttgac accatcaccat gcgcggagg tctatacccc gctccggtct gggccgccat ccatgctcgg tcaacggcct	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccatctgcag ccaccaccat tagcttcacg gggcagcatc gcccaccgac ccgccacggg tcctccgcac	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc caccaccacc gacgtggcgg gacgtggcgg cacaactccc aagatgctca gagcagcacc catccccacg	acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca atgagcgcgg gcatgggcca agcaagggct cccccaacgg tcacgcccac	ence: Synth ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc agtctcggac ccaccagcgc gctggcctcc gagcctctcg cccccactat cttcgaagcc ctcggccggc	etic 60 120 180 240 300 360 420 480 540 600 660 720 780	
<pre>&lt;2233 offilial polyn &lt;400&gt; SEQUI atgaacgcgc gtgcccgccc ggcggcagcc ggcggcagcg ggcccctgc accacctaca aagttccccc ctggcgggca atgaataacc cccctctcca gcccacccgg atggtgcca atggtgcca cacgggcaac</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga atcccaccat ccaccttgac accatcacca acgtgagcgg tctatacccc gctccggtct gggccgccat ccatgctcgg tcaacggcct	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag ccaccaccat tagcttcacg gggcagcatc gcccaccgac ccgccaccggg tcctccgcac agcccgggag	ion of Art: ggcgagctgc agcccccacg cgctccatgg tgcgagactc ccgctgcctc caccaccacc ctcatgcggg gacgtggccg cacaactccc aagatgctca gagcagcacc catccccacg caccaccctt	ificial Sequ acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca atgagcgcgg gcatgggcca agcaagggct cccccaacgg tcaccgccacc cccacctgaa cggtgaccgg	ence: Synth ccatgagccg cgtggcgcac cctgctggac cagcctggcc agtctcggac gagcatgccc gagcctccg gctggcctcc gagcctctcg cccccactat cttcgaagcc ctcggccggc cgcccaggc	etic 60 120 180 240 300 360 420 480 540 600 600 660 720 780 840	
<pre>&lt;2233 Official polyn &lt;400&gt; SEQUI atgaacgcgc gtgcccgccc ggcggcagcc ggcggcagcg accacctaca aagttccccc ctggcgggca atgaataacc cccctctcca gcccacccgg atggtgccca cacgggcaac agcaatggaa</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga atcccaccat ccaccttgac accatcacca acgtgagcgg tctatacccc gggccgccat ccatgctcgg tcaacggcct tcctgggcac	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccctctgcag ccaccaccat tagcttcacg gggcagcatc gcccaccgac ccgccacggg tcctccgcac agcccgggag	ion of Art: ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc caccaccacc ctcatgcggg gacgtggccg cacaactccc aagatgctca gagcagcacc catccccacg cccaaccctt gagatcaata	ificial Sequ acggggtgag cgcgcagctc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca atgagcgcgg gcatgggcca agcaagggct cccccaacgg tcacgccaca cccactgaa cggtgaccgg ccaaagaggt	ence: Synth ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc agtctcggac ccaccagcgc gctggcctcc gagcctctcg cccccactat cttcgaagcc ctcggccggc cgcccaggc ggcgcaggtc	etic 60 120 180 240 300 360 420 480 540 600 660 720 780 840 900	
<pre>&lt;400&gt; SEQUI atgaacgcgc gtgcccgccc cgcggcagcc ggcggcagcc ggcgcagcg accacctaca aagttccccc ctggcgggca atgaataacc cccctctcca gcccacccgg atggtgcca cacgggcaac agcaatggaa atcaccaccg</pre>	INFORMATIC nucleotide ENCE: 77 agctgaccat ctgccgacct acctgccccc gcggcggaga atcccaccat ccaccttgac acgtgagcgg tctatacccc ggcccggtct gggccgccat ccatgctcgg tcaacggcct tcctgggcac	DN: Descript ggaagcgatc gctgggcggc cgcgcacccg ttaccaccac gaccatggcc ccaccatgcag ccaccaccat tagcttcacg gggcagcatc gcccaccgac ccgccacggg tcctccgcac agcccgggag gcagatggaa ctacagcatc	ggcgagctgc agcccccacg cgctccatgg caccaccggg tgcgagactc ccgctgcctc caccaccacc ctcatgcggg gacgtggccg cacaactccc aagatgctca gagcagcacc catccccacg cccaaccctt gagatcaata ccccaggcca	acggggtgag cgcgcagctc gcatggcgtc gcatggcgtc cccctgagca ccccaggtat ccatctccac acccgcacca atgagcgcgg gcatgggcca agcaagggct cccccaccga tcacgcccac cccacctgaa cggtgaccgg tcatggcca	ence: Synth ccatgagccg cgtggcgcac cctgctggac cagcctggcc gagcatgccc gagcatgccc gagcctccg gctggcctcc gagcctctcg ccccactat cttcgaagcc ctcggccggc cgccagggc ggcgcaggtc gagggtgctc	etic 60 120 180 240 300 360 420 480 540 600 660 720 780 840 900 960	

121

aaatccggcc gggagacctt ccggaggatg tggaagtggc tgcaggagcc ggagttccag 1080 cgcatgtccg cgctccgctt agcagcatgc aaaaggaaag aacaagaaca tgggaaggat 1140 agaggcaaca cacccaaaaa gcccaggttg gtcttcacag atgtccagcg tcgaactcta 1200 catgcaatat tcaaggaaaa taagcgtcca tccaaagaat tgcaaatcac catttcccag 1260 cagctggggt tggagctgag cactgtcagc aacttcttca tgaacgcaag aaggaggagt 1320 ctggacaagt ggcaggacga gggcagctcc aattcaggca actcatcttc ttcatcaagc 1380 acttgtacca aagca 1395 <210> SEQ ID NO 78 <211> LENGTH: 891 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 78 atgatgtett atettaagea acegeettae geagteaatg ggetgagtet gaceaetteg 60 ggtatggact tgetgeacee etcegtggge taccegggge eetgggette ttgteeegea 120 180 qccacccccc qqaaacaqcq ccqqqaqaqq acqacqttca ctcqqqcqca qctaqatqtq ctggaagcac tgtttgccaa gacccggtac ccagacatct tcatgcgaga ggaggtggca 240 ctgaaaatca acttgcccga gtcgagggtg caggtatggt ttaagaatcg aagagctaag 300 tgccgccaac aacagcaaca acagcagaat ggaggtcaaa acaaagtgag acctgccaaa 360 aagaagacat ctccagctcg ggaagtgagt tcagagagtg gaacaagtgg ccaattcact 420 cccccctcta gcacctcagt cccgaccatt gccagcagca gtgctcctgt gtctatctgg 480 ageccagett ceateteece actgteagat ceetgteea ectectete etgeatgeag 540 aggteetate ceatgaeeta taeteagget teaggttata gteaaggata tgetggetea 600 acttectact ttggggggcat ggaetgtgga teatatttga eccetatgea teaceagett 660 cccggaccag gggccacact cagtcccatg ggtaccaatg cagtcaccag ccatctcaat 720 cagteeccag ettetette caeccaggga tatggagett caagettggg ttttaactea 780 accactgatt gcttggatta taaggaccaa actgcctcct ggaagcttaa cttcaatgct 840 gactgettgg attataaaga teagacatee tegtggaaat teeaggtttt g 891 <210> SEQ ID NO 79 <211> LENGTH: 1554 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEOUENCE: 79 atggcggccc ttcccggcac ggtaccgaga atgatgcggc cggctccggg gcagaactac 60 ccccgcacgg gattcccttt ggaagtgtcc accccgcttg gccaaggccg ggtcaatcag 120 ctqqqaqqqq tcttcatcaa tqqqcqaccc ctqcctaacc acatccqcca caaqataqtq 180 gagatggeee accatggeat eeggeeetgt gteateteee gaeagetgeg tgteteeeae 240 ggetgegtet ceaagattet ttgeegetae caggagaeeg ggteeateeg geetggggee 300

atcggcggca	gcaagcccag	acaggtggcg	actccggatg	tagagaaaaa	gattgaggag	360
tacaagaggg	aaaacccagg	catgttcagc	tgggagatcc	gggacaggct	gctgaaggat	420
gggcactgtg	accgaagcac	tgtgccctca	gtgagttcga	ttagccgcgt	gctcagaatc	480
aagttcggga	agaaagagga	ggaggatgaa	gcggacaaga	aggaggacga	cggcgaaaag	540
aaggccaaac	acagcatcga	cggcatcctg	ggcgacaaag	ggaaccggct	ggacgagggc	600
tcggatgtgg	agtcggaacc	tgacctccca	ctgaagcgca	agcagcgacg	cagteggaee	660
acattcacgg	ccgagcagct	ggaggagctg	gagaaggcct	ttgagaggac	ccactaccca	720
gacatataca	cccgcgagga	gctggcgcag	aggaccaagc	tgacagaggc	gcgtgtgcag	780
gtctggttca	gtaaccgccg	cgcccgttgg	cgtaagcagg	caggagccaa	ccagctggcg	840
gcgttcaacc	accttctgcc	aggaggcttc	ccgcccaccg	gcatgcccac	getgeeeee	900
taccagctgc	cggactccac	ctaccccacc	accaccatct	cccaagatgg	gggcagcact	960
gtgcaccggc	ctcagcccct	gccaccgtcc	accatgcacc	agggcggggct	ggetgeageg	1020
gctgcagccg	ccgacaccag	ctctgcctac	ggagcccgcc	acagettete	cagctactct	1080
gacagcttca	tgaatccggc	ggcgccctcc	aaccacatga	acccggtcag	caacggcctg	1140
tctcctcagg	tgatgagcat	cttgggcaac	cccagtgcgg	tgcccccgca	gccacaggct	1200
gacttctcca	tctccccgct	gcatggcggc	ctggactcgg	ccacctccat	ctcagccagc	1260
tgcagccagc	gggccgactc	catcaagcca	ggagacagcc	tgcccacctc	ccaggcctac	1320
tgcccaccca	cctacagcac	caccggctac	agcgtggacc	ccgtggccgg	ctatcagtac	1380
ggccagtacg	gccagagtga	gtgcctggtg	ccctgggcgt	cccccgtccc	cattccttct	1440
cccaccccca	gggcctcctg	cttgtttatg	gagagctaca	aggtggtgtc	agggtgggga	1500
atgtccattt	cacagatgga	aaaattgaag	tccagccaga	tggaacagtt	cacc	1554
<210> SEQ : <211> LENG <212> TYPE <213> ORGAJ <220> FEATU <223> OTHEI polyi	ID NO 80 TH: 903 : DNA VISM: Artif: JRE: & INFORMATIC Ducleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	aence: Synthe	etic
<400> SEQU	ENCE: 80					
atgagttgcc	aagcttttac	ttcggctgat	acctttatac	ctctgaattc	tgacgcctct	60
gcaactctgc	ctctgataat	gcatcacagt	gctgccgagt	gtctaccagt	ctccaaccat	120
gccaccaatg	tgatgtctac	agcaacagga	cttcattatt	ctgttccttc	ctgtcattat	180
ggaaaccagc	catcaaccta	tggagtgatg	gcaggtagtt	taaccccttg	tctttataaa	240
tttcctgacc	acaccttgag	tcatggattt	cctcctatac	accagcctct	tctggcagag	300
gaccccacag	ctgctgattt	caagcaggaa	ctcaggcgga	aaagtaaatt	ggtggaagag	360
ccaatagaca	tggattctcc	agaaatcaga	gaacttgaaa	agtttgccaa	tgaatttaaa	420
gtgagacgaa	ttaaattagg	atacacccag	acaaatgttg	gggaggccct	ggcagctgtg	480
catggctctg	aattcagtca	aacaacaatc	tgccgatttg	aaaatctgca	gctcagcttt	540
aaaaatgcat	gcaaactgaa	agcaatatta	tccaaatggc	tggaggaagc	tgagcaagta	600
ggagctttgt	acaatgaaaa	agtgggagca	aatgaaagga	aaagaaaacg	aagaacaact	660

-continued

ataagcattg ctgctaaaga tgctctggag agacactttg gagaacagaa taaaccttct	720
tctcaagaga tcatgaggat ggctgaagaa ctgaatctgg agaaagaagt agtaagagtt	780
tggttttgca accggaggca gagagaaaaa cgggtgaaaa caagtctgaa tcagagttta	840
ttttctattt ctaaggaaca tcttgagtgc agatcaggcc tcatgggccc agctttcttg	900
tac	903
<210> SEQ ID NO 81 <211> LENGTH: 1080 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	tic
<400> SEQUENCE: 81	
atggegggae acetggette agattttgee ttetegeece etceaggtgg tggaggtgat	60
gggccagggg ggccggagcc gggctgggtt gatcetcgga cetggetaag ettecaagge	120
cctcctggag ggccaggaat cgggccgggg gttgggccag gctctgaggt gtgggggatt	180
cccccatgcc ccccgccgta tgagttctgt ggggggatgg cgtactgtgg gccccaggtt	240
ggagtggggc tagtgcccca aggcggcttg gagacctctc agcctgaggg cgaagcagga	300
gteggggtgg agageaacte egatggggee teeeeggage eetgeaeegt eaceeetggt	360
gccgtgaagc tggagaagga gaagctggag caaaacccgg aggagtccca ggacatcaaa	420
gctctgcaga aagaactcga gcaatttgcc aagctcctga agcagaagag gatcaccctg	480
ggatatacac aggeegatgt ggggeteace etggggggtte tatttgggaa ggtatteage	540
caaacgacca tetgeegett tgaggetetg cagettaget teaagaacat gtgtaagetg	600
cggcccttgc tgcagaagtg ggtggaggaa gctgacaaca atgaaaatct tcaggagata	660
tgcaaagcag aaaccctcgt gcaggcccga aagagaaagc gaaccagtat cgagaaccga	720
gtgagaggca acctggagaa tttgttcctg cagtgcccga aacccacact gcagcagatc	780
agccacatcg cccagcagct tgggctcgag aaggatgtgg tccgagtgtg gttctgtaac	840
cggcgccaga agggcaagcg atcaagcagc gactatgcac aacgagagga ttttgaggct	900
getgggtete ettteteagg gggaeeagtg teettteete tggeeeeagg geeeeattt	960
ggtaccccag gctatgggag ccctcacttc actgcactgt actectcggt ccctttccct	1020
gagggggaag cettteecee tgtetetgte accaetetgg geteteceat geatteaaac	1080
<210> SEQ ID NO 82 <211> LENGTH: 1440 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	tic
<400> SEQUENCE: 82	
atggetteag acageatatt tgagteattt eettegtaee eacagtgett eatgagagaa	60
tgcatacttg gaatgaatcc ttctagagac gtccacgatg ccagcacgag ccgccgcttc	120
acgccgcctt ccaccgcgct gagcccaggc aagatgagcg aggcgttgcc gctgggcgcc	180
ccggacgccg gcgctgccct ggccggcaag ctgaggagcg gcgaccgcag catggtggag	240

-continued	
gtgetggeeg accaeeeggg cgagetggtg egeaeegaea geeeeaaett eetetgetee	300
gtgctgccta cgcactggcg ctgcaacaag accctgccca tcgctttcaa ggtggtggcc	360
ctaggggatg ttccagatgg cactctggtc actgtgatgg ctggcaatga tgaaaactac	420
tcggctgagc tgagaaatgc taccgcagcc atgaagaacc aggttgcaag atttaatgac	480
ctcaggtttg tcggtcgaag tggaagaggg aaaagcttca ctctgaccat cactgtcttc	540
acaaacccac cgcaagtcgc cacctaccac agagccatca aaatcacagt ggatgggccc	600
cgagaacctc gaagacatcg gcagaaacta gatgatcaga ccaagcccgg gagcttgtcc	660
ttttccgagc ggctcagtga actggagcag ctgcggcgca cagccatgag ggtcagccca	720
caccacccag cccccacgcc caaccctcgt gcctccctga accactccac tgcctttaac	780
cctcagcctc agagtcagat gcaggataca aggcagatcc aaccatcccc accgtggtcc	840
tacgatcagt cctaccaata cctgggatcc attgcctctc cttctgtgca cccagcaacg	900
cccatttcac ctggacgtgc cagcggcatg acaaccctct ctgcagaact ttccagtcga	960
ctctcaacgg caccegacet gacagegtte agegaceege gecagtteee egegetgeee	1020
tccatctccg acccccgcat gcactatcca ggcgccttca cctactcccc gacgccggtc	1080
acctegggea teggeategg catgteggee atgggetegg ceaegegeta ceaeacetae	1140
ctgccgccgc cctaccccgg ctcgtcgcaa gcgcagggag gcccgttcca agccagctcg	1200
ccctcctacc acctgtacta cggcgcctcg gccggctcct accagttctc catggtgggc	1260
ggegageget egeegeegeg cateetgeeg ceetgeacea aegeeteeae eggeteegeg	1320
ctgctcaacc ccagcctccc gaaccagagc gacgtggtgg aggccgaggg cagccacagc	1380
aactccccca ccaacatggc gccctccgcg cgcctggagg aggccgtgtg gaggccctac	1440
<210> SEQ ID NO 83 <211> LENGTH: 852 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synth polynucleotide	hetic
<400> SEQUENCE: 83	
atgtcgatgc tgccgtcgtt tggctttacg caggagcaag tggcgtgcgt gtgcgaggtt	60
ctgcagcaag gcggaaacct ggagcgcctg ggcaggttcc tgtggtcact gcccgcctgc	120
gaccacctgc acaagaacga gagcgtactc aaggccaagg cggtggtcgc cttccaccgc	180
ggcaacttcc gtgageteta caagateetg gagageeace agttetegee teacaaceae	240
cccaaactgc agcaactgtg gctgaaggcg cattacgtgg aggccgagaa gctgtgcggc	300
cgacccctgg gcgccgtggg caaatatcgg gtgcgccgaa aatttccact gccgcgcacc	360
atctgggacg gcgaggagac cagctactgc ttcaaggaga agtcgagggg tgtcctgcgg	420
gagtggtacg cgcacaatcc ctacccatcg ccgcgtgaga agcgggagct ggccgaggcc	480
accggcetea ceaecaecea ggteageaae tggtttaaga accggaggea aagagaeegg	540
gccgcggagg ccaaggaaag ggagaacacc gaaaacaata actcctcctc caacaagcag	600
aaccaactct ctcctctgga agggggcaag ccgctcatgt ccagctcaga agaggaattc	660
tcacctcccc aaagtccaga ccagaactcg gtccttctgc tgcagggcaa tatgggccac	720
gccaggaget caaactatte teteeeggge ttaacageet egeageeeag teaeggeetg	780

cagacccacc agcatcagct ccaagactct ctgctcggcc ccctcacctc cagtctggtg	840
gacttgggggt cc	852
<210> SEQ ID NO 84 <211> LENGTH: 873 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthet polynucleotide	tic
<400> SEQUENCE: 84	
atgtccatgc tgcccacctt cggcttcacg caggagcaag tggcgtgcgt gtgcgaggtg	60
ctgcagcagg gcggcaacat cgagcggctg ggccgcttcc tgtggtcgct gcccgcctgc	120
gagcacette acaagaatga aagegtgete aaggeeaagg eegtggtgge etteeaeege	180
ggcaacttee gegageteta caagateetg gagageeace agttetegee geacaaceae	240
gccaagctgc agcagctgtg gctcaaggca cactacatcg aggcggagaa gctgcgcggc	300
cgacccctgg gcgccgtggg caaataccgc gtgcgccgca aattcccgct gccgcgctcc	360
atctgggacg gcgaggagac cagctactgc ttcaaggaaa agagtcgcag cgtgctgcgc	420
gagtggtacg cgcacaaccc ctacccttca ccccgcgaga agcgtgagct gacggaggcc	480
acgggcetea ceaceacae ggteageaae tggtteaaga aeeggeggea gegegaeegg	540
geggeegagg ceaaggaaag ggagaacaae gagaaeteea attetaacag ceacaaeeeg	600
ctgaatggca gcggcaagtc ggtgttaggc agctcggagg atgagaagac tccatcgggg	660
acgccagacc actcatcatc cagccccgca ctgctcctca gcccgccgcc ccctgggctg	720
ccgtccctgc acagectggg ccacectecg ggeeceageg cagtgeeagt geeggtgeea	780
ggcggaggtg gagcggaccc actgcaacac caccatggcc tgcaggactc catcctcaac	840
cccatgtcag ccaacctcgt ggacctgggc tcc	873
<210> SEQ ID NO 85 <211> LENGTH: 804 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthet polynucleotide	tic
<400> SEQUENCE: 85	
atgeegeget eetteetggt caagaageat tteaaegeet eeaaaaagee aaaetaeage	60
gaactggaca cacatacagt gattatttcc ccgtatctct atgagagtta ctccatgcct	120
gtcataccac aaccagagat cctcagctca ggagcataca gccccatcac tgtgtggact	180
accgctgctc cattecacge ccagetacee aatggeetet eteetette eggatactee	240
tcatctttgg ggcgagtgag tccccctcct ccatctgaca cctcctccaa ggaccacagt	300
ggctcagaaa gccccattag tgatgaagag gaaagactac agtccaagct ttcagacccc	360
catgccattg aagctgaaaa gtttcagtgc aatttatgca ataagaccta ttcaactttt	420
tctgggctgg ccaaacataa gcagctgcac tgcgatgccc agtctagaaa atctttcagc	480
tgtaaatact gtgacaagga atatgtgagc ctgggcgccc tgaagatgca tattcggacc	540
cacacattac cttgtgtttg caagatctgc ggcaaggcgt tttccagacc ctggttgctt	600

126

caaggacaca	ttagaactca	cacggggggag	aagcettttt	cttgccctca	ctgcaacaga	660
gcatttgcag	acaggtcaaa	tctgagggct	catctgcaga	cccattctga	tgtaaagaaa	720
taccagtgca	aaaactgctc	caaaaccttc	tccagaatgt	ctctcctgca	caaacatgag	780
gaatctggct	gctgtgtagc	acac				804
<210> SEQ : <211> LENG <212> TYPE <213> ORGAI <220> FEATU <223> OTHEN polyn	ID NO 86 FH: 1398 : DNA VISM: Artif: JRE: R INFORMATIC ucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	ience: Synth	etic
<400> SEQUI	ENCE: 86					
atggcggagg	agcaggacct	atcggaggtg	gagetgagee	ccgtgggctc	ggaggagccc	60
cgctgcctgt	ccccggggag	cgcgccctcg	ctagggcccg	acggcggcgg	cggcggatcg	120
ggcctgcgag	ccagcccggg	gccaggcgag	ctgggcaagg	tcaagaagga	gcagcaggac	180
ggcgaggcgg	acgatgacaa	gttccccgtg	tgcatccgcg	aggccgtcag	ccaggtgctc	240
agcggctacg	actggacgct	ggtgcccatg	cccgtgcgcg	tcaacggcgc	cagcaaaagc	300
aagccgcacg	tcaagcggcc	catgaacgcc	ttcatggtgt	gggctcaggc	agcgcgcagg	360
aagctcgcgg	accagtaccc	gcacctgcac	aacgctgagc	tcagcaagac	gctgggcaag	420
ctctggaggc	tgctgaacga	aagtgacaag	cgccccttca	tcgaggaggc	tgagcggctc	480
cgtatgcagc	acaagaaaga	ccacccggac	tacaagtacc	agcccaggcg	gcggaagaac	540
gggaaggccg	cccagggcga	ggcggagtgc	cccggtgggg	aggccgagca	aggtgggacc	600
gccgccatcc	aggcccacta	caagagcgcc	cacttggacc	accggcaccc	aggagagggc	660
tcccccatgt	cagatgggaa	ccccgagcac	ccctcaggcc	agagccatgg	cccacccacc	720
cctccaacca	ccccgaagac	agagctgcag	tcgggcaagg	cagacccgaa	gcgggacggg	780
cgctccatgg	gggagggcgg	gaagceteae	atcgacttcg	gcaacgtgga	cattggtgag	840
atcagccacg	aggtaatgtc	caacatggag	acctttgatg	tggctgagtt	ggaccagtac	900
ctgccgccca	atgggcaccc	aggccatgtg	agcagctact	cagcagccgg	ctatgggctg	960
ggcagtgccc	tggccgtggc	cagtggacac	tccgcctgga	tctccaagcc	accaggcgtg	1020
gctctgccca	cggtctcacc	acctggtgtg	gatgccaaag	cccaggtgaa	gacagagacc	1080
gcggggcccc	aggggccccc	acactacacc	gaccagccat	ccacctcaca	gatcgcctac	1140
acctccctca	gcctgcccca	ctatggctca	gccttcccct	ccatctcccg	cccccagttt	1200
gactactctg	accatcagcc	ctcaggaccc	tattatggcc	actcgggcca	ggcctctggc	1260
ctctactcgg	ccttctccta	tatggggccc	tcgcagcggc	ccctctacac	ggccatctct	1320
gaccccagcc	cctcagggcc	ccagtcccac	agccccacac	actgggagca	gccagtatat	1380
acqacactot	cccqqccc	-	-		-	1398
<210> SEQ : <211> LENG <212> TYPE <213> ORGAN	ID NO 87 IH: 951 : DNA NISM: Artif:	icial Seque	nce			

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

<400> SEQUENCE: 87	
atgtacaaca tgatggagac ggagctgaag ccgccgggcc cgcagcaaac ttcgggggggc	60
ggcggcggca actecaeege ggeggeggee ggeggeaaee agaaaaaeag eeeggaeege	120
gtcaagcggc ccatgaatgc cttcatggtg tggtcccgcg ggcagcggcg caagatggcc	180
caggagaacc ccaagatgca caactcggag atcagcaagc gcctgggcgc cgagtggaaa	240
cttttgtcgg agacggagaa gcggccgttc atcgacgagg ctaagcggct gcgagcgctg	300
cacatgaagg agcacccgga ttataaatac cggccccggc ggaaaaccaa gacgctcatg	360
aagaaggata agtacacgct gcccggcggg ctgctggccc ccggcggcaa tagcatggcg	420
ageggggteg gggtggggege eggeetggge gegggegtga aceagegeat ggaeagttae	480
gegeacatga aeggetggag caaeggeage taeageatga tgeaggacea getgggetae	540
ccgcagcacc cgggcctcaa tgcgcacggc gcagcgcaga tgcagcccat gcaccgctac	600
gacgtgagcg ccctgcagta caactccatg accagctcgc agacctacat gaacggctcg	660
cccacctaca gcatgtccta ctcgcagcag ggcacccctg gcatggctct tggctccatg	720
ggttcggtgg tcaagtccga ggccagetcc ageceectg tggttaeete tteeteecae	780
tccagggcgc cctgccaggc cggggacctc cgggacatga tcagcatgta tctccccggc	840
gccgaggtgc cggaacccgc cgcccccagc agacttcaca tgtcccagca ctaccagagc	900
ggcccggtgc ccggcacggc cattaacggc acactgcccc tctcacacat g	951
<210> SEQ ID NO 88	
<pre>&lt;211&gt; LENGTH: 1338 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	ic
<pre>&lt;211&gt; LENGIN: 1336 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	60
<pre>&lt;211&gt; LENGTR: 1336 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	ic 60 120
<pre>&lt;211&gt; IINGIN: 1336 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	60 120 180
<pre>&lt;211&gt; IEMGIN: 1336 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	cic 60 120 180 240
<pre>&lt;211&gt; IIVIE: DNA &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	51C 60 120 180 240 300
<pre>&lt;211&gt; IEMGIN: 1336 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	cic 60 120 180 240 300 360
<pre>&lt;211&gt; IING TYPE: DNA &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;2223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	cic 60 120 180 240 300 360 420
<pre>&lt;211&gt; IIYPE: DNA &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	fic 60 120 180 240 300 360 420 480
<pre>&lt;211&gt; ILMIN 1336 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;2223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	cic 60 120 180 240 300 360 420 480 540
<pre>&lt;211&gt; IENERIA: ISSS &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet polynucleotide &lt;400&gt; SEQUENCE: 88 atgcgacctg ttcgagagaa ctcatcaggt gcgagaagcc cgcgggttcc tgctgatttg gcgcggagca ttttgataag cctacccttc ccgccggact cgctggccca caggccccca agctccgctc cgacggagtc ccagggcctt ttcaccgtgg ccgctccagc cccgggagcg ccttctcctc ccgccacgct ggcgcacctt cttcccgccc cggcaatgta cagccttctg gagactgaac tcaagaaccc cgtagggaca cccaacaag cggcggcgc gaactcgggc gcgcgcagca gcggtggtg gagcggagt ggcgggggta cagaccaga ccgtgtgaaa cggcccatga acgccttcat ggtatggtc cgcggcacc ggcgacaaat ggcctggag aaccccaaga tgcacaattc tgagatcagc agcgggcagc gacctggag accccaaga tgcacaattc tgagatcagc agcgcgcacc ggcgactg gaactgcgg accgacgccg gagagcgacc attcatcgac gaggccaagc gacttcgcg cgtgcacatg accgacgccg agaagcgacc attcatcgac gaggccaagc gacttcgcg cgtgcacatg</pre>	cic 60 120 180 240 300 360 420 480 540 600
<pre>2112 IENGIN: 1336 2112 TYPE: DNA 2113 ORGANISM: Artificial Sequence 2203 FEATURE: 2203 OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	cic 60 120 180 240 300 360 420 480 540 600
<pre>&lt;211&gt; IINGAN: INF 212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet polynucleotide &lt;400&gt; SEQUENCE: 88 atgcgacctg ttcgagagaa ctcatcaggt gcgagaagcc cgcgggttcc tgctgatttg gcgcggagca ttttgataag cctaccettc ccgccggact cgctggccca caggccccca agetccgete cgacggagte ccagggectt tteaccgtgg ccgetecage cccgggageg cetteteete ccgecaeget ggegeacett etteecgeee eggeaatgta cageettetg gagactgaae teaagaacee egtagggaca eccaecaag eggeggege gaaeteggge ggeggeagea geggtggtge gageggaggt ggegggggta cagaecegge gaaeteggge ggegecaega acgeetteat ggtatggtee egeggeage ggegeaaaat ggeeetggag aaceecaaga tgeacaatte tgagatcage aagegettgg gegeegaetg gaaeteggg acceaega tgeaeatte tgagatcage aagegettgg gegeegaetg gaaeteggag aaceecaaga tgeaeatte tgagatcage aagegettgg gegeegaetg gaaeteggag aacegaegeeg agaagegaee atteategae gageecaage gaettegeee egtegeaeag aaggagtate cggaetaeaa gtacegaeeg egeegeaaga ccaagaeegeegee egeegeeget</pre>	cic 60 120 180 240 360 420 480 540 600 660 720
<pre>&lt;211&gt; LENDA &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthet polynucleotide &lt;400&gt; SEQUENCE: 88 atgcgacctg ttcgagagaa ctcatcaggt gcgagaagcc cgcgggttcc tgctgatttg gcgcggagca ttttgataag cctaccctc ccgccggact cgctggccca caggccccca agctccgctc cgacggagtc ccagggcctt ttcaccgtgg ccgctccage cccgggagcg ccttctcctc ccgccacgct ggcgcacctt cttcccgccc cggcaatgta cagccttctg gagactgaac tcaagaaccc cgtagggaca cccaacaag cggcggcgc gaactcgggc ggcggcagca gcggtggtgc gagcggagt ggcggaggta cagaccagga ccgtgtgaaa cggccccaga acgccttcat ggtatggtcc cgcgggcagc ggcgcaaaat ggccctggag aaccccaaga tgcacaattc tgagatcagc agcggcagc gacctggag aaccccaaga tgcacaattc tgagatcagc agcggcagc gacctggg gagagtatc cggactacaa gtaccgacg cgcgcaaga ccaagacgt gctcaagaa aggagtatc cggactacaa gtaccgacg cgccgcaaga ccaagacgt gctcaagaaa gataagtact ccctgcccag cgcccagcagt ccggtgggg tgggcggcg cgccactg gccgcggcg cagccgtg cgcccag cgccgcag ccgcgcgc ccc aaggagtatc ccggccag cgccccag cgccgccagc cgccgccgc gccgcgccg cagccgctg cgcccagcagt ccggtgggg cgccaaga gataagtact ccctgcccag cgcccagcagt ccggtgggg tgggccagc cgccgccgc gccgcggccg cagccgtgc cgccagcagt ccggtggcg tgggccagc cctggacagc gccgcggccg cagccgctg cgccaaga ccggtgggcg ccgcgccacg gccgcggccg cagccgctg cgccaaga ccaagacgt cctggacaga accgacgccg cagccgctg cgccaaga ccggtgggg ccgcgccacg gccgcggccg cagccgctg cgccacag ccgcgcgcg ccgcgccg gccgcggccg cagccgctg cgccacag ccgcgcgcg ccgcgccg gccgcggccg cagccgctg cgccacaga ccggtgggg tgggccagcg cctggacag gccgcggccg cagccgctg cgccacaga ccggtgggg tgggccagcg cctggacac gccgcggccg cagccgctg cgccacaga ccggtggg tgggccagcg cctggacag gccgcggccg cagccgctg cgccacaga ccggtgggg tgggccagcg cctggacag gcgcgcg cagccgctg cgccacaga ccggtggg tgggccacag cctggacacg gcgcgcg cagccgccg cgccgcacag ccgcgccg ccgccgcc gccgcgccg cagccgccg cgccgccg ccgccgccg ccgccgcc gccgcgccg cagccgccg cgccgccg ccgccgcg ccgcgccg gccgcgcg cgccgccg ccgccgccg ccgccgcg ccgccg</pre>	ic 60 120 180 240 300 360 420 480 540 600 660 720 780
<pre>2112 bENGIN: 1338 2112 bENGIN: 1338 2112 STPE: DNA 2113 ORGANISM: Artificial Sequence 2200 FEATURE: 2200 STEATURE: 2213 OTHER INFORMATION: Description of Artificial Sequence: Synthet</pre>	cic 60 120 180 240 300 360 420 480 540 600 660 720 780

50nt Indoa	
cgctacgaca tggccggcct gcagtacagc ccaatgatgc cgcccggcgc tcagagctac	960
atgaacgtcg ctgccgcggc cgccgccgcc tcgggctacg ggggcatggc gccctcagcc	1020
acagcageeg eggeegeege etaegggeag cageeegeea eegeegege egeagetgeg	1080
geogeageog ceatgageet gggeeceatg ggeteggtag tgaagtetga geecageteg	1140
cegeegeeeg ceategeate geacteteag egegegtgee teggegaeet gegegaeatg	1200
atcagcatgt acctgccacc cggcgggggac gcggccgacg ccgcctctcc gctgcccggc	1260
ggtcgcctgc acggcgtgca ccagcactac cagggcgccg ggactgcagt caacggaacg	1320
gtgccgctga cccacatc	1338
<210> SEQ ID NO 89 <211> LENGTH: 813 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	Ptic
<400> SEQUENCE: 89	
atgttacagg cgtgcaaaat ggaagggttt cccctcgtcc cccctcagcc atcagaagac	60
ctggtgccct atgacacgga tctataccaa cgccaaacgc acgagtatta cccctatctc	120
agcagtgatg gggagageca tagcgaccat tactgggact tecaceecca ceaegtgeae	180
agegagtteg agagettege egagaacaae tteaeggage teeagagegt geageeeeeg	240
cagetgeage agetetaeeg ceacatggag etggageaga tgeaegteet egataeeeee	300
atggtgccac cccatcccag tcttggccac caggtctcct acctgccccg gatgtgcctc	360
cagtacccat ccctgtcccc agcccagccc agctcagatg aggaggaggg cgagcggcag	420
ageeeecac tggaggtgte tgaeggegag geggatggee tggageeegg geetgggete	480
ctgcctgggg agacaggcag caagaagaag atccgcctgt accagttcct gttggacctg	540
ctccgcagcg gcgacatgaa ggacagcatc tggtgggtgg acaaggacaa gggcaccttc	600
cagttetegt ecaageacaa ggaggegetg gegeaeeget ggggeateea gaagggeaae	660
cgcaagaaga tgacctacca gaagatggcg cgcgcgctgc gcaactacgg caagacgggc	720
gaggtcaaga aggtgaagaa gaagctcacc taccagttca gcggcgaagt gctgggccgc	780
gggggeetgg eegageggeg eeaceegeee eac	813
<210> SEQ ID NO 90 <211> LENGTH: 789 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	≆tic
<400> SEQUENCE: 90	
atgetegeee tggaggetge acagetegae gggeeacaet teagetgtet gtaeceagat	60
ggcgtettet atgaeetgga cagetgeaag eatteeaget accetgatte agaggggget	120
cetgaeteee tgtgggaetg gaetgtggee ceacetgtee eageeaeeee etatgaagee	180
ttegaeeegg cageageege ttttageeae eeecaggetg eeeagetetg etaegaaeee	240
cccacctaca gccctgcagg gaacctcgaa ctqqccccca qcctqqaqqc cccqqqqcct	300
5 5 55 5 55 55 55 55 55 55	

				-contir	nued			
ggcctccccg	cataccccac	ggagaacttc	gctagccaga	ccctggttcc	cccggcatat	360		
gccccgtacc	ccagccctgt	gctatcagag	gaggaagact	taccgttgga	cagccctgcc	420		
ctggaggtct	cggacagcga	gtcggatgag	gccctcgtgg	ctggccccga	ggggaaggga	480		
tccgaggcag	ggactcgcaa	gaagctgcgc	ctgtaccagt	tcctgctggg	gctactgacg	540		
cgcggggaca	tgcgtgagtg	cgtgtggtgg	gtggagccag	gcgccggcgt	cttccagttc	600		
tcctccaagc	acaaggaact	cctggcgcgc	cgctggggcc	agcagaaggg	gaaccgcaag	660		
cgcatgacct	accagaagct	ggcgcgcgcc	ctccgaaact	acgccaagac	cggcgagatc	720		
cgcaaggtca	agcgcaagct	cacctaccag	ttcgacagcg	cgctgctgcc	tgcagtccgc	780		
cgggccttg						789		
<210> SEQ 3 <211> LENG <212> TYPE <213> ORGA <220> FEAT <223> OTHEI polyi	ID NO 91 TH: 744 : DNA NISM: Artif: URE: R INFORMATI( nucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	ience: Synthe	tic		
<400> SEQU	ENCE: 91							
atgacgtgtg	ttgaacaaga	caagctgggt	caagcatttg	aagatgcttt	tgaggttctg	60		
aggcaacatt	caactggaga	tcttcagtac	tcgccagatt	acagaaatta	cctggcttta	120		
atcaaccatc	gtcctcatgt	caaaggaaat	tccagctgct	atggagtgtt	gcctacagag	180		
gagcetgtet	ataattggag	aacggtaatt	aacagtgctg	cggacttcta	ttttgaagga	240		
aatattcatc	aatctctgca	gaacataact	gaaaaccagc	tggtacaacc	cactcttctc	300		
cagcaaaagg	ggggaaaagg	caggaagaag	ctccgactgt	ttgaatacct	tcacgaatcc	360		
ctgtataatc	cggagatggc	atcttgtatt	cagtgggtag	ataaaaccaa	aggcatcttt	420		
cagtttgtat	caaaaaacaa	agaaaaactt	gccgagcttt	gggggaaaag	aaaaggcaac	480		
aggaagacca	tgacttacca	gaaaatggcc	agggcactca	gaaattacgg	aagaagtggg	540		
gaaattacca	aaatccggag	gaagctgact	taccagttca	gtgaggccat	tctccaaaga	600		
ctctctccat	cctatttcct	ggggaaagag	atcttctatt	cacagtgtgt	tcaacctgat	660		
caagaatatc	tcagtttaaa	taactggaat	gcaaattata	attatacata	tgccaattac	720		
catgagctaa	atcaccatga	ttgc				744		
<210> SEQ 3 <211> LENG <212> TYPE <213> ORGAJ <220> FEAT <223> OTHEJ polyn	ID NO 92 TH: 612 : DNA NISM: Artif: URE: R INFORMATI( nucleotide	icial Sequer DN: Descript	nce tion of Art:	ificial Sequ	ience: Synthe	tic		
<400> SEQU	ENCE: 92							
atgcaatcat	atgettetge	tatgttaagc	gtattcaaca	gcgatgatta	cagtccagct	60		
gtgcaagaga	atattcccgc	tctccggaga	agetetteet	tcctttgcac	tgaaagctgt	120		
aactctaagt	atcagtgtga	aacgggagaa	aacagtaaag	gcaacgtcca	ggatagagtg	180		
aagcgaccca	tgaacgcatt	catcgtgtgg	tctcgcgatc	agaggcgcaa	gatggctcta	240		
qaqaatccca	gaatgcgaaa	ctcagagatc	agcaagcaqc	tgggatacca	gtggaaaatg	300		

-continued

cttactgaag ccgaaaaatg gccattcttc caggaggcac agaaattaca ggccatgcac	360
agagagaaat acccgaatta taagtatcga cctcgtcgga aggcgaagat gctgccgaag	420
aattgcagtt tgetteeege agateeeget teggtaetet geagegaagt geaaetggae	480
aacaggttgt acagggatga ctgtacgaaa gccacacact caagaatgga gcaccagcta	540
ggecaettae egeceateaa egeagecage teacegeage aaegggaeeg etaeageeae	600
tggacaaagc tg	612
<210> SEQ ID NO 93 <211> LENGTH: 1557 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthe polynucleotide	etic
<400> SEQUENCE: 93	
atggeegaeg eagaegaggg etttggeetg gegeaeaege etetggagee tgaegeaaaa	60
gacetgeeet gegattegaa accegagage gegetegggg ecceeageaa gteeeegteg	120
teecegeagg eegeetteae eeageaggge atggagggaa teaaagtgtt tetecatgaa	180
agagaactgt ggctaaaatt ccacgaagtg ggcacggaaa tgatcataac caaggctgga	240
aggeggatgt tteeccagtta caaagtgaag gtgaegggee ttaateecaa aaegaagtae	300
attettetea tggacattgt acetgeegae gateacagat acaaattege agataataaa	360
tggtctgtga cgggcaaagc tgagcccgcc atgcctggcc gcctgtacgt gcacccagac	420
tcccccgcca ccggggcgca ttggatgagg cagctcgtct ccttccagaa actcaagctc	480
accaacaacc acctggaccc atttgggcat attattctaa attccatgca caaataccag	540
cctagattac acatcgtgaa agcggatgaa aataatggat ttggctcaaa aaatacagcg	600
ttctgcactc acgtctttcc tgagactgcg tttatagcag tgacttccta ccagaaccac	660
aagatcacgc aattaaagat tgagaataat ccctttgcca aaggatttcg gggcagtgat	720
gacatggagc tgcacagaat gtcaagaatg caaagtaaag aatatcccgt ggtccccagg	780
agcaccgtga ggcaaaaagt ggcctccaac cacagtcctt tcagcagcga gtctcgagct	840
ctctccacct catccaattt ggggtcccaa taccagtgtg agaatggtgt ttccggcccc	900
teccaggace teetgeetee acceaaceea tacceaetge eecaggagea tageeaaatt	960
taccattgta ccaagaggaa agaggaagaa tgttccacca cagaccatcc ctataagaag	1020
ccctacatgg agacatcacc cagtgaagaa gattccttct accgctctag ctatccacag	1080
cagcagggcc tgggtgcctc ctacaggaca gagtcggcac agcggcaagc ttgcatgtat	1140
gccagetetg egececeeag egageetgtg eccageetag aggaeateag etgeaaeaeg	1200
tggccaagca tgccttccta cagcagctgc accgtcacca ccgtgcagcc catggacagg	1260
ctaccctacc ageacttete egeteactte acetegggge eeetggteee teggetgget	1320
qqcatqqcca accatqqctc cccacaqctq qqaqaqqqaa tqttccaqca ccaqacctcc	1380
	1440
	1500
accellage coolgaget collacter catggogtge caaggaetet atcocctcat	1500
cagtaccact ctgtgcacgg agttggcatg gtgccagagt ggagcgacaa tagcttg	1557

```
-continued
```

<210> SEQ ID NO 94 <211> LENGTH: 1350 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide <400> SEQUENCE: 94 atgttgtgga aaataaccga taatgtcaag tacgaagagg actgcgagga tcgccacgac 60 gggagcagca atgggaatcc gcgggtcccc cacctctcct ccgccgggca gcacctctac 120 ageccegege cacceetete ceacaetgga gtegeegaat ateageegee accetaettt 180 240 ccccctccct accaqcaqct qqcctactcc caqtcqqccq acccctactc qcatctqqqq gaagegtaeg eegeegeeat caaceceetg caceageegg egeeeaeagg eageeageag 300 caggeetgge ceggeegeea gageeaggag ggagegggge tgeeetegea ceaegggege 360 420 ccggccggcc tactgcccca cctctccggg ctggaggcgg gcgcggtgag cgcccgcagg gatgeetace geogeteega cetgetgetg ceecacgeae acgeeetgga tgeogeggge 480 ctqqccqaqa acctqqqqct ccacqacatq cctcaccaqa tqqacqaqqt qcaqaatqtc 540 gacgaccage acctgttget geacgateag acagteatte geaaaggtee cattteeatg 600 660 accaaqaacc ctctqaacct cccctqtcaq aaqqaqctqq tqqqqqccqt aatqaacccc actgaggtet tetgeteagt eeetggaaga ttgtegetee teagetetae gtetaaatae 720 aaagtgacag tggctgaagt acagaggcga ctgtccccac ctgaatgctt aaatgcctcg 780 ttactgggag gtgttctcag aagagccaaa tcgaaaaatg gaggccggtc cttgcgggag 840 aagttggaca agattgggtt gaatcttccg gccgggaggc ggaaagccgc tcatgtgact 900 ctcctgacat ccttagtaga aggtgaagct gttcatttgg ctagggactt tgcctatgtc 960 tgtgaageeg aattteetag taaaceagtg geagaatatt taaceagaee teatettgga 1020 ggacgaaatg agatggcagc taggaagaac atgctattgg cggcccagca actgtgtaaa 1080 gaattcacag aacttctcag ccaagaccgg acaccccatg ggaccagcag gctcgcccca 1140 gtcttggaga cgaacataca gaactgcttg tctcatttca gcctgattac ccacgggttt 1200 ggcagccagg ccatctgtgc cgcggtgtct gccctgcaga actacatcaa agaagccctg 1260 attgtcatag acaaatccta catgaaccct ggagaccaga gtccagctga ttctaacaaa 1320 accctggaga aaatggagaa acacaggaaa 1350 <210> SEQ ID NO 95 <211> LENGTH: 28 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 95 agaccacgcc tctgtcatgt accaaatc 28 <210> SEQ ID NO 96 <211> LENGTH: 27 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
-continued		

primer
<400> SEQUENCE: 96
ggtcagcagc atcgtggtca acataac 27
<210> SEQ ID NO 97 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 97
tctccgtggt cctgaagcag acata 25
<pre>&lt;210&gt; SEQ ID NO 98 &lt;211&gt; LENGTH: 28 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic     primer </pre>
<400> SEQUENCE: 98
agecatgtgg tetetetggt tgtgtatg 28
<210> SEQ ID NO 99 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 99
tttgtgggcc tgaagaaaac t 21
<210> SEQ ID NO 100 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
<400> SEQUENCE: 100
cttgaatccc gaatggaaag gg 22
<pre>&lt;210&gt; SEQ ID NO 101 &lt;211&gt; LENGTH: 21 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic</pre>
<400> SEQUENCE: 101
tacagcatgt cctactcgca g 21
<210> SEQ ID NO 102 <211> LENGTH: 22 <212> TYPE: DNA

<sup>&</sup>lt;213> ORGANISM: Artificial Sequence

## -continued

<220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 102 gagtccattg ctgttggaac cg 22 <210> SEQ ID NO 103 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 103 21 cageggaaac eccaacagtt a <210> SEQ ID NO 104 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 104 gacctccaca gagaagtcga g 21 <210> SEQ ID NO 105 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 105 agtccactga gtaccggaga c 21 <210> SEQ ID NO 106 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 106 20 cgagagctac acgttcacgg <210> SEQ ID NO 107 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 107 21 aqccaacctt aactqaqqaq t <210> SEQ ID NO 108 <211> LENGTH: 21

									_
	$\sim$	$\sim$	n	г.	п.	n	11	$\frown$	$\sim$
-	<b>L</b>	<u> </u>		L.,	-1-	11	u.	_	~

<212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 108 tgatcctgac tgcgatgaga g 21 <210> SEQ ID NO 109 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 109 ggcaacgtgg ccttttctac 20 <210> SEQ ID NO 110 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 110 21 gaagtttcgc agacctgaca t <210> SEQ ID NO 111 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 111 atcgctctcc tgctaacagt c 21 <210> SEQ ID NO 112 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 112 ctgagacccg agcagagttt g 21 <210> SEQ ID NO 113 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 113 cacgatetea tacetggeet gette 25

	$\sim$	$\sim r$	۰÷		n	11	$\sim$	~
-	<u> </u>	л.	ւս	-1-	11	u	-	u.
	-			_		_	_	

<210> SEQ ID NO 114 <211> LENGTH: 26 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 114 tggagcagga caggttcagt ctttca 26 <210> SEQ ID NO 115 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 115 23 aggttgctgc tggtgaggtc att <210> SEQ ID NO 116 <211> LENGTH: 26 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 116 gtttgagtgg tgccgtactg gtagga 2.6 <210> SEQ ID NO 117 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 117 agggctgtcc tgaataagca g 21 <210> SEQ ID NO 118 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 118 23 gtgtatatcc cagggtgatc ctc <210> SEQ ID NO 119 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 119 qaqqaaqaqq taaccacaqq q

-continued

<210> SEQ ID NO 120 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 120 22 atgtccctct tgtcgccaac ct <210> SEQ ID NO 121 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 121 21 gagggtcagt agaacatgcg t <210> SEQ ID NO 122 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 122 21 tgcctttttc ttagggcaga g <210> SEQ ID NO 123 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 123 catttcacgc atctggcgtt c 21 <210> SEQ ID NO 124 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 124 gggtgtcgag ggaaaaatag g 21 <210> SEQ ID NO 125 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 125

-continued

ggcaagttga ttggagggat g 21 <210> SEQ ID NO 126 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 126 cttgtctgtt cttctgaccc c 21 <210> SEQ ID NO 127 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 127 agtggcagtt acccattect g 21 <210> SEQ ID NO 128 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 128 22 gtatgcacca ttcaactcct cg <210> SEQ ID NO 129 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 129 21 ctcgtactgg atgggtgaac t <210> SEQ ID NO 130 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer <400> SEQUENCE: 130 21 tgaatetega egtteteete e

1. An isolated polynucleotide or vector comprising:

(a) a polynucleotide encoding a transcription factor (TF) open reading frame (ORF);

(b) a nucleic acid barcode, and

(c) an optional vector comprising (a) and (b);

wherein the nucleic acid barcode is located 3' to the TF ORF.

**2**. The polynucleotide or vector of claim **1**, wherein the TF ORF encodes a developmentally critical TF.

**3**. A TF screening library comprising a polynucleotide or vector of claim **1**.

**4**. A TF screening library comprising a polynucleotide or vector of claim **2**.

**5**. The TF screening library of claim **3**, wherein the developmentally critical TF is a TF selected from the TFs listed in Table 1.

6. The polynucleotide or vector of claim 1 wherein at least one nucleic acid or vector further comprises a nucleic acid encoding an expression control element.

7. A viral packaging system comprising the polynucleotide or vector of claim 1 and a packaging plasmid.

**8**. A method for producing a viral particle, the method comprising transfecting a packaging cell line with the system of claim **7** under conditions suitable to package the vector or the TF screening library into a viral particle.

9. A viral particle produced by the method of claim 8, and optionally a carrier.

**10**. An isolated cell comprising the polynucleotide or vector of claim **1**, and optionally a carrier.

11. A kit comprising the polynucleotide or vector of claim 1 and optionally instructions for use.

**12**. A method of performing a high throughput gene activation screen, the method comprising:

(a) transducing a target cell with the viral particle of claim 9; and

(b) performing scRNA-seq on the transduced target cell to identify the nucleic acid barcode.

13. The method of claim 12, further comprising determining a fitness effect in the transduced target cell.

14. The method of claim 12, further comprising identifying a co-perturbation network.

**15**. The method of claim **12**, further comprising identifying a functional gene module.

**16**. The method of claim **12**, wherein the target cell is a stem cell, optionally an embryonic stem cell (ESC) or an induced pluripotent stem cell (iPSC).

17. A method of driving differentiation of a stem cell into an endothelial cell, the method comprising inducing ectopic expression of ETV2 in a stem cell under conditions suitable to support differentiation of the stem cell into an endothelial cell.

18. The method of claim 17, wherein ectopic expression of ETV2 is induced by transducing the stem cell with a vector comprising a nucleic acid encoding ETV2 and a nucleic acid encoding an expression control element, and optionally wherein the stem cell has been genetically modified.

**19**. The method of claim **17**, further comprising genetically modifying the stem cell or the endothelial cell.

**20**. An endothelial cell produced by the method of claim **19**, and optionally a carrier.

21. A method of treating a subject thereof, the method comprising administering the endothelial cell of claim 20 to the subject.

\* \* \* \* \*