



US 20210108193A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2021/0108193 A1**

Mali et al.

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

(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.**

<b>C12N 15/10</b>	(2006.01)
<b>C12N 5/071</b>	(2006.01)
<b>C12N 15/86</b>	(2006.01)
<b>A61K 35/44</b>	(2006.01)

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

#### 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.**

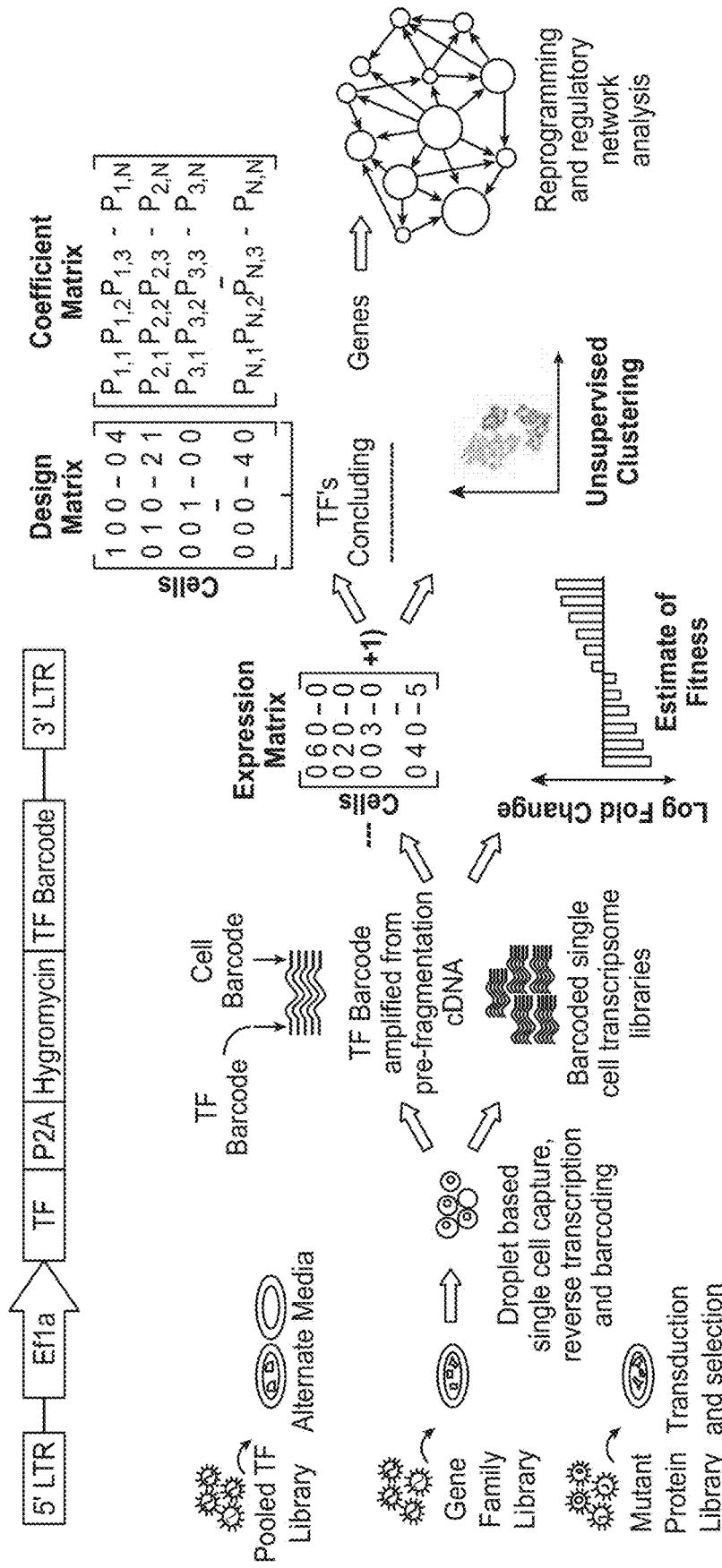


FIG. 1A

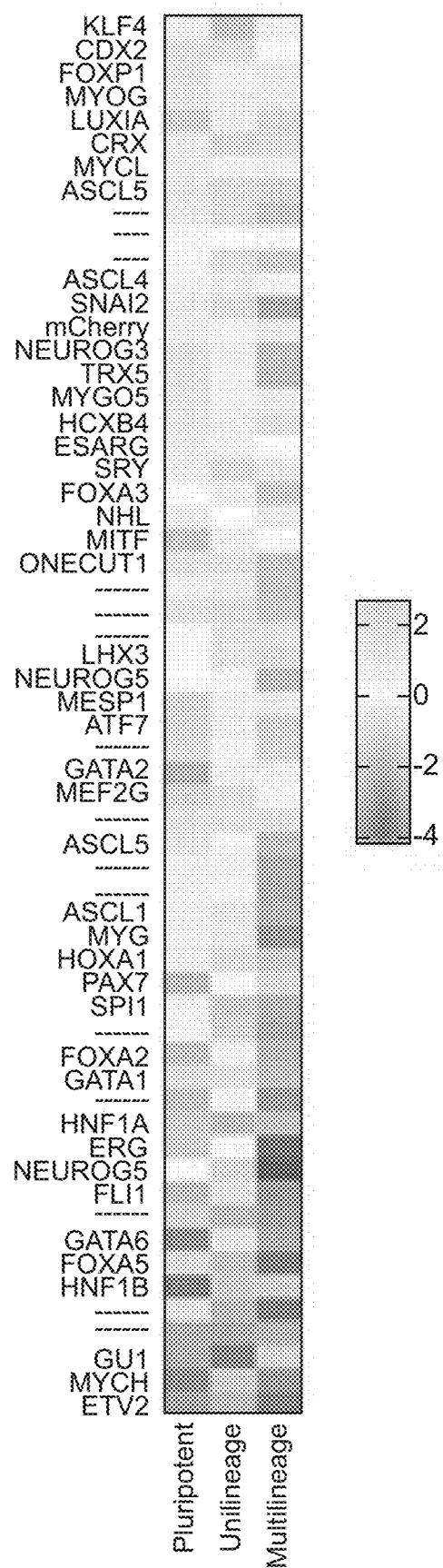


FIG. 1B

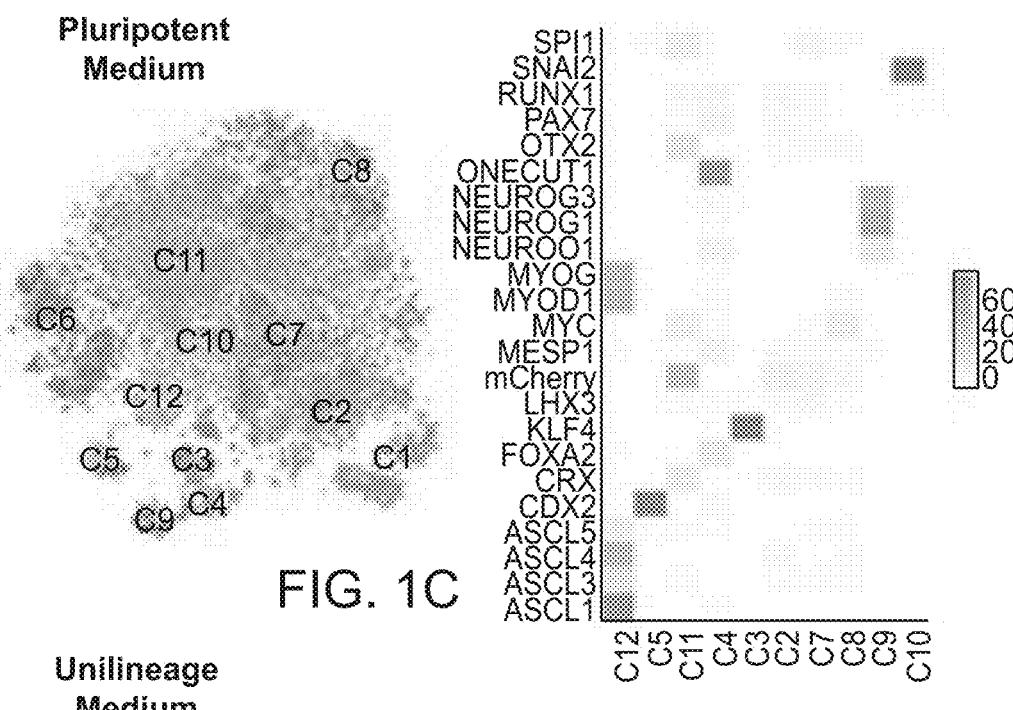


FIG. 1C

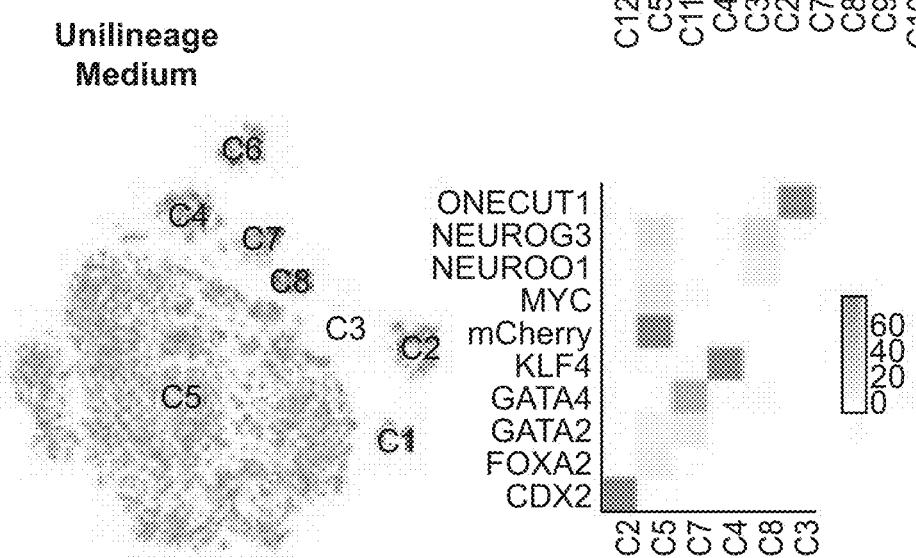


FIG. 1D

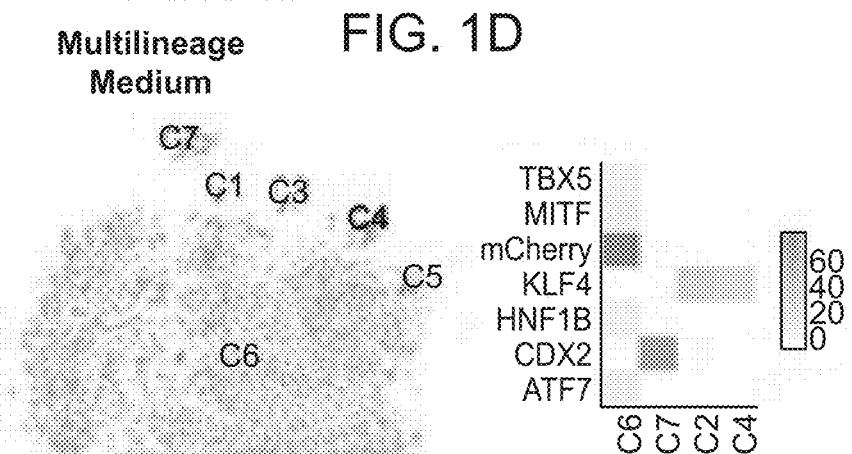


FIG. 1E

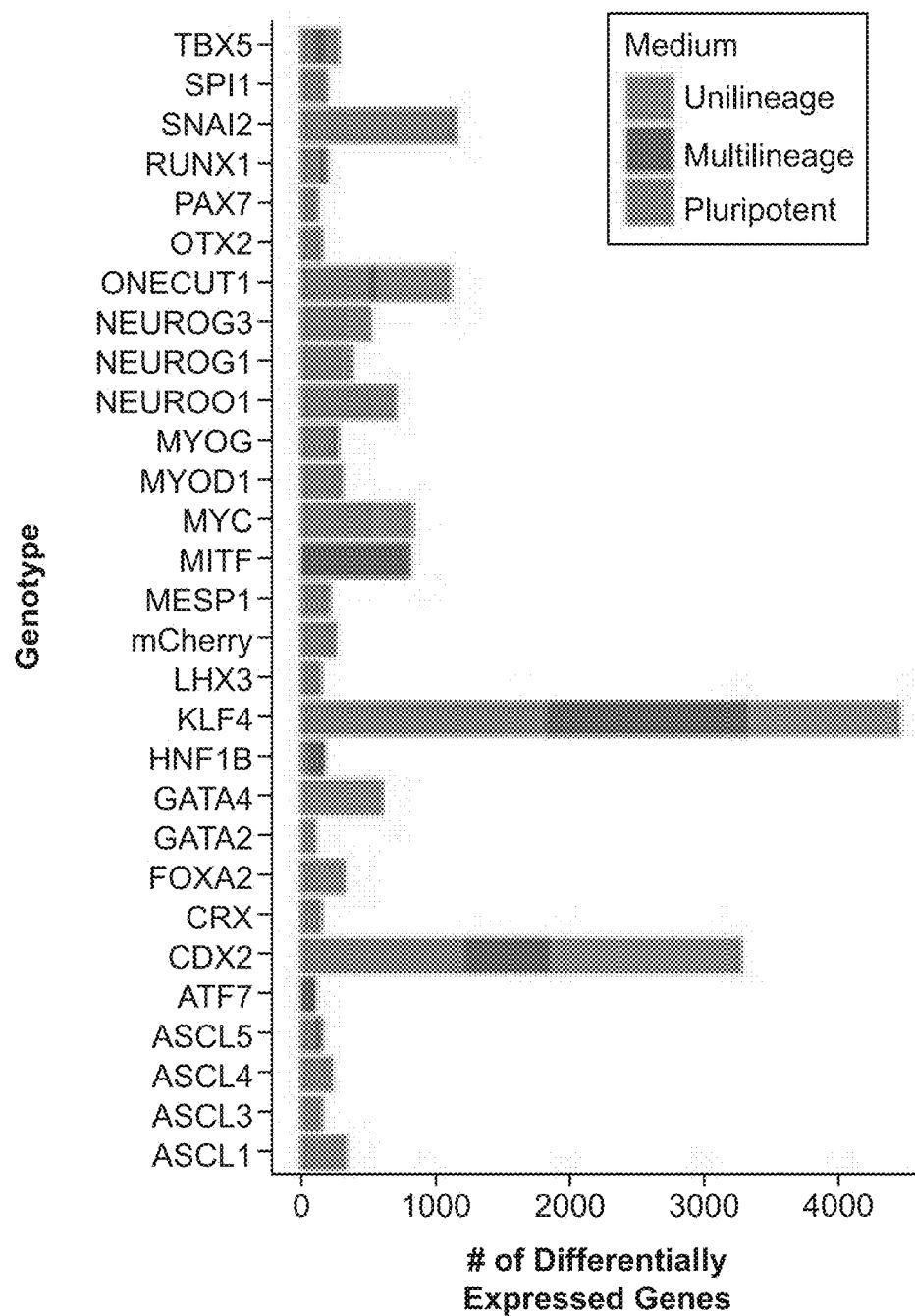
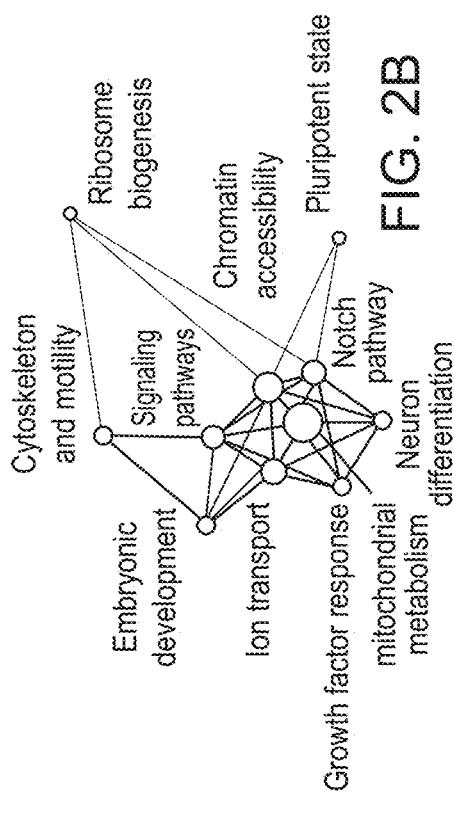
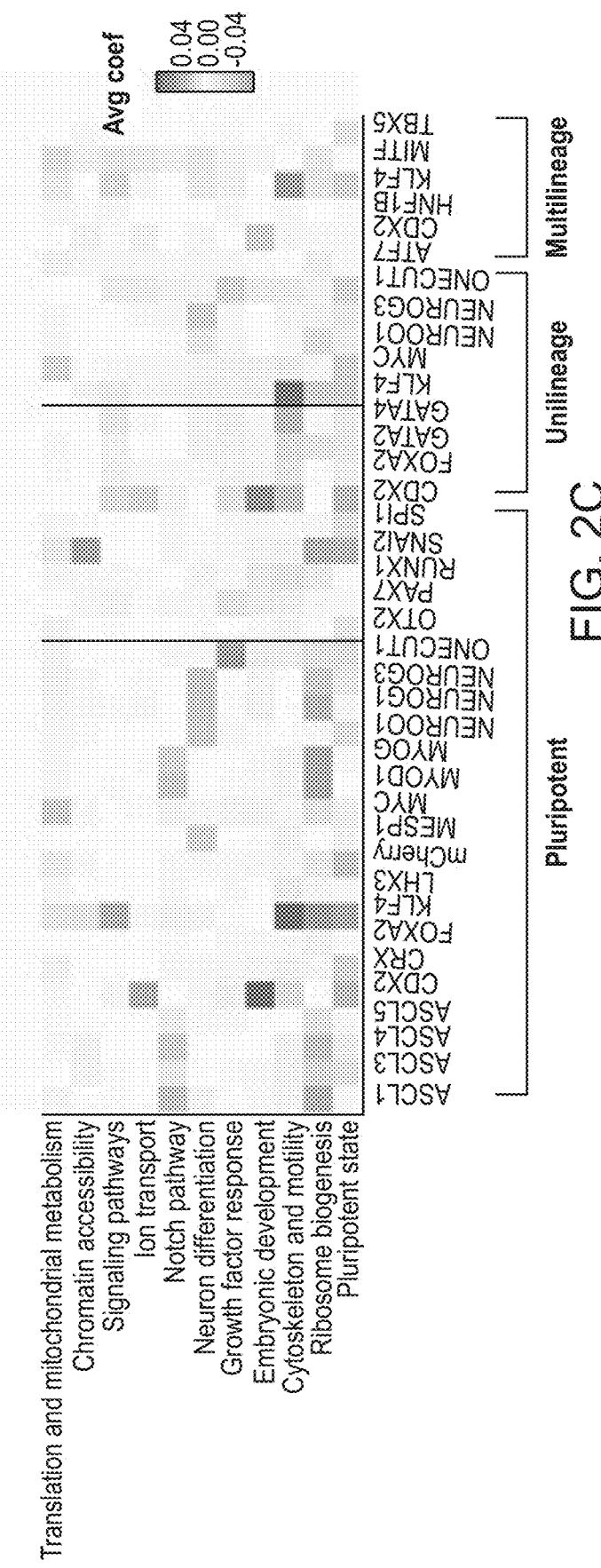
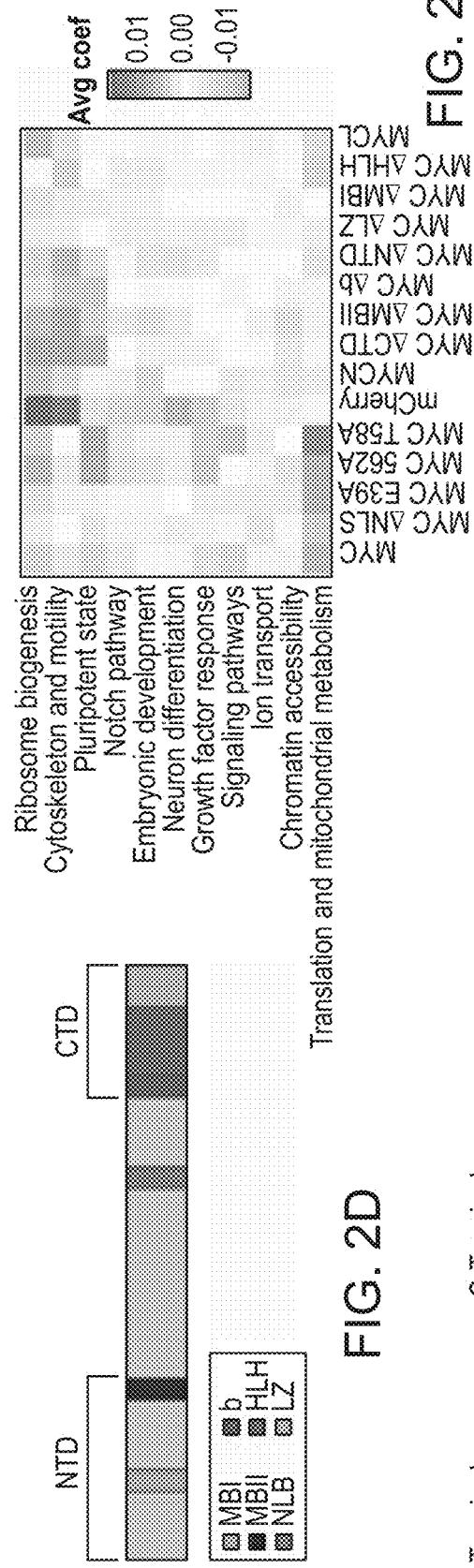
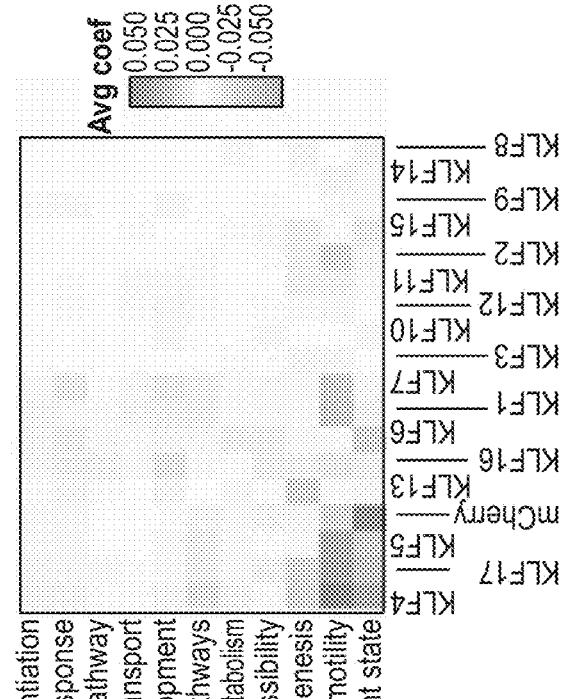
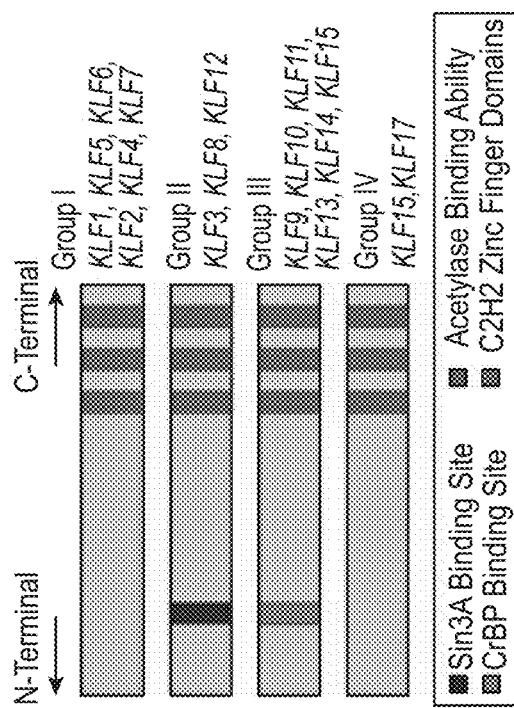
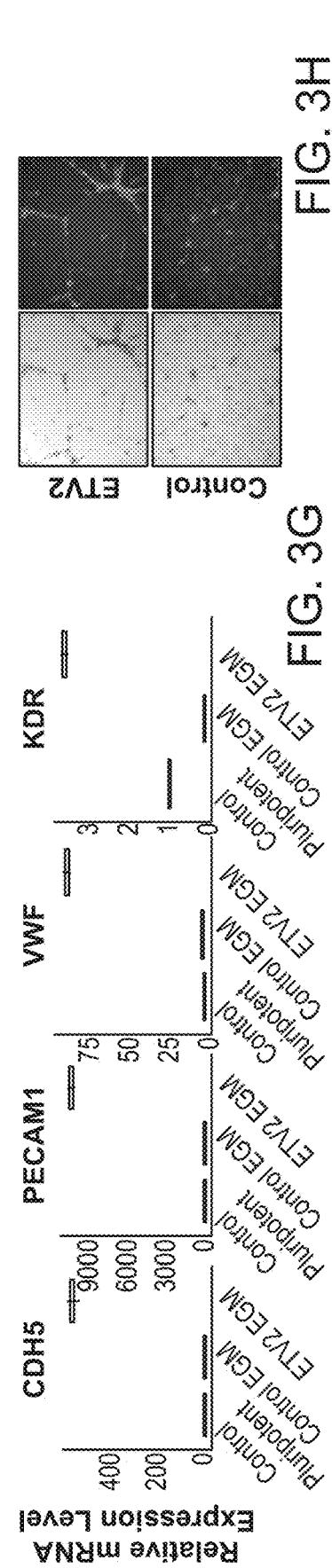
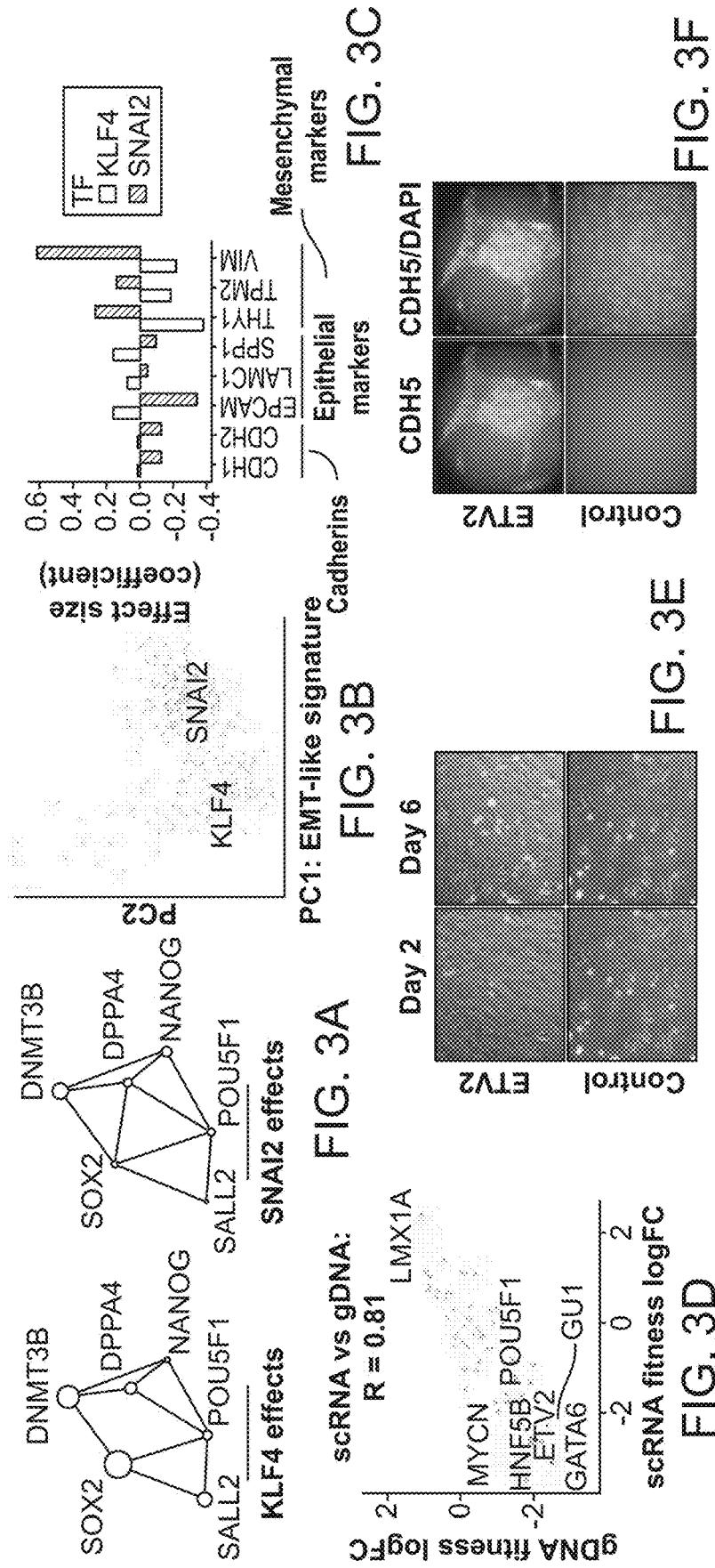


FIG. 1F

Coefficient Matrix	
TFs	Genes
$P_{1,1} P_{1,2} P_{1,3} \dots P_{1,N}$	Gene1
$P_{2,1} P_{2,2} P_{2,3} \dots P_{2,N}$	Gene2
$P_{3,1} P_{3,2} P_{3,3} \dots P_{3,N}$	Gene3
$\vdots$	Gene4
$P_{N,1} P_{N,2} P_{N,3} \dots P_{N,N}$	Gene5 Gene6 Gene7

**FIG. 2A**

**FIG. 2B**

**FIG. 2C**

**FIG. 2E****FIG. 2G****FIG. 2F**



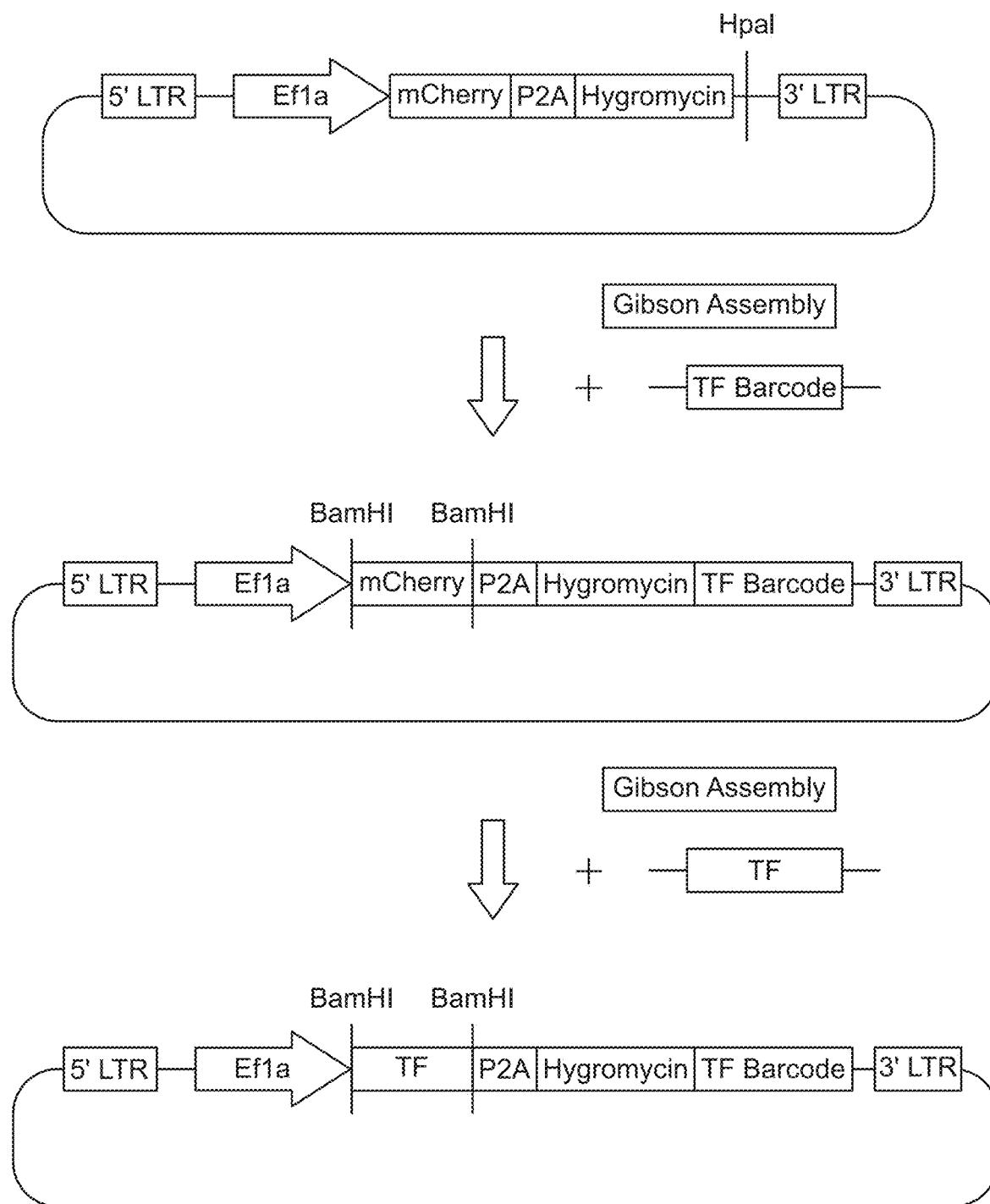


FIG. 4

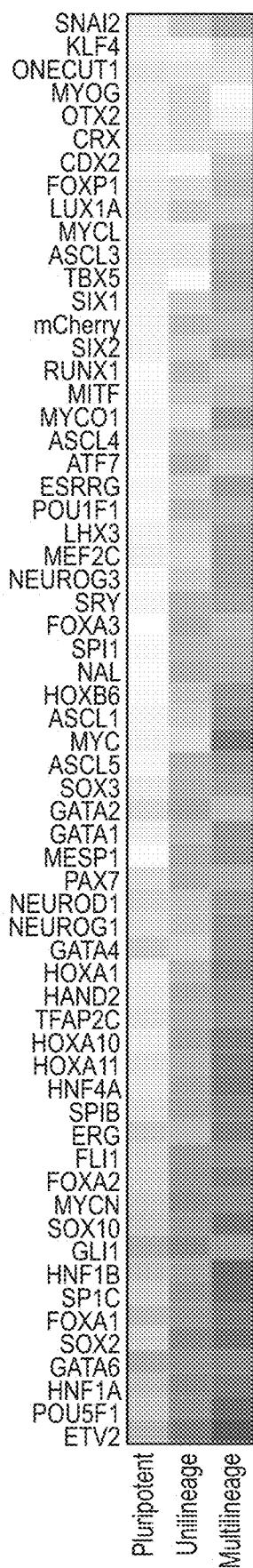


FIG. 5A

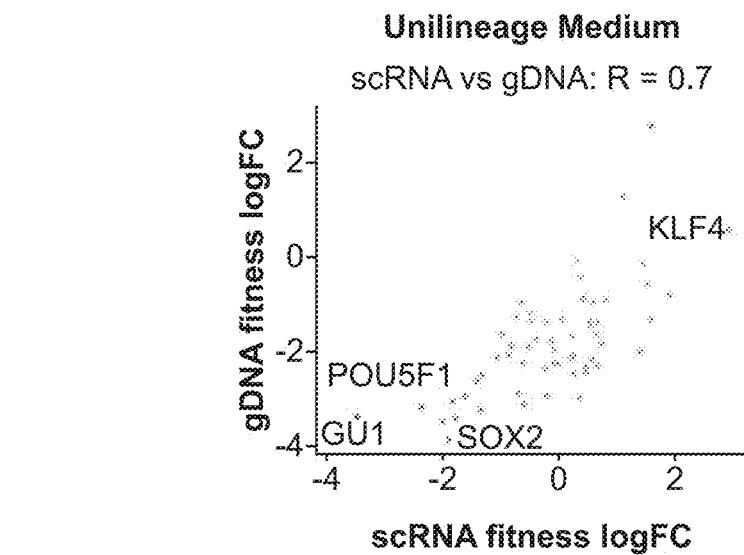


FIG. 5B

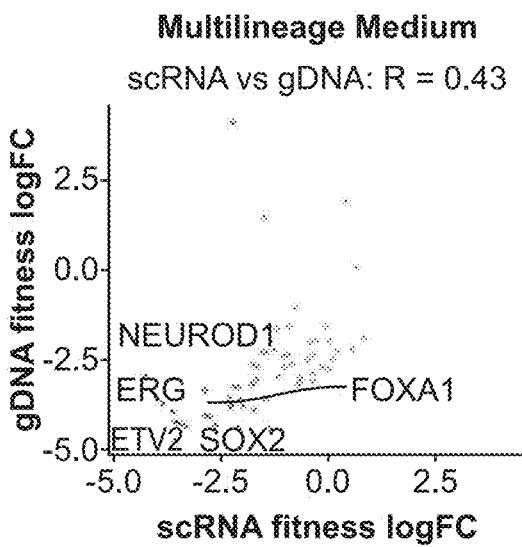
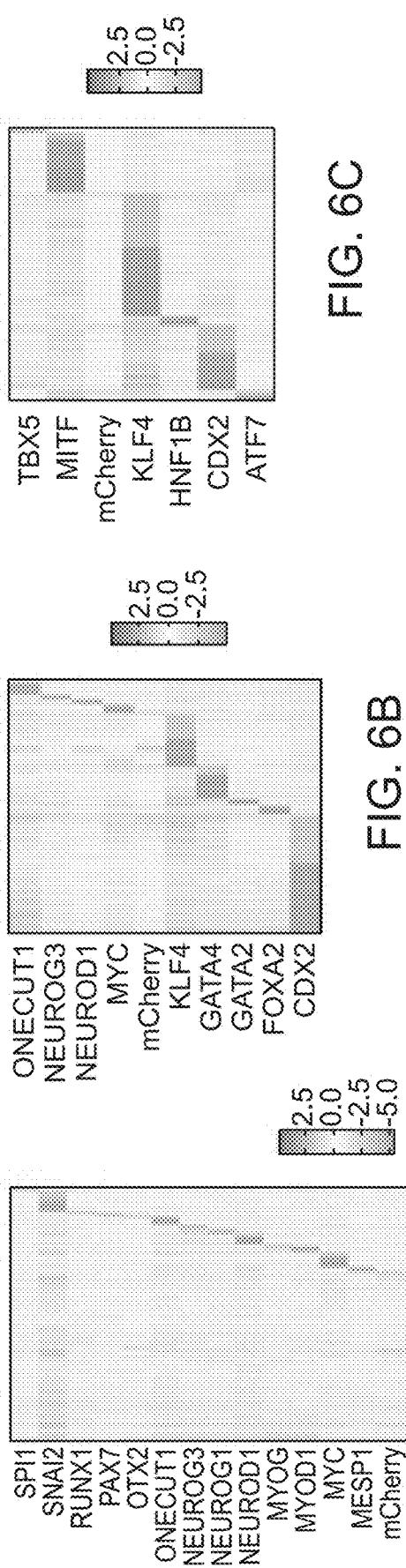


FIG. 5C



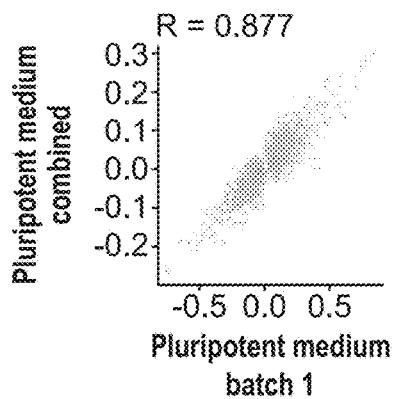


FIG. 7A

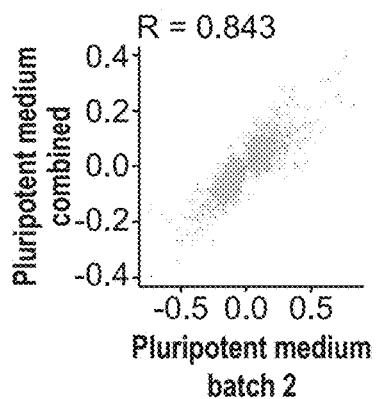


FIG. 7B

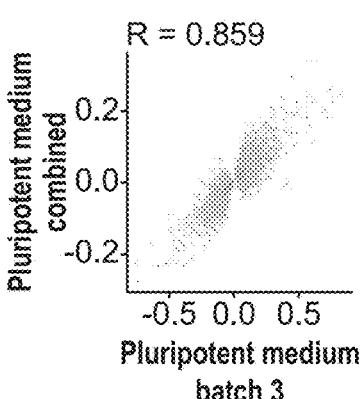


FIG. 7C

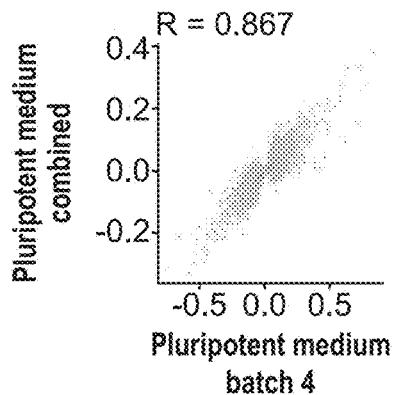


FIG. 7D

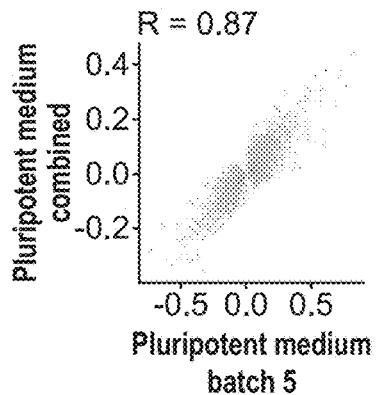


FIG. 7E

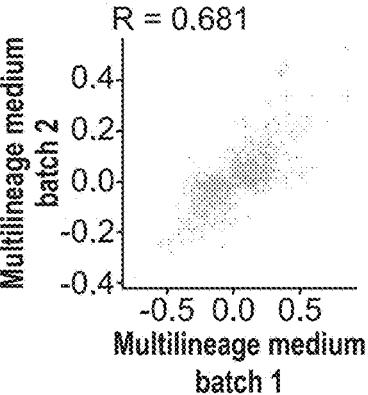


FIG. 7F

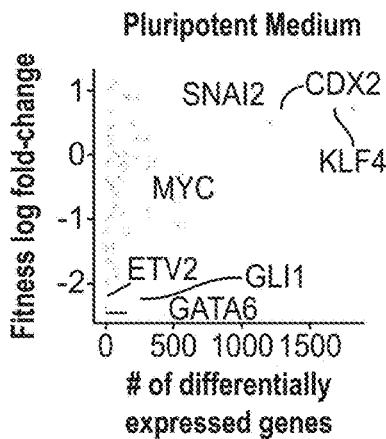


FIG. 8A

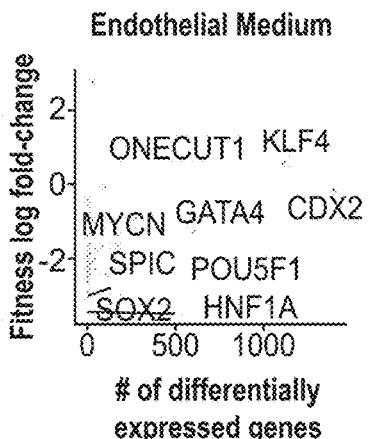


FIG. 8B

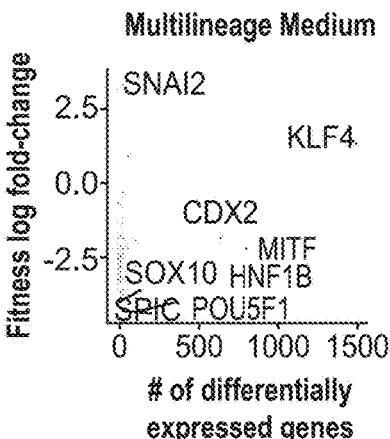


FIG. 8C

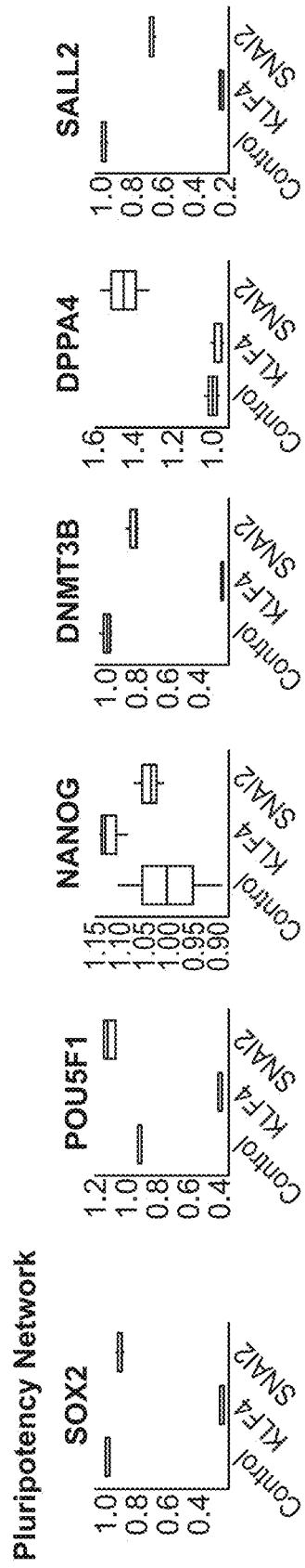


FIG. 9A

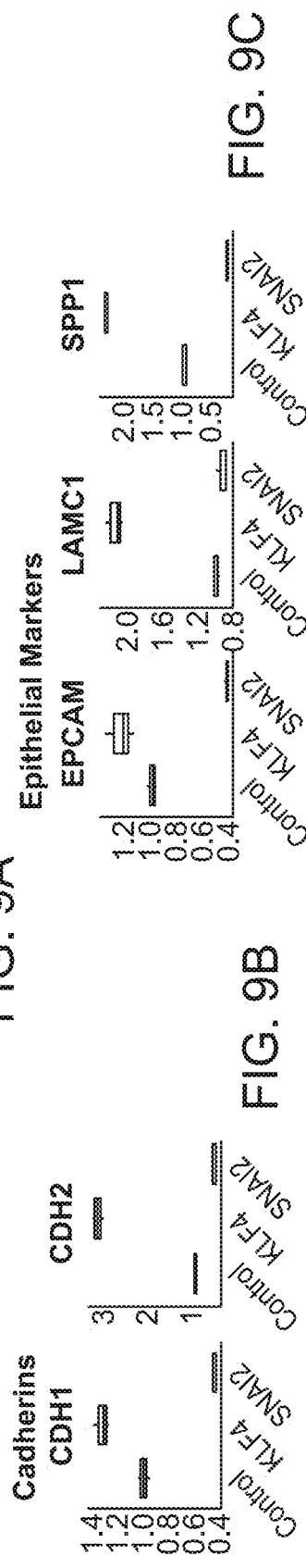


FIG. 9B

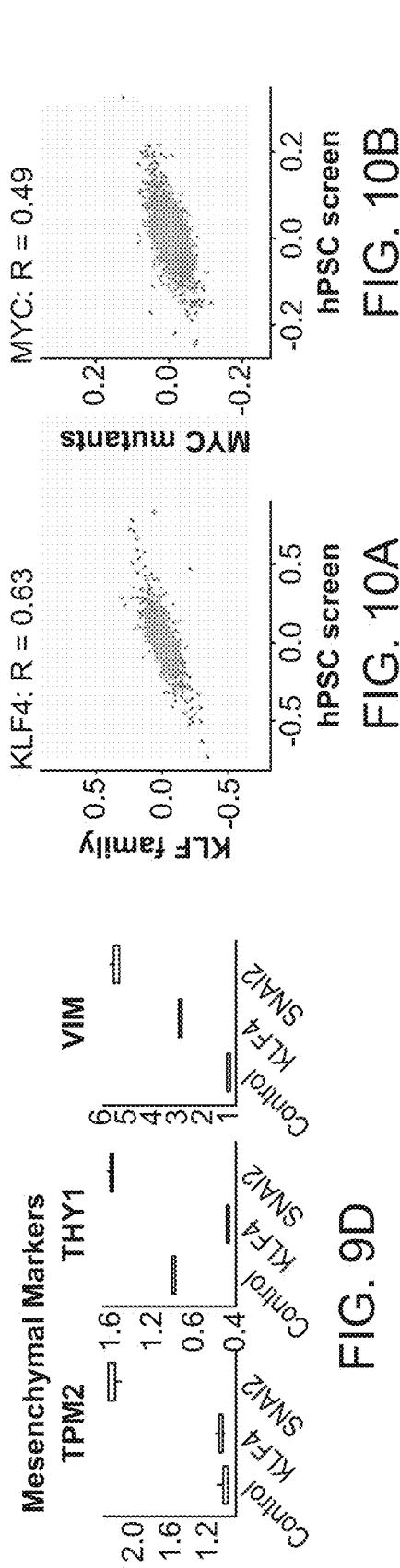


FIG. 9D

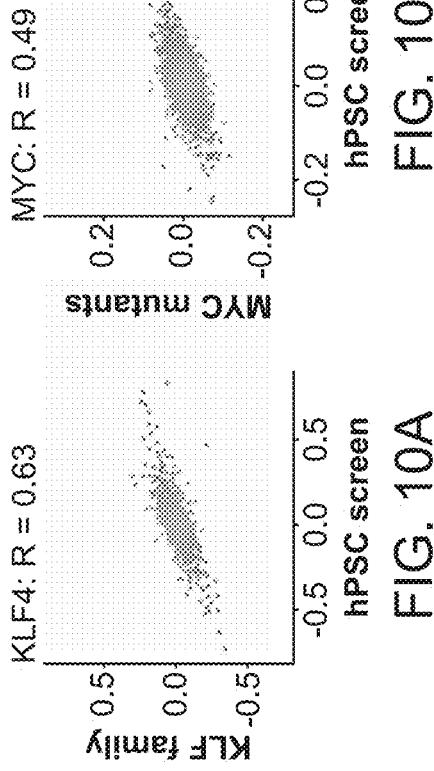


FIG. 10A



FIG. 10B

## 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 a translation elongation factor, optionally wherein the translation elongation factor is Efla.

[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 a TF 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 a TF 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 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° C. to 0° C., from -80° C. to 0° C., from -20° C. to 0° C., or from 0° C. to 10° 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 amino-terminal 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 gene-set 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 mCherry-transduced 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. 5A-5C: Fitness analysis from genomic DNA and correlation with fitness from scRNA-seq genotyped cell counts (FIG. 5A) Log fold-change of TF read counts amplified from genomic DNA vs plasmid library control (FIG.

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

[0027] FIGS. 6A-6D: Differential gene expression analysis of significant TFs (FIG. 6A) Heatmap of differentially expressed genes for significant TFs in hPSC medium. (FIG. 6B) Heatmap of differentially expressed genes for significant TFs in endothelial growth medium. (FIG. 6C) Heatmap of differentially expressed genes for significant TFs in multilineage medium (FIG. 6D) 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, |coef|>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. 8A-8C: Correlation between fitness and transcriptomic effects. (FIG. 8A) Correlation of the number of differentially expressed genes for each TF vs the fitness effect (log-FC) for hPSC medium (FIG. 8B) Correlation of the number of differentially expressed genes for each TF vs the fitness effect (log-FC) for endothelial medium (FIG. 8C) 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 post-transduction 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. 10A-10B: Correlation of KLF4 and MYC effects across samples. (FIG. 10A) Correlation of KLF4 effects in the KLF family screen with KLF4 effects in the hPSC screen. (FIG. 10B) 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, 3<sup>rd</sup> 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, 5<sup>th</sup> 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, 3<sup>rd</sup> edition (Cold Spring Harbor Laboratory Press (2002)); Sohal (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: Springer-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 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 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, dextrose, 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 HNF1 $\beta$ , 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 mark-

ers 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 insulin-dependent 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 bar-coded 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, MESPI, MITF, MYC, MYCL, MYCN, MYOD1, MYOG, NEUROD1, NEUROG1, 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. Encapsulation 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 translation elongation factor, optionally wherein the translation elongation factor is Efla.

[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 a TF 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 a TF 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 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:

(SEQ ID NO: 1)

```

1   ttccctgtgc agataagccc agcttagccc agctgacccc agaccctctc ccctcactcc
61  ccccatgtcg caggatcgag accctgagggc agacagcccc ttcaccaagc ccccccgcggc
121 gccccatca ccccgtaaac ttctcccgcc ctccggccctg ccctcacca gcccgtgttt
181 ccccaagcct cgctccaagc ccacgcccacc cctgcagecag ggcagcccc gaggccagca
241 cctatccccg aggctggggt cgaggctcgcc ccccgccccct gcctctgcaa cttgagcctg
301 gctgcgaccc ctgctctgac gtctcgaaaa attccccctt gcccaggccc ttgggggagg
361 gggtgcatgg tatgaaatgg ggctgagacc cccggctggg ggcagaggaa cccggccagag
421 aaggagccaa attaggcttc tgttccctg atctggcact ccaagggac acggccgacag
481 cgacagcaga gacatgctgg aaaggtacaa gctcatccct ggcaagcttc ccacagctgg
541 actggggctc cgcgttaactg caccagaag ttccatgggg ggccggagccc gactctcagg
601 ctcttccctg gtccggggac tggacagaca tggcgtgcac agcctgggac tcttggagcg
661 ggcgcctcgca gaccctgggc cccgeccctc tggcccccggg ccccatcccc gccgcggct
721 cccgaaggcgc cgcggggccag aactgcgtcc cctgtgggggg agaggccacc tcgtggtcgc
781 ggcgcggcggc cgcggggagc aacaccagct gggactgttc tggggggccc gacggcgata
841 cctactgggg cagtggccctg ggccggggagc cgcgcacggg ctgttaccatt tcgtggggcg
901 ggcccgccggg cccggactgt accacccctt ggaacccggg gctgcattgcg ggtggcacca
961 cctcttggaa ggggttccag agctcagctc tcaccgtttt ctccgaaccg agcccgcaagt
1021 cggaccgtgc cagtttggct cgtgcggccca aaactaaccg cctggggccccc attcagctgt
1081 ggcagttccct cttggggactg ctccacgcg gggccgttag cagctgcattc cttggactg
1141 gcaacagccg cggatccag ctgtgcgacc ccaaaggggt ggctggctg tggggccggc

```

- continued

```

1201 gcaagagaaa gccgggcatg aattacgaga agctgagccg gggccttcgc tactactata
1261 gccgcacat cgtgcgcaag agcggggggc gaaagtacac gtaccgcttc gggggccgcg
1321 tgcccagcct agcctatccg gactgtgcgg gaggcggacg gggagcagag acacaataaa
1381 aattcccggt caaacctcaa aaaaaaaaaa aaa

```

**[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° C. to 10° 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

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 over-express 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 multi-tissue human cDNA pool or directly synthesized as double-stranded DNA fragments, and individually cloned into the backbone vector (FIG. 4). The final library consisted of 61 developmentally critical or pioneer TFs (Table 1). Appli-

cants 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 ( $FDR < 10^{-6}$ ), 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. 2C). 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. 2C). 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 AMBI	ATGCCCTCAACGTTAGCTTACCAACAGGAACATATGACC TCGACTACGACTCGGTGCAGCGTATTCTACTGCGACGA GGAGGAGAACTTCTTACCAAGCAGCAGCAGCAGAGCAGCT GCAGCCCCGGGGGGATCAGGTAGCGGTAGCGCCGCGCTC CGGGCTCTGCTGCCCTCTAACGGTGGGTCAACACCTCTC CCCTTCGGGGAGACAAACGACGGGGTGGCGGAGCTTCT CCACGGGGGACCCAGCTGGAGATGGTGACCGAGCTGCTGG GAGGAGACATGGTGAACCGAGATTTCATCTGGCACCCGG ACGACGAGACACTTCATCAAAAACATCATCATCCAGGACTG TATGTGGAGCGGTTCTGGCCGCCGCAAGCTCGTCTCA GAGAAGCTGGCTCTTACCAAGGCTCGCGCAAAGACAGC GGCAGCCCGAACCCCGCCGCCACAGCGCTGCTCCA CCTCCAGCTTGTACCTGCAGGATCTGAGCGCCGCCCTC AAGATGCATCGACCCCTGGTGGCTTCTCCCTACCCCTCTC AACGACAGCAGCTGCCAAGTCTCTGCCCTCGAAGACT CCAGCGCTTCTCCGTCCTCGGATTCTCTGCTCTCCCTCG ACGGAGCTCCCGCAGGGCAGGCCAGGGCCAGCTGGTGC TCCATGAGGAGACACCGCCCCACCCAGCAGGACTCTG AGGAGGACAAGAAGATGAGGAGAAATCGATGTTGTT CTGTGGAAAAGAGGCAGGCTCTGGAAAAGGTCAGAGT CTGGATCACCTCTGAGGCCACAGCAAACCTCTCA CAGCCCACACTGGCTCTCAAGAGGTGCCACGTCTCCACACAT CAGCACAACTACCCAGGGCTCCCTCACTCCGAAGGACT ATCCTGCTGCCAAGAGGGTCAAGTTGGACAGTGTCAAGAGT CTCTGAGACAGATCAGCAACAAACCGAAAAATGACCCAGGCC CAGGTCTTCGGACACCGAGGGAAATGTCAGAGGCAAC ACACAAAGCTTGGAGGCCAGGGAGAACTGCAAGAGGCC ACGGAGCTTTTGCCCTGCGTGACCGAGATCCGGAGTTG GAAAACATGAAAGGCCCAAGGTAGTTATCTTAAA AAAGCCACAGCATACATCTGTCGCGTCAAGCAGAGGAG CAAAGCTCATTTCTGAAGAGGACTTGTGCGGAAACGAC GAGAACAGTTGAAACACAAACTGAAACAGCTACGGAACT CTTGTGCG	2	Deletion of MYC Box I
c-MYC AMBI I	ATGCCCTCAACGTTAGCTTACCAACAGGAACATATGACC TCGACTACGACTCGGTGCAGCGTATTCTACTGCGACGA GGAGGAGAACTTCTTACCAAGCAGCAGCAGCAGAGCAGCT GCAGCCCCGGGGCCAGCGAGGATATCTGGAGAAATT CGAGCTGCTGCCCTGGCCCGCCCTGTCGGCTAGCCGCCG	3	Deletion of MYC Box II

- continued

GENE	SEQUENCE	SEQ ID NO:	MUTATION
	TCGGGCTCTGCCCTCCTACGTTGGTACACCCCTT CTCCCTCGGGAGACAACGACGGCGGTGGCGGGAGCTT CTCCACGGCGACCAGCTGGAGATGGTACCGAGCTGCTG GGAGGAGACATGGTAAACAGAGTTCATCTGCAGACCCG CACGACGAGACCTTCAAAACATCGATCAGTAGC GGTCGCTCTAGAGAAGCTGGCCTCTACCAAGGCTGC GCAAAGACAGCGCCAGCCGAACCCGCCGGCCACA GCGTCTGCTCACCTCCAGTTGACCTGCAAGGATCTGAG CGCGGCCCTCAGAGTCAGACAGCAGCTGCCAACG CCTACCCCTCAACAGACAGCAGCTGCCAACG CCTCGAAGACTCCAGGCCCTCTCCGTCCTGGATTCT CTGCTCTCTCGACGGAGTCTCCCGCAGGGCAGCCCC AGCCCTGGTCTCCATGAGGAGACACGCCAACACAG CAGCAGCTCTAGGGAGAACAGAAGATGAGGAAGAAA CGATGTTGTTCTGTGGAAAAGAGGCAAGGCTCTGGCAA AGGTCAAGAGCTGGATCACCTCTGCTGGAGGCCACAGCA AACCTCTCACAGCCACTGTTCTCAAGAGGTGCCACGT CTCCACACATCAGCACAACATACGCAGCGCTCCCTCACT CGGAAGGACTATCCTGCTGCCAGAGGGTAAGTTGGAC AGTGTAGAGTCTGAGACAGATCAGCAACACCGAAA TGCACAGCCCCAGGTCTCGACACCGAGGGAAATGTC AAGAGGCAACACACAACAGTCTGGAGGCCAGAGGAG AACGAGCTAAACGGAGCTTTTGCCCTGGTGAACCAGA TCCCGGAGTTGAAAATGAAAAGGCCCAAGGTAG TTATCCTAAAAAGCCACAGCATACATCTGTCCGTC AGCAGAGGAGCAAAGCTCATTTCTGAAGAGGACTTGT GGGAAACGAGAACAGTTGAAACACAAACTTGAACA GCTACGGAACCTTGTGCG		
MYC ΔNLs	ATGCCCTCAACGTTAGCTTACCAACAGGAACATGACC TCGACTACGACTCGGTGAGCGGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCGAGCAGCAGCGAGCGAGCT GAGCCCCCGGCCAGCGAGGATACTGGAGAAATT CGAGCTGCTGCCACCCGCCCTGTCCCCTAGGCCCGC TCGGGCTCTGCTGCCCTCTACGTTGGTACACCCCTT CTCCCTCGGGAGACAACGACGGCGGTGGCGGGAGCTT CTCCACGGCCACCAGCTGGAGATGGTACCGAGCTGCTG GGAGGAGACATGGTAAACAGAGTTCATCTGGACCCCG GACGACGAGACCTTCAAAACATCATCTCCAGGACT GTATGAGCTGGCTCTCGGCCAGGCCAACGCTCTGCTC AGAGAAGCTGGCTTCTACCGGCTGCGCGAAAGACAG CGGCACCCGAACCCGCCCGGCCACAGGCTCTGCTCC ACCTCAGCTTGTACCTGCAGGATCTGAGGCCGCGCCT CAGAGTCATGACCCCTCGGTGGCTTCCCCTACCC AACGACAGCAGCTGCCAACGCTTGTGCCCTCGCAAGACT CCAGGCCCTCTCGTCTCGGATTCTGTCTCCCTG ACGGAGTCTCCCGCAGGGCAGCCCCGAGCCCTGGTGC TCCATGAGGAGACACGCCAACCCAGCAGCGACTCTG AGGAGGAACAGAAGATGAGGAAGAAAATCGATGTTTT CTGTGAAAAGAGGCAAGGCTCTGGAAAAGTCAGAGT CTGGATCACCTCTGCTGGAGGCCACAGCAAACCTCTCA CAGCCACTGTTCTCAAGAGGTGCCCCACGTTCCACACAT CAGCACAACATCGAGCGCTCCCTCACTCGGAAGGACT ATGGATCAGGTAGCGTAGTGTCAAGATCTGAGACAGA TCAGCAACAACGGAAATGCAACAGCCCCAGGTCTCGG ACACCGAGGAGATGCAAGAGGGAAACACAAACGCT TGGAGCGCCAGGGAGAACAGGCTAAAACGGAGCTTT TTGCCCTGCTGACAGATCCGGAGTTGAAAACAATGA AAAGGCCCAAGGTAGTTCTTAAAGGCCACAGC ATACATCTGTCCGTCCAAGCAGGAGAACAGTCATT TCTGAAGAGGACTTGTGCGGAAACGACGAGAACAGTTG AAACACAAACTTGAACAGCTACGGAAACTTGTGCG	4	Deletion of nuclear localization signal sequence
MYC Δb	ATGCCCTCAACGTTAGCTTACCAACAGGAACATGACC TCGACTACGACTCGGTGAGCGGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCGAGCAGCAGCGAGCGAGCT GAGCCCCCGGCCAGCGAGGATACTGGAGAAATT CGAGCTGCTGCCACCCGCCCTGTCCCCTAGGCCCGC TCGGGCTCTGCTGCCCTCTACGTTGGTACACCCCTT CTCCCTCGGGAGACAACGACGGCGGTGGCGGGAGCTT CTCCACGGCCACCAGCTGGAGATGGTACCGAGCTGCTG GGAGGAGACATGGTAAACAGAGTTCATCTGCCACCCG GACGACGAGACCTTCAAAACATCATCTCCAGGACT	5	Deletion of basic motif

- continued

GENE	SEQUENCE	SEQ ID NO:	MUTATION
	GTATGTGGAGCGGCCCTCGGCCGCGCCAAGCTCGTCTC AGAGAACGCTGGCCTCTACCAGGCTCGCGCGAAAGACAG CGGCAGCCGAACCCGCCCGCCGACAGCGCTGCTCC ACCTCCAGCTTGATCTGCAGGATCTGAGCGCCGCC CAGAGTGCATCGACCCCTCGGTGGCTTCCCTACCCCTC AACGAACAGCAGTCGCCAAGTCCCTGCCCTCGCAAGACT CCAGCGCCTCTCCCTCGGTGGATTCTCTGCTCTCTCG ACCGAGTCCTCCCCGAGGGCACGCCGAGCCCCTGGTGC TCATGAGGAGACACGGCCACCCAGCAGCGACTCTG AGGAGGAACAAGAAGATGAGGAAGAAATCGATGTTGTT CTGTGGAAAAGAGGAGGGCTCGCAAAAGTCAGAGT CTGGATCACCTCTGCTGGAGGCCACAGCAAACCTCTCA CAGCCCACGTGCTCAAGAGGTGCCACGTCCACACAT CAGCACAACTACGCAGCGCCTCCACTCGGAAGGACT ATCCCTGCTGCCAGAGGGCTAAGTTGAGACTGTCAGAGT CCTGAGACAGATCAGCAACACCGAAAATGCAACAGCCC CAGGTCTCGACACCGAGGAAATGTCGGATCAGGTAG CGGTGAGCTAAACGGAGCTTTTGCCCTGCGTGACCAG ATCCCCGAGTTGAAAAACAATGAAAAGGCCCAAGGTA GTTATCCTTAAAGGCAACAGCATACATCTGTCCGTCC AAGCAGAGGAGAAAAGCTATTCTGAAGAGGACTTGT TGGGAAACGACGAGAACAGTTGAAACACAAACTTGAAC AGCTACGGAACCTTGTGCG		
MYC ΔHLH	ATGCCCTCAACGTTAGCTTACCAACAGGAACATGACC TCGACTACGACTCGGTGAGCGCTATTCTACTGCGACGA GGAGGAGAACTCTACCGAGCAGCAGCAGAGCGAGCT GCAGCCCCGGCCAGCGAGGATATCTGGAGAAATT CGAGCTGCTGCCACCCGCCCTGTCCCTAGGCCCGC TCGGGCTCTGCTGCCCTCTACGGTGGTACACCCCT CTCCCTTCGGGAGAACCGACGGCGGTGGGGAGCTT CTTCACGGCCGACAGCTGGAGATGGTGACCGAGCTGCTG GGAGGAGACATGGTGAACCGAGGTTTACATGCCACCG GACGACGAGACCTTACAAACATCATCCAGGACT GTATGTGGAGCGCCTCTCGGCCGCGCCAAGCTCGTCTC AGAGAACGCTGGCCTTACCCAGGCTGCGCGAAAGACAG CGGCAGCCGAACCCGCCCGCCACAGCGCTGCTCC ACCTCAGCTTGATCTGCAGGATCTGAGCGCCGCC CAGAGTGCATGCCCTCGGTGGCTTCCCTACCCCTC AACGAACAGCAGCTGCCAAGTCCCTGCGCAAGACT CCAGCCTCTCCCTCGGATTCTCTGCTCTCC ACGGAGTCCTCCCGCAGGGCAGCCCGAGCCCTGGTGC TCATGAGGAGACACGCCAACCCAGCAGCGACTCTG AGGAGGAACAAGAAGATGAGGAAGAAATCGATGTTGTT CTGTGGAAAAGAGGAGGGCTTGGAAAAGTCAGAGT CTGGATCACCTCTGCTGGAGGCCACAGCAAACCTCTCA CAGCCCACGTGCTCAAGAGGTGCCACGTCCACACAT CAGCACAACTACGCAGGCCCTCCACTCGGAAGGACT ATCCCTGCTGCCAGGGCTAAGTTGAGACTGTCAGAGT CCTGAGACAGATCAGCAACACCGAAAATGCAACAGCCC CAGGTCTCGACACCGAGGAAATGTCAGAGGACT ACACAACGCTTGGAGGCCAGAGGAGGAACGATCAGG TAGCGGTCAAAGCTATTCTGAAGAGGACTTGTGCGG AAACGACGAGAACAGTTGAAACACAAACTTGAACAGCTA CGGAACCTTGTGCG	6	Deletion of helix-loop-helix motif
MYC ΔLZ	ATGCCCTCAACGTTAGCTTACCAACAGGAACATGACC TCGACTACGACTCGGTGAGCGCTATTCTACTGCGACGA GGAGGAGAACTCTACCGAGCAGCAGCAGAGCGAGCT GCAGCCCCGGCCAGCGAGGATATCTGGAGAAATT CGAGCTGCTGCCACCCGCCCTGTCCCTAGGCCCGC TCGGGCTCTGCTGCCCTCTACGGTGGTACACCCCT CTCCCTTCGGGAGAACGACGGCGGTGGGGAGCTT CTTCACGGCCGACCGAGCTGGAGATGGTGACCGAGCTG GGAGGAGACATGGTGAACCGAGGTTTACATGCCACCG GACGACGAGACCTTACAAACATCATCCAGGACT GTATGTGGAGCGCCTCTCGGCCGCCAAGCTCGTCTC AGAGAACGCTGCCCTTACCCAGGCTGCGCGAAAGACAG CGGCAGCCGAACCCGCCGCCAGCGCTGCTCC ACCTCAGCTTGATCTGCAGGATCTGAGCGCCGCC CAGAGTGCATGCCCTCGGTGGCTTCCCTACCCCTC AACGAACAGCAGCTGCCAAGTCCCTGCGCAAGACT CCAGCCTCTCTCGGATTCTCTGCTCTCC	7	Deletion of leucine zipper motif

- continued

GENE	SEQUENCE	SEQ ID NO:	MUTATION
	ACGGAGTCCTCCCCGAGGGCAGCCCCGAGCCCCCTGGTGC TCCATGAGGAGACACCGCCCACACCAGCAGCAGACTCTG AGGAGGAACAAAGAAGATGAGGAAGAAATCGATGTTGTTT CTGTGAAAAGGGCAGGCTCTGGCAAAAGGTCAAGAGT CTGGATCACCTTCTGCTGGAGGCCACAGCAAACCTCCTCA CAGCCCACTGGTCTCAAGAGGTGCCACGTCTCCACACAT CAGCACAACACTACGGCAGCGCTCCACTCGGAAGGACT ATCCTGCTGCCAACAGGGCTAAGTTGACAGTGTCAAGAGT CCTGAGACAGATCAGAACACCGAAAATGCACCGAGCCC CAGGTCTCGAACCCGAGGAGAATGTCAGAGGGCAAC ACACAACGTCCTGGAGCCAGAGGGAGGAACGAGCTAAA ACGGAGCTTTTGCCCTGCGTGACCCAGATCCGGAGTTG GAAAACAATGAAAGGCCCAAGGTAGTTACCTTAAA AAAGCCACAGCATACATCCTGTCGGTCCAAGCAGAGGAG		
MYC ΔNTD	ATGGGATCAGGTAGCGGCTCGTCTCAGAGAACGCTGGCCT CCTACCAAGGCTGCGCCAAAGACAGCGGGCACCCGAACC CCGCCCGCGCCACAGCGTCTGCTCCACCTCCAGCTTGT CTCGCAGGATCTGAGCGCCGCGCTCAGAGTGCATCGAC CCCTCGGTGTTTCCCTACCCCTCAACAGCAGCAGCAGCT CGCCCAAGTCTGCGCTTCGAAGACTCCAGCGCTTCTC TCCGTCTCGGATTCTCTGCTTCCTCGACGGAGTCTCC CGCAGGGCAGCCGAGCCCCCTGGTCTCCATGAGGAGA CACCGCCACACACAGCAGGACTCTGAGGAGGAACAAAG AAAGATGAGGAAGAAATCGATGTTGTTCTGAGAAGA GGCAGGCTCTGGCAAAGGTCAAGACTGATGACCTTC TGCTGGAGGACAGCAAACCTCCACAGCCACTGGTC CTCAAGAGGTGCCACGTCCACACATCAGCACAACACTACG CAGCGCTCCCTCCTCACTCGGAAGGGATATCTGCTGCCAA GAGGGTCAAGTGGAGCAGTGTCAAGACTCTGAGACAGAT CAGCAACAACGAAATGCAACAGCCCCAGGTTCTGG CACCGAGGAGAAATGTCAGAGGGGAACACACAGCTT GGAGGCCAGAGGGAGAACAGACTAAACGGAGCTTTT TGCCCTGCGTGACCAAGATCCGGAGTTGGAAAATGA AAAGGCCCAAGGTAGTTACCTAAAAAAAGCCACAGC ATACATCTGCGTCCAAGCAGAGGAGCAGAACAGTCATT TCTGAAGAGGACTTGTGCGGAAACGACGAGAACAGTTG AAACACAAACTTGAACAGCTACGGAACTCTGTGCG	8	Deletion of amino-terminal domain: Housing MYC Box I and II
MYC ΔCTD	ATGCCCTCAACGTTAGCTTACCAACAGGAACATGACC TCGACTACGACTCGGTGAGCGTATTCTACTGCGACGA GGAGGAGAACTCTACCAAGCAGCAGCAGAGCGAGCT GCAGCCCCCGGCCAGCGAGGATATCTGGAAGAAATT CGAGCTGCTGCCACCCGCCCTGCTCCCTACCGCT TCGGGCTCTGCTGCCCTCACTGTGCGTCAACCCCT CTCCCTCGGGAGAACAGCAGGGCGTGGGGAGCTT CTCCACGGCCGACCAGTGGAGATGGTGAACCGAGCTG GGAGGAGACATGGTAACCAAGAGTTCATGCCACCG GACGACAGACCTTCAAAACATCATCATTCCAGGACT GTATGAGGGCTCTCGCCGCGCAAGCTGCTCTC AGAGAAGCTGCCCTTACCAAGGCTGCGCGAACAGACAG CGGCAGCCCCAACCGCCCGCCACAGCGCTGCTCC ACCTCCAGCTTGTACCTGCAAGGATCTGAGCGCCGCGCT CAGAGTGCATGCCCTCGGGCTTCCCTACCCCTC AACGACAGCAGCTGCCAACAGTCTGCGCTCGCAAGACT CCAGGCCCTCTCGGTCTCGGATTCTCTGCTCTCC ACGGAGTCTCCCGCAGGGCAGGCCAGGCTCTGGTGC TCCATGAGGAGACACGGCCACACCCAGCAGCAGCTG AGGAGGAACAAGAAGATGAGGAAGAAATCGATGTTGTT CTGTGAAAAGGGCAGGCTCTGGAAAAGGTCAAGAGT CTGGATCACCTCTGCTGGAGGGCACAGCAAACCTCCTCA CAGCCCACTGGTCTCAAGAGGTGCCACGTCTCCACACAT CAGCACAACACTACGGCAGCGCTCCACTCGGAAGGACT ATCCTGCTGCCAACAGGGTCAAGTTGGACAGTGTCAAGAGT CTGAGACAGATCAGAACACCGAAAATGCACCGAGCCC CAGGTCTCGAACACCGAGGAGAATGTC	9	Deletion of carboxy-terminal domain: Housing basic helix-loop-helix leucine zipper motif, governing heterodimerization with MAX protein
MYC Glu39A1a	ATGCCCTCAACGTTAGCTTACCAACAGGAACATGACC TCGACTACGACTCGGTGAGCGTATTCTACTGCGACGA GGAGGAGAACTCTACCAAGCAGCAGCAGAGCGAGCT GCAGCCCCCGGCCAGCGAGGATATCTGGAAGAAATT CGAGCTGCTGCCACCCGCCCTGCTCCCTAGCGCCG TCGGGCTCTGCTGCCACCCGCCCTGCTCCCTAGCGCCG TCGGGCTCTGCTGCCCTACGTTGCGTCAACCCCT	10	Point mutation changing Glutamic Acid to Alanine at amino acid 39

- continued

GENE	SEQUENCE	SEQ ID NO:	MUTATION
	CTCCCTTGGGGAGACAACGACGGCGGTGGCGGGAGCTT CTCCACGGCCGACCAGCTGGAGATGGTGACCGAGCTGCTG GGAGGAGACATGGTGAACCAAGAGTTTCATCTGGCACCCG GACGACGAGACCTTCAAAACATCATCATTCCAGGACT GTATGGAGCGGCTTCGCGCCGCGCAAGCTGCTCTC AGAGAAGCTGCCCTTACCAAGGCTGCGCGAACAGACAG CGGCAGCCCCGAAACCCGCCCGGGCACAGGGCTGCTCC ACCTCCAGCTTGACCTCCAGGATCTGAGGCCCGCCGCT CAGAGTGCATGACCCCTCGGGCTTCCCCTACCCCTC AACGACAGCAGCTCGCCCAAGTCTGCGCCTCGAACAGACT CCAGCGCCTTCTCCGCTCCGGATTCCTGCTCTCCCTCG ACGGAGTCCCTCCCGCAGGGCAGGCCGAGCCCCGAGCCCTGGTGC TCCATAGGGAGACACGGCCACACCCAGCAGCAGCTCTG AGGAGGAACAAGAAGATGAGGAAGAAATCGATGTTGTT CTGTGAAAAGAGGGCAGGCTCTGGCAAAAGGTCAAGAGT CTGGATCACCTCTGCTGGAGGCCACAGCAAACCTCTCA CAGCCCACTGGTCTCAAGAGGTGCCCCACGTCTCCACACAT CAGCACAACATACGCAGCGCTCCCTCACTCGGAAGGACT ATCCCTGCTGCCAGAGGGCTCAAGTTGAGCAGTGTCAAGAGT CCTGAGACAGATCAGCAACACCGAAAATGCAAGCAGGCC CAGGTCTCGCACCCAGGAGAATGTCAGAGGGCAAC ACACAACGTCTGGAGCGCCAGAGGGAAACGAGCTAAA ACGGAGCTTTTTGCCCCGCTGACCCAGATCCGGAGTTG GAAAACAATGAAAAGGCCCCAAGGTAGTTATCCTTAAA AAAGCCACAGCATACATCCCTGCGTCCAAGCAGAGGAG CAAAGCTCATTCTGAAGAGGACTTGTGCGGAAACGAC GAGAACAGTTGAAACACAAACTGAAACAGCTACGGAACT CTTGTGCG		
MYC Thr58Ala	ATGCCCTCAACGTTAGCTTACCAACAGGAACATGACC TCGACTACGACTCGGTGAGCGGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCAAGCAGCAGCAGAGCGAGCT GCAAGCCCCGGCCAGCGAGGATACTGGAGAAATT CGAGCTGCTGCCGCCGCCGCCCCCTGCTCCCTAGGCCCGC TCCGGCTCTGCTGCCCTTACGTGCGGTACACCCCT CTCCCTTGGGAGACAACGACGGCGGTGGCGGGAGCTT CTCCACGGCCGACCAGCTGGAGATGGTGACCGAGCTGCTG GGAGGAGACATGGTGAACCAAGAGTTTCACTGGCACCCG GACGACGAGACCTTCAAAACATCATCATTCCAGGACT GTATGGAGGGCTCTCGGCCGCGCCAAAGCTGCTCTC AGAGAAGCTGCCCTTACCAAGGCTGCGCGAACAGACAG CGGCAGCCGAACCCGCCGCCAGCGAGCTCTGCTCC ACCTCCAGCTTGACCTGCAAGGATCTGAGGCCGCCGCT CAGAGTGCATGACCCCTCGGTGGCTTCCCCTACCCCTC AACGACAGCAGCTGCCCAAGTCTGCGCCTCGAACAGACT CCAGCCCTTCTCCGCTCCGGATTCTGCTCTCCCTG ACGGAGTCCCTCCCGCAGGGCAGGCCGAGCCCCCTGGTGC TCCATAGGGAGACACGCCAACCCAGCAGCAGCTCTG AGGAGGAACAAGAAGATGAGGAAGAAATCGATGTTGTT CTGTGAAAAGAGGGCAGGCTCTGGAAAAGGTCAAGAGT CTGGATCACCTCTGCTGGAGGCCACAGCAAACCTCTCA CAGCCCACTGGTCTCAAGAGGTGCCCCACGTCTCCACACAT CAGCACAACATACGCAGCCCTCCCTCACTCGGAAGGACT ATCCCTGCTGCCAACAGGGTAAGTTGAGCAGTGTCAAGAGT CCTGAGACAGATCAGCAACACCGAAAATGCAACAGGCC CAGGTCTCGCACCCAGGAGAAATGCAAGAGGGCAAC ACACAACGTCTGGAGGCCAGAGGGAAACGAGCTAAA ACGGAGCTTTTTGCCCCGCTGACCCAGATCCGGAGTTG GAAAACAATGAAAAGGCCCCAAGGTAGTTATCCTTAAA AAAGCCACAGCATACATCCCTGCGTCCAAGCAGAGGAG CAAAGCTCATTCTGAAGAGGACTTGTGCGGAAACGAC GAGAACAGTTGAAACACAAACTGAAACAGCTACGGAACT CTTGTGCG	11	Point mutation changing Threonine to Alanine at amino acid 58
MYC Ser62Ala	ATGCCCTCAACGTTAGCTTACCAACAGGAACATGACC TCGACTACGACTCGGTGAGCGGTATTTCTACTGCGACGA GGAGGAGAACTTCTACCAAGCAGCAGCAGAGCGAGCT GCAAGCCCCGGCCGCCAGCGAGGATACTGGAGAAATT CGAGCTGCTGCCGCCGCCGCCCCCTGGCCCTAGGCCCGC TCCGGCTCTGCTGCCCTTACGTGCGGTACACCCCT CTCCCTTGGGAGACAACGACGGCGGTGGCGGGAGCTT CTCCACGGCCGACCAGCTGGAGATGGTGACCGAGCTG GGAGGAGACATGGTGAACCAAGAGTTTCACTGGCACCCG GACGACGAGACCTTCAAAACATCATCATTCCAGGACT GTATGGAGGGCTCTCGGCCGCGCCAAAGCTGCTCTC AGAGAAGCTGCCCTTACCAAGGCTGCGCGAACAGACAG CGGCAGCCGAACCCGCCGCCAGCGAGCTCTGCTCC ACCTCCAGCTTGACCTGCAAGGATCTGAGGCCGCCGCT CAGAGTGCATGACCCCTCGGTGGCTTCCCCTACCCCTC AACGACAGCAGCTGCCCAAGTCTGCGCCTCGAACAGACT CCAGCCCTTCTCCGCTCCGGATTCTGCTCTCCCTG ACGGAGTCCCTCCCGCAGGGCAGGCCGAGCCCCCTGGTGC TCCATAGGGAGACACGCCAACCCAGCAGCAGCTCTG AGGAGGAACAAGAAGATGAGGAAGAAATCGATGTTGTT CTGTGAAAAGAGGGCAGGCTCTGGAAAAGGTCAAGAGT CTGGATCACCTCTGCTGGAGGCCACAGCAAACCTCTCA CAGCCCACTGGTCTCAAGAGGTGCCCCACGTCTCCACACAT CAGCACAACATACGCAGCCCTCCCTCACTCGGAAGGACT ATCCCTGCTGCCAACAGGGTAAGTTGAGCAGTGTCAAGAGT CCTGAGACAGATCAGCAACACCGAAAATGCAACAGGCC CAGGTCTCGCACCCAGGAGAAATGCAAGAGGGCAAC ACACAACGTCTGGAGGCCAGAGGGAAACGAGCTAAA ACGGAGCTTTTTGCCCCGCTGACCCAGATCCGGAGTTG GAAAACAATGAAAAGGCCCCAAGGTAGTTATCCTTAAA AAAGCCACAGCATACATCCCTGCGTCCAAGCAGAGGAG CAAAGCTCATTCTGAAGAGGACTTGTGCGGAAACGAC GAGAACAGTTGAAACACAAACTGAAACAGCTACGGAACT CTTGTGCG	12	Point mutation changing Serine to Alanine at amino acid 58

- continued

GENE	SEQUENCE	SEQ ID NO:	MUTATION
	GACGACGAGACCTTCATAAAAACATCATCATCCAGGACT GTATGGAGCGGCTTCTCGGCCGCGCAAGCTCGTCTC AGAGAAAGCTGGCTCTTACCAAGGGCTCGCGCAAAGACAG CGGCAGCCGAACCCCGCCCGGGCCACAGCGTCTGCTCC ACCTCCAGCTTGTACCTGCAGGATCTGAGGCCCGCCCT CAGAGTGCATGACCCCTCGGTGGTCTTCCCCTACCCCTC AACGACAGCAGCTCGCCCAAGTCCCTCGCGCTCGAAGACT CCAGGCCCTTCTCCGCTCCGATTCTCTGCTCTCCTCG ACGGAGTCCTCCCGAGGGCAGCCCCGAGGCCCTGGTGC TCCATGAGGAGACACGCCAACCCAGCAGCAGGACTCTG AGGAGGAACAAAGAAGATGAGGAAGAAATCGATGTTGTTT CTGTGGAAAAGAGGCAGGGCTCTGGCAAAAGGTCAAGAGT CTGGATCACCTCTGCTGGAGGCCACAGAACCTCCTCA CAGGCCACTGGTCTCAAGAGGTGCCACGTCTCCACACAT CAGCACAACACTCGCAGGCCCTCCCTCACTCGGAAGGACT ATCCTGCTGCCAAGAGGGTCAAGTTGGACAGTGTGTCAGAGT CCTGAGACAGATCAGAACACCGAAAATGCCAACGCC CAGGTCTCGACACCGAGGAGAATGTCAAGAGGGGAAC ACACAAACGTCTGGAGCCAGAGGAGGAACGAGCTAAA ACGGAGCTTTTGCCCTCGCTGACCAGATCCGGAGTTG GAAAACAATGAAAGGCCCCAAGGTAGTTATCCTTTAAA AAAGCCACAGCATACATCCTGTCGTCGAAGCAGGAGG CAAAGCTCATTTCTGAAGAGGACTTGTGGGAAACGAC GAGAACAGTTGAAACACAAACTTGAAACAGTACCGAACT CTTGTGCG		

[0114] Additionally, the consistent and strong effects of KLF4 overexpression motivated the investigation of the full KLF zinc finger transcription factor family (FIG. 2F) 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. 2G), 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. 2G).

#### KLF Family Library

#### [0115]

GENE	SEQUENCE	SEQ ID NO:
KLF1	ATGGCGACTGCGGAGACAGCACTTCCATCAATCTCAACACTCACTGCACTG GGGCCATTTCAGATAACCCAGGAGCATTCTTAAGTGGTGGCGGTCCGAA GAGGCTCAAGACATGGGACTCTGGTCGCCGAGATCCCACCGAACCTCCCTCG CATGTCAAAGTGAAGATCAGCCTGGCGAGGAAGAGGATGACGAAAGGG GTGCCGACGCCACTGGGACTTGGATCTTCCTTACCAATTCTCTGGTCC GGAACCTGGGGGGCACACAGCTGCGCTCTCGCTCCCTCAGAACGG GCGGGGCTCAGTACCCACCCCTCCGAAACTGGGAGCCTATGCTGGGG GTCTGGACTGGGCTGGGTCTGGTAGTGGAGGACATTCTGGCTGGG TACCCCCCCTTTGAGGGCCCGCCTCCGGACGCCCTTGTGGGACCGGGC TCGCTCTGACCGGCTCGGAACCAAAGGCCCTCGCGCTGCAGCCGTGT ACCCCGGACCGGAGCCGATCTCAGGGGATACTTCCACGGACCGGA CTCAGCGTTCAGGGCTTCGGGCCATACGGATTGTTGAGCGGCTAC CCGGCTATGTATCCCCTCCCGTCAACAGGACACTTCAATTGTTCCGG GGTCTTCAAGGGCTCGGCCGGGCTCTCAGTCCAGTTCCCTCAGT TGTCTGGGACGGGAACTGTGGACTTGGGACTTGGGGACTGCAGAGGA CCCAGCGTTATAGCAGAGACAGGGCAAGTAAAGGGGCCAGGAAGCT GGGCAGGAAACGCCAGCGCACACTGTGCCCATCCAGGTTGCCGT AAATCTTACAGAAGCGATCATTTAAAGCACATCTCGCACACAC GGGGAGAGGCCCTACGCCGTACTGGGAAGGTTGCCGTGGAGATTG CTAGATCTGACGAGCTCACCCGGATTATCGAAAACACACTGGCCAGCGA CCGTTCCGGTGCCTGCAACTCTGCCAAGGGGTTGAGTCGCTCAGATCATCTG GCTTGCATATGAAAGCGACACCTT	13
KLF2	ATGGCCCTTAGTGAACCCATTCTCCCAAGCTTTCCACGTTCGCGTCTCCTT GCCGAGAGAGAGGCCCTCAGGAAAGGTGGCGAGGGCTGAACCGAGCT GGAGGTACGGATGATGATCTAACAGTGTGCTCGATTTCTACTCTCAATG GGACTGGACGGCTGGAGGGAGCGAGCTCTGAACCCACCCAC TCCGCCCCCAGCGTTTACTACCCGGAGCCAGGTGCGCCGCCATATT AGCCCCGGGGTGGCTGGAGCTCCCGCCATGAAATTGGATGC	14

- continued

GENE	SEQUENCE	SEQ ID NO:
	CCCGCTGGCCCGCGCTGCATGGTAGATTCTGCTCGGCCCTGGGTG ACTCGTTAAGGCTAACCTCCTGAGGCTATGGTGGAGGTGGCTACGGAT GTGGCCCCGGCTACCCGAGGACGGAGGTCTAACGGGGAAAGGGCA CCTGGCCCCGGCTGCAAGCTGTATGGGGGCCCCGGTGGGAGGCTCCCC GCCCTGATACACCCCCCTTACTCCAGATGGACCAGCTGACTTCCC ACCTGGCCCCAGAGCGAGTTCCCCCTCATTTGGAGGACCGGGTTGG CGCCCGAGGTCTGGACTTCACTACGGCCCTCTGGCCCCCACTTTGGT CTTTCGACGATGCTGCTGCCCCAGCAGCCTTGGGCTTGCCCC GCAGCGAGGGACTGCTCACGCCAACCGGAAGCCCCCTGGAGCTCTGA AGCCAAGCCGAAGCGAGGACGCAGATCATGGCCGCGAAGCGGACAGCT ACGCATACTGCTCATATGGGGCTGGAAAAAACCTACACAAGAGTTC ACACCTTAAAGCGCACCTCGCACACACAGGGAGAAACCATATCATT GTAACTGGGACGGATGTGGATGAAATTGCTCGTCTGATGAGCTTACGA GACATTATGAAAGCATACGGACATCGGCCCTTCAATGCCATTTGTG ACAGAGCTTCCCGGTCTGACCACCTCGCTCTGCACATGAAGAGGCACA TG	
KLF3	ATGCTCATGTTGACCCAGTTCTGTCAGCAAGCAAGAGGCCATGGACCCGTG TCAGTGTATACCCATCATTAATTACATGGAATCATGAAAGCCTAACAGTAT GGGGCATCTACTCCACACATTGCGTGGAGAAGTTTTCAGACCCAGAA GGTCTGTCGACGGAATACAGATGGAGCAGCTACGGTGAAGCAA GCGGAGTTCACCCCCCTCGCTGGGAATTGCCCTCTCTGAAGTTCC GTCCTCACACGGAGAGCTCGCTGGGTTGAGCATGCTTCTCCAGGCC ACCGATAAAAATACTCACCCCCCTCTCAGGGTGCAGCCTTGGCGT GCCGTGTCATGCCACAGTGTGGCAGCTGCCCTCGGGCATGGAAT ACGGAGCCGGGGATCTGCCGTATCCAGGCCGGTGGTGTGCAGGCC TCCCTTATGTAACAAGTCACCTCCAGCAGCCTCTATGGTCTCCATTAC GGAGGAGATGGAAAATTCCAGTAGTACATGCAAGTACCTGTAAATTGAAT CATATGAGAGCTTATACAGAAAAAAATTAAATAGAACCTGGGATC GAACACAGAGGACAGATTATTCTGAAGAAATGTCACCCCCCTTAATG AACTCAGTGTCCCCCCCAGGACATTGTTGCAAGAGAATCACCTCGTC ATCGTCAGCTGGGAAGAGACCTTACCTGTGGAAATCCCCGGATACTCAA AGGAAGCGAGGATACACAGATGTGATTATGATGGATGCAACAAAGTGA CACTAAAGCTCCACTTGAAGACACAGAACAGAACACACAGGAGAAA AACCTAACAAATGTACATGGGAAGGGTGACATGGAAGTTGCTCGGT GATGAACTAACAAAGACATTCCGAAACATACTGGGAATCAAACCTTCCA GTGCCGGACTGTGACCGCAGCTCTCCGTTCTGACCATCTGCCCTCCAT AGGAACGCCACATGCTAGTC	15
KLF5	ATGCTACAAGGGTGTGAGCATGAGCGCCGCTGGGACCCGTGCC GCCGCCGGCGCGCAGGACGAGCCGTGTCGCCAGCTAACGGGTGC TGGGGCCCGGAATCGGCCCGCACGCCAGGCCAGGCCGCCGCCAGGC CTGAAGCACGCCACCACGCCGCCAGGCCAGGCCGCCGCCAGGC CCCCAGCGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGC ACCTGGTCAGACAGATGTGAATGGAGAAGTATCTGACACCTCAGCTT CCTCCAGTTCTATAATTCCAGACATAAAAAGTATAGACGAGACAGTG TCAGTCGTAGACCGATTCTCACTGACACTGAAGGGTACCTACAGTATC AACATGAACTCTCTCTCTGACATCACTCACCTGAGAAACTGGCTCTAC AAATCCAGAGACCGTGCACACACATCAAGACAGAACCTGTGCCAT TTTCAGCACCAAGAGTGAACAGACTGCCCTCTCGGCCGCCACCCAGC CCTCCCTGAGTTCAACAGTATTCAGCTCACACAGACCGCAGCTCCAGA GGTGAACAAATTTCTCATCAAACAGAACTTCTACACCCAGATCTCATCT TTCTGCCCCCTACCCAGCAGGCCACCTGTACCGCTACTGAAATACACCGGA TCTAGATATGCCAGTTCTACAAATCAGACAGCAGCAATGGACACTCTAA TGTTCATGTCAGTGCATGGCAGGCCAAACACACACACCTGCTGTT CCGAGACACTGCAAGAACATTCCAGGCCATGCCCTTGCAACATACAC AATGCCAACTGAGTTCTCCACACAGGCCACTTACTTCCCCGTCC ACCAAGCTAGAGGCTGGAGTCCAGATGAGACAGCAGAGATGCTCCAGA ATTTAACCCACCTCATCTGCTACAATTGCTTCTAAACTGGCAAT TCACAAATCCAAATTACCCACCCACTGCCAGTTAACTCACAACATCCA ACCTGTCAGATAACATAGAAGGAGTAACCCGATTGGAGAAAGACGCA TCCACTACTGCGATTACCTGGTGCACAAAAGTTATACCAAGTCTTCTC ATTTAAAGCTCACCTGAGGACTCACACTGGTGAAGACAGCAGATG ACCTGGGAAGGCTGCAGTGGAGGTTCGCGCAGTGGATGAGCTGACCC CCACTACCGGAAGCACACAGGCCAGGCCCTCCAGTGCAGGGGTGTGCA ACCCAGCTCTCGCGCTCTGACCACCTGCCCTGCATATGAAGAGGCC AGAAC	16
KLF6	ATGGACGTGCTCCCCATGTGCAGCATCTCCAGGAGCTCCAGATCGTCAC GAGACGGCTACTCTCGCCGCTCTGAGGAGTACTGCAACAG ACCTGCTAGAGCTGAACTTCCAGAGGCCAGGCCCTGCTATGTTCA GCCTCAGAAATCAAATTGACAGCCAGGAAGATCTGGAACCAAATCAT TCTGGCTCGGGAGAAAAGGAGGAATCCGAACTGAAGATATCTCCAGTC	17

- continued

GENE	SEQUENCE	SEQ ID NO:
	CTCCAGAGGACACTCTCATCAGCCGAGCTTTGTACAACTTAGAGACCA ACAGCCTGAACTCGATGTCAGCAGCAGAATCCTCTGACAGCTCCGAGGA CTTCTCCACGGCCAAGTTCACCTCCGACCCATTGGCGAAGTTTGCTCA GCTCGGAAAATTGAGCTCTCTGCACCTCCACGCCCTCCATCTCTCCGG AACTGAGCAGGGAACCTCTCAACTGTGGGGTGCCTGCCGGAGCTG CCCTGCCAGGGAGGTGCGCAGCGGACTTCGGGAAGCAGGTGACA GGGAATGGCGATGCCCTCCCGACGGCAGGAGGGTGCACCGGTGCC ACTTTAACGCTGAGGAAGTTACACAAAAGCTCCACTTGAAGACCA CACCAAGCGACACAGGAAAGCCTTACAGATGTCATGGGAAGG GTGTGAGTGGCTTGCAGAAGTGTAGGTAAACAGGCCTCCGAA AGCACACGGGGCAAGCCTTTAAATGCTCCACTGTGACAGGTGTTTT CCAGGTCTGACCATGGCCCTGACATGAAGAGGCACCTC	
KLF7	ATGGACGTGTTGGCTAGTTATAGTATATTCCAGGAGCTACAACGTCCAC GACACGGCTACTTCTCAGCTTACCATCCCTGGAGGAGACCTGGCAGCAG ACATGCCCTGAATTGAAACGCTACCTACAGACGGAGCCCGGAGGATCTC AGAGACCTTGGTGGAGGACTTGGACTGTTCCCTCACGCTTCCCCCTCCCC GTGCAATTGGAAAAGCTTCCGTCGCTTAGACCCCCCTGCTGCTCCCCGTGGA AGCGCCCATCTGTGAGAAAGAGCTGGCAGTGGACATCTGCTCTCGGA CAAGTTGCTATCTGAGACCTGCCCTCAGCTCCAGCCGGCAGCTTCTCT AGACAGCTACACAGCGTCACACCAGGCCAGCTAACGCAGTGACCTCAT TAACGCCCATCTGCCCTGAGCTCAGGCCATCTGGTCAAACACCTCAC AAACTCTCTGCCGTGGATGGCAGCTGAGCTTGAAGAATGGTGGCCAAG AAGGTGCTCTCAGCTCCGTAAGGTGGAGGGTGCACACAGTGCAGC AGCGTGCAGGCTGGCTGGGGCCGTTAAGAGTGGACAGAGCGACAGTGCAC AAGGAGGGCTAGGGCTGAAGCATGCCCAGAACAGAAGAGGGTCA CCGCTGTCAAGTTAACGGTGGCCGAAAGTTATACAAAAGCTCCACTT AAAGGCCACAGGAGACTCACACAGGTGAGAACGCTTATAAGTGTCTCAT GGGAGGGATGTGAGTGGCTTTCGACAGCAGTGTAGCTCACAGGAC TACAGGAAACACACAGGTGCAAAGCCTTCAATGCAACCACTGCGACAG GTGTTTCCAGGTCTGACCATCTGCCCTCACATGAAGAGACATATC	18
KLF8	ATGGTGCATATGGATAACCTCATAAACAATTGGAGGTCAAACCTTAATTCA GAAGGTGGCTCAATGCAGGTATTCAGGACTCTGTTCTGTGCGAAC AGAGATCCCCCTGAGATAGAATACAGAAGTAATATGACTTCTCCAACACTC CTGGATGCCAACCCATGGAGAACCCAGCACTGTTAATGACATCAAGATT GAGGCCAGAAGAACCTTTGGCTAGTGAATTCAAGCTGCCAACGTGAA CCAGGTGACCTCTCTTACAAGCCCAAGGCTCTCCAGCCCTGCTAGC ATGCTACAAGCTCCAATACGTCCTCCAGGCCACAGTCTCTCCCCAGACC CTTGTGGTGCACGTCACATCTGACATGAGCACTTCAGCAACATTCT ACTGTTCTGACCCCCAGGCTGTCCTGACCTCCTCAGAGCACTGGTAGC CAGCAGATCTACATGTCATTCAACTATCCCCCTCAGTCAGTGTGCAAAT AAGATGGTGGCCCTGAAGACCATCCAGGTGAGCTGCTGAGTCTCTGCCATG GTGTACTACTTTGCTCTGAGATGGGGCCCTGAGCCATTACAGTCCCA CTCATGGAGGAGATGGTAAAATGCTGGATCACTGAAAGTGTGACCCAC CTCCATGTCTCCACTGGAATTCCAAGTGACAGTGGAGGAGTACAATTGA GAGTGGACTCTCAGGCTTCCAGACTCTGCAAGGGACTACAGCAAGAACAG CAGCAATGCCCAATGCAGGGAGAGAGCTGCTGACTTGAAGAGAAGA GGAGTCAAGATGTGACTTGCAGGATGCGACAGCAAAGTGTACACCAAAG CTCTCACCTGAAAGCTACCCGAGAATCCATACAGGAGAGAACCTTATA AATGACACCTGGATGGCTCTGGAAATTGCTCGCTCAGATGAGCTCA CTCGCAATTCCGAAAGCACACAGGCATCAAGCCTTTCGGTGACAGACT GCAACCGCAGCTTCTCGTGCACCATGCTGCCCTGCATGCCGTGCCA TGACACCATG	19
KLF9	ATGTCGGCGGCCGCCACATGGACTTCGTTGGCTGCCAGTGTCTGGTTCC ATTTCGAACCGCGCTGCGGTGCGGAGGATGGGTCGCTCCGGACGCCGA GCGCTGCGACTACCTGAGCGAGGTGACCAAGGAGCACGGTGACCCGG GGGACACTGGAGGGATTACTGCACACTGGTACCCATGCCAACAGCTTG TTGGACCTGAACAAGTACCGACCATCGACAGCCCCCTCGTGTGCGAC AGTCTGGAAAGTCCAGATGAGGATATGGATCCGACAGCAGCTGAC CGAATCTGGTCGAGTCCCTCCACAGGCCGGAGGAGACAGGATCTG GCACCGGCCAGCCGCTCTCCCTCCATCTGGAGTGGCTGGAGG GGAAACACGCCTCCGAAAAGAGGACAAGTGCCTTACAGTGGCTGTGG AAAGTCTATGAAATCTCCCATCTCAAGGCCATTACAGATGTCATACA GGTGAACGGCCCTTCCCTGCACTGGCCAGACTGCCTTAAAAGTTCTC CGCTCAGACAGGCTGACCCGCCACTACCGACCCACACTGGGAAAGCA GTTCGCTGCCGTGAGAAGCGCTTACAGGAGGAGTGCACACCTCAC AAAGCACGCCGGCGACACCGAGTCCACCCAGCATGATCAAGCGAT CGAAAAAGGCCTGCCAACGCTT	20

- continued

GENE	SEQUENCE	SEQ ID NO:
KLF10	ATGCTCAACTCGGTGCCCTCTCCAGCAGACTGCGGAGGAAAGAATGGA AATGATTTCGAAAGGCCAAAAGGAGAGTATGTTACCTGGAACAAAACCTG CAGAGAAAAGTGATTTGAGCTGAGAAGCACTTATGTCATGAGCTGC AGTTGGAAAGTCTGATTTAAGAAATACGTTGAAACAGACCTGTACACCA GTATCTGATTTGTAGAGGAAGAGAACTGCTTCCGGGAAACACCTGATTT CATACAATCCAGCATTGTTGACTCACCCTACAGTCCTTGACTTTG AACCTCTCAAGTCATACTGTAGGGACCCAGCCATCTACTGTACACT TCAAGTCAGTCACTGCAAACTTCACATGGCCACCTTCAAAG AGGAAGAAAAGAGCCCAGTATCTGCCCAAACCTCCCAAAGCTCAGGCA ACAATGTTGATTCTGCATACAGTGATGCCAGTATGTAACCCAGACCC TGCCCAATGAAAGCAGCCAGCATCTCAACTATCAGAACAAATTCTTTAGA AGAAAGAACCCCCCTAAATGTTGAGGCTGCAAGAAAGAACATACATGTC CGCTGTCAACCAACAGATCCAATGAGAGAAAACACAGTGGCAGATG TTGATGAGAAAGCAAGTGTGCACTTATGACTTTCTGTGCCCTCCCTAG AGACGGTCTATCGCAGGCTCAGCAGGCCCTGTGTCcccAACACAGAAGT CAGTGTGTTCTCCACCTGCACTATGCAAGGGGAGTGCACCTATG CGGTCACTGCAAGATGGTCCCCCTCTGCCAACAACCTGTTGTGACAA CAGTGTGTTCCAGCACTCTCCAGGCCACAGCCACAGCGTTGCCCCCTG TTGTTCTGGGACACAAAGTCCCAAGGCGCTGTGATGTTGTGTT CCCAGCCCGTTGTCAGAGTTCAAGCCTCGGGTGGTGAAGCCGAATGGC ACCAGACTCTCTCCATTGCCCCCTGCTCTGGGTTTCCCTCAGCAGCAA AAGTCACTCTCAAGATTCAAGGATAAGGAGTCACATCTGTAGCC ACCCAGGATGTCAGAACATACTTTAAAGTTCCATCTGAAGGCCAC ACGGAGGACGACACAGGAGAAAAGCCTTTCAGCTGTAGCTGAAAGGTTG TGAAAGGAGTTGCCCCTCTGATGAACTGTCCAGACACAGGGAAACCC ACACGGGTGAGAAGAAATTGCGTGCCTCATGTTGACCCGGTTCATG AGGAGTGAACATTGACCAAGCATGCCGGGCCATCTATCAGCCAAGAA GCTACCAAATGGCAAGATGCAAGCTAAATGACATTGCTCTAC CTCAACCCCTGCTCCACACAG	21
KLF11	ATGCATACTCCTGATTGCTGGACCTGACGACGCCGAGCCGTGGACATT ATGGACATTGTGAATCTATACTCGAAAGAAAGACATGATTAGAGGG AACTACATGCTCTACCTCGAGCAAACAGACATGGAGGCGTAGAACAGCTC TGGTGTGATGTCAGTGGGTAGAGATCCAGAAGGGGACTTGTGTTA GAATGAGCGCTACTCCAGTCCGATAGCGGCAGCTAACAACTACTG TTCATATGGACGCGCCACCCCTGAGCTGCCAAAGACTTCAAGCCTCT CAACTCTTGCATCACTCACCAACAGTCCCCGATCTGTGCAACCATCAA CCCCGACCCCTGTTAGCCCCCAAGTTACAGATTAAAGGCGTGTACCGGA CCGATGTTCTGCAGAGTTAGCGGGCATTGAGCGGAGGG GCTGAACGAGGCTGTGGGACTTGAACCCGATCCGAGTTCTCTGTAGA GCCAAGGGTACTAGTGTATTGCGCATACGGCGAGAGTCCGGCAGCTGT TTCCCACCATACAAACCCCCAGACTGTGCGCTTAGTGAATCCGGGAAAGG GAGAACACAGCTGTGGGCACTTGAGGACACTTCAAGATACACATTGAC AGATACTTGTGTCACCAACCTGGTGTGATGTCACCTGTTGCACAA GTCCGGGGTCTCTCTGACTGACAAAGGTCAACAAGGGGATGGCTG GCGCTGTCACACATGCACTTAAACTACGAAATGATTGCTTAGG AAAACACAGCTGTGGGCACTTACGTTGAGTGTGCTCCACCTGCTGTG CAGATGATCCCTGTAACCGGGCAATCATCTATGTTGCGCTGCGTTTGAAG CCCCCCCACAACTGTCCGGTGTACTGTTGCGGATCTTGGCAAGCA GCGCCGCCCGCAACCGTGTGGGGCCCGTGTCCCGAGGGTGC GTCATGTTGTTCTCCCAAGGGGCCCTCCGCCACAGCTCCGTG GCGAATGTCATGGCTGCCGGAAACACGAAATTGTTGCCCTTGCAACCGCT CCAGTTCTACAAAGGCTCACAGAATTGTTGCGGCCAACAGTCACCTCTCA CGAACAGGAACTATGTTGCTCTTCCAGGTTGCAAGAAAACATATTTC AAATCTCTCATGAAACACATCTGGACCCCATACAGGAGAGAACGCT TTAATTGAGCTGGGATGGTGTGATAAAAATTGCAAGAAGTGTG GCTCAAGTCGACATGCCAGGACGATACCGGGAAAAAAATTGCTTGT CAGTTGTGACAGAAGATTATGAGGTCGACCATCTACCAAGCACGCG GACGCCACATGACTACAAAGAAAATTCTGGCTGCAAGCCGAGGTGGG AAACTCAACCGAATGCTTCCGCTGAATCCCCGGCAGCCGCTGGTAAGT ATGCTGCCAGTGCC	22
KLF12	ATGAACATTACATGAAAGCGCAAGACGATAAAAGAACATCAATACATTG GAACCGAATGTTGATGTTGGATGCCATGCCGCTGTACGGTAAAACCG AGCTCTGGAGTCGAACAAGGATCCCCAAACGTCACAAACTACCCGAT ATGGAGGCGAGTGCCTCTGCTCAACATGTAAGGGAGAGCCGCTGA GGACTCTCTCCCTGAGATCATTTCAGACACAGACTGAGCCCTGAGATCT TTCAATTAAACAAAGCCAGAACATCTCCTACTGCGGTAAGTTCTCTCCGT AAGTATGACAGCAAGTGCATCTAGTCAGGATCAGGACTAGCAGTTC TTCATCTAGTAGACTTGCTAGTTACCAACGGTGTACAAGTGTGTTCTAG GCCAGCAGCAGCTAACGGTACTGACTCCGGTCACTCGTGGCAAGCG CTAGTGGCGTGGGTGGCCAACAAATTCTCCATATTACCCCGTGCCTC	23

- continued

GENE	SEQUENCE	SEQ ID NO:
	CGCTAGTCGATGAATCTCAGAGCAACAAGCTTAGTCACGTACATAGGA TCCCCGTGTCGTCAGTCAGTTCCGTCGTCACAGCTGTGCGATCCCC TGGGAATGTCAATAATACTATAGTGTGTCAGTCAGGGATGGTAGGG CCATGGGAAAGCACAGATGGACCCCCCGCGCTTGTCAACGGAGACAGTCTA AATCCGATAGTGACGACGATGATTGCTAACGTAACACTGGACTCTGTGA ACGAGACCGGGAGTACCGCTCTGCAATCGCTAGGGCGTACAGGAGGTC CACCCAAGGCTGTGTCAGAGTCCGAGGTAACAGGATGAATAATCGAAA ATTTCCCTGTACCATCAGCCCATTCTATAGACTCCACTCGGAGACACCG ACGAAGTGAATCACCCGACTCCAGAAAAAGGAGGATACTCGCTGTGACT TTGAGGGCTGAAACAGGCTACACAAAAAGTTCACACCTCAAGGCCAT CGACGGGAGCATACTGGGGAAAACCGTACAAAATGCACCTGGGAGGGATG CACGTGGAAATTGCAACGCTCTGACGAGTTGACAGCCACTATCGAAAGC ATACCGGCTAAAGCGTTTAAATGCGTGATTGGACAGGGATTAGGCC GCTGTGATCACCTGCTCTCACCGGAGGCACATGCTTGT	
KLF13	ATGGCTGCGGCTGCATATGGATCATTTGCGGCTGAGTCGCTGGTGTCA ATGCTAGTAGAGCCGTGTTACACGGTCCAGAGAAGGCCAGAATCAGG CCAGAGGGCGCCCGCGCTGCAACACCGACGCTGCGCTCGGGTCGAGG AGGCCGACGGAGGACAGTCGCTACTTTCTGAGTAGCGAGAAATA TTGCGAGATCTGAATCAACAGGCTCCAGCACCTGCGCCCGCTGAAACGCC GGAGGGCGCGCTGCGAGAAAGGCCAGAACACCATGCGCTTGCAGGCCAC CTGCGCCAGAACCCACAAGTCCAGGTGCCAGGGTGCAGGCCGCTGCCCT CCTTACCGGCCCTGCTGAACCAAAGCAGGAGGCCAGGTTGCAACCTGAA GCGGAACCCGGCCCTGCAAGGCTTGGGAACCTTGGCTTGAGGAGCGGG TGAGGGCGGCCCGAGCAGGGCCGACCTGGAAATACCGCAAAGGAAACAT AAATGCCATTATGCTGGTTGCGAAAAGGTTATGAAAGTCATCCACCTG AAAGCACACCTCCGACTCACACGGGTGAGCGACCTTTGCGTGTTCCTGG CAAGACTGCAATAAAAGTTGCTAGATCTGATGAAACTTGCAACGGCATTAT CGAATCATAACCGGTTGAAAGAAGTCTCATGCCCTATATGTGAGAAACG GTTCATGCGCTGACCATTGACGAAACATGCAAGACGACATGCTAATT TCATCGGGGATGTTGCAAGAGACGGGAGGGGAAGTAGGACTGGAAGTC TCTCGGACTATTCCGATCCGACGCTTCCACCAACGATTAGCCCCGAA GCAGTCCC	24
KLF14	ATGTCAGGCCAGTCGATGCCCTGATTACTTCGGGCGAGTGTCTTGT CCATGTCAGGGGCTGCGTGTGATCAGAGACACCAAGACGGGAGGGA GGGGAGGGCGAGCTGGATCTGAAGTTCGGCGCGCTCCACCTGAATCAGC GCTTCCGGCCCTGGCTCTCGAGTCGGCTAGCGTGCCTAACCTCCACA AGTGCCTGCTCCGAGTCGGAGGGCGAGCAGCCCGCATCTCCGTG AGCATCAGTGGCGCAGTCGGAGCTCGGGAGGGCTCTGG AAAACAGCGGAGAGGCCCCCGAGCTCAAGCGCTTTCCGATCCAATC CCTTCAGTGTCAACACCCATGCTCGAGCTCCGCGCCGCTGGAGCT GCGCAGTGTGCGCACCTGAAAGTCATCGATGCGCCGCGCTCCCATCT GCGCAGTGTCTCCGGTCAACCCCGAGCATCTGGCGGCTTTAGTGGTGA GCTCTGGGGCGGGTCCCGCCCTGGCGGATAAAGCTCTCGCAGGGC AGTTTACGCCGAGCAAACGCCATCAATGCCCTTCTGGTTGTACA AAAGCATACTATAAGTCATCCCATCTCAAGAGTCACCAAGAGGACGCATAC AGGTGAGGACCTTTAGTGTGACTGGCTGATTTGCAAGAAAGAATTAC GCGGAGCGCAGCAACTTGCGCCGCACTACCGCACTCACACTGGAGAAAGA GGTTCTTGTCCCTGTGTCCTCAAGCAGTTCTACGCAGTGTACTTGAC AAAACATGCTAGGAGACATCCAACATACCATCCGACATGATAGAGTATC GAGGTAGGGAGCACACCTAGAATTGATCCTCGCTGACTAGTGAAGTC GAGTCAGTGCAGTGGAAAGCGGACCGGGTCCCGCGCCCTATTACAC CTGTCTT	25
KLF15	ATGGGGACCAACTACTTCAGTGGACGAGAACTTCGTCGCCAAATG CCAGTTGGGTATCTGGGTGATAGGCTGGTGGCGGGCATATCACATG CTGCCCTACCCGCTCTGAAGATGACAGCGATGCCCTCAGCCCTGCTCC TGTTCAGTCCCAGCTCTCAAGCCCTGCTCTGCTATGGTGGAGGCTG GGCACCGAGAGCAGCAGCTTGGACTTCTTATTGCTCCAGGCCAG CTGGCGAGTGGCGGGGAGCGGGAGCTGGGGCAGCAGTGGCC CGTGGCTGGGGGGCTGGCGAAGGGCAGGGCCCTGTGAAGGGGAGC ATTTCGCTTGGCCAGTTCCCTTGGGTGATCTGATGACGTCCTCACGCC CTTCCAGCCTACCTGGAGGAGATTGAAGAGTTCTGGAGGAGAACATGG AGCTGGAGTCAAGGAGGTCCTGAGGGCAACACCAAGGACTTGGATGCC TGCAGCCAGCTCTAGCTGGGCCACACAAGGCCACCTCCATCTGGTCC AGCGGGAGAGAGCGCTTGGCCCTTCCACAGGGTGGTGCCAGTGAGGAGG TGCCTGGGGCCAGGTGGGGCCACGCGCTGATGGCCCATCCAGTGT GCTGCAGATCCAGGGCGCTGTGAAGCAGGAATGGGCACAGGGCCTG CCTCCCTGGCAAGCCCAAGAGATGTCAGGGTTGCCCCAGTGGTACCCCTC ACATCCAGGGCAAGCCCACTCGCACTCGTCCCCAGTGGTACCCCTC ACTTGAACCTGCCCTCCAAGTTGTGCGCATTGCCCCCTGTGCCCATG CAAGCCTGTTGGATCGGGACCCCTGGGGCTGGCCCTGCGGCTCCCAT	26

- continued

GENE	SEQUENCE	SEQ ID NO:
	GGGCCAGAAGTCCCCAAGAACCCAGCCGAGAACATCATCAAAATGCACA AATGTACTTCTCCCTGGCTGCAGCAAGATGTACACAAAAGCAGGCCACCTCA AGGCCACACTGGCGCCGGCACACGGGTGAGAAGGCCCTGGCACCTGG CCAGGCTGGCGCTGGAGGTTCTCGCGCTTGACAGCTGTCGGCACAG GGCTCGCACTCAGGTGAAGCGTACAGTGTCTGTGCGAGAAGA AGTTCGCGGGAGCGACCCTCTCAAGCACATCAAGGTGCACCGCTTC CGCGAGCAGCCGCTCCGTGCGCTCGTGAAC	
KLF16	ATGTCAGCCGGTGGCTGCGTGGATTATTTGCAAGCAGATGTGCTGATG GCAATTTCATCCGGTGCAGTAGTTCATCGCGGAAGACAGGGTCTGAGGGT GCGGGGCTCGGGGGGGTTGGATGTTGGCGCAGGGAGGCC TTCTCGGAAACACTGGCCCTCTCTCCCGGGCAGGAGGCC GGTCTGGTGCAGCTGGCTCTCACCTGTTGGCAGCTCCATACTGGC TGACCTGCGAGGGGGCCAGGCGCTGCACCTGGTGGCGCAGTCCAGCAA GTTCCAGCTCGGGCGTCTCCCGAGTAGTGGCGAGCTCCGGCGGG CACCTCTGCGCTAAATCACACCGATGCCCTTCCCAGACTGCGCGA AGGGTATTATAAGTCCAGTCATTGAAATCACACTTGAGGACACATACCG GCGAGAGACCTTTCGCGACTGGCAGGGTTGATAAGAAATTGCG AGAACGACGAACATGGCCCATCACCGCACACAGGGGAAAAAA GATTCTCATGCCCACTCTGCTAAGCGTTACCGCGAAGCGACCATCTTG CAAAGCACCGTAGGAGACACCCCTGGGTTCCACCCGACCTCTGGCGACGA CCTGGCGCCCGGTCTACTAGCCGCTGACTCATTGCGTGCTCTCGCA GGTCCCTGCTCCGAGCCCGCACCGTCCCCAGCTCTGGCGGCTT	27
KLF17	ATGTACGGCGACCGCAGGCTGAGATGGAACAGGAGGCTGGGGAGCTGAG CCGGTGGCAGCGCGCACCAGGTGCCAGGATAACAGAGAACTCAGCGC CCATCTGAACATGCTTCTCATCTGGAAGCTGGAGTGCACACCTCTTG GAACCAAGGCTTACCAAGCATTGCACTTCTCACAGCGCAGAGATGCT GGGGTCCCCTTGTTGCTGTTGAGGCGCGGGCAGAATGTGAATGAAG GGGGCCACAGTTCAGTATGCCACTGCGTGGCTGAGGCTGAGTACTGCC CCCAAGCGACTCTACTCTTCCCGATGTTACTGTCAGAGAATGTC CCCCTCAGCAAGAGATGACGATTTCAGTGGGCCCAACTAATGCCCGTAG GAGAGCCCAATATTCAAGGGTACCGCAGGCCCTGGTGGAACTTAAGG ATGCCCTCAATGGGCTGGCAGTCTGGCTTCCACTGGAATCCAATAATG TCCACACTGGAAACCTTCCAGTGGCTTACCCCTGGCCTCTCGAGACT TCTGACGAAACATTGTTGGGCCACTGTGCTTCCACTGAGGCCCAGGCA GTGCTCCCTCCATGGCTCAGATGTTGCCCCGCAAGATGCCCATGACCTT GGATGCCCGCAGTGAGTCCAGTCATTGCTGGTTTAGGATCTCAGGAC TCTCTGTCAGTCAGCCAGACTCTAAGAAGGCCATTCTACCAAGAGCAG CCCGACACTGTCACAGCAGTAGAGAAGAACCTCAGGCCCTAGGAAGG GACTGTAGAAGGGCTCCCTCAGAGGCAAGGCCCTACTGCTGCAACTACG AGAACTGCCAAAAGCTTACAAACGCTCCACCTCGTGAGGCCACAG CGCAAGCACACAGGTGAGAGGCCATTCTTGCAACTGGAAAGTTGTC ATGAGCTTTCTCGTGTAGGCTTAGACGACATATGCGGGTACACAC CAGATATGACCATATAATGTGATCAGTGCAAGCCGGAGTTCATGAGGT CTGACCATCTCAACCAACACAGAGACTCATGGCCGGGACCTCAGAC CCACAGGCCAACACAATGGAGAGCAGGACAGTCCTCTGCTGCTGG TCCT	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. 3C), 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 ETV2-transduced 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 feeder-free 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 EF1α 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 polyadenylation 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 polyadenylation 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>O 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, 2x Gibson assembly master mix (New England Biolabs), H<sub>2</sub>O 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 µg 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

$\mu\text{M}$ , 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  $\mu\text{g}/\text{ml}$ , 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  $\mu\text{g}/\text{ml}$ , 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 1xPBS with 0.04% BSA at a concentration between 600-2000 per  $\mu\text{l}$ . Samples were loaded on the 10 $\times$  Chromium system and processed as per manufacturer's instructions (10 $\times$  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 $\times$  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  $\mu\text{l}$  reactions for each sample. 2  $\mu\text{l}$  of the cDNA was input per reaction with Kapa HiFi Hotstart ReadyMix (Kapa Biosystems). The PCR primers used were, NexteraI7\_TF\_Barcod\_F: GTCTCGTGGCTCGGAGATGTGTATAAGA-GACAGAGAACTATTCCTGGCTGTTACG 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  $\mu\text{l}$  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  $\mu\text{l}$  reactions for each sample. 2  $\mu\text{g}$  of genomic DNA was input per reaction with Kapa HiFi Hotstart ReadyMix. The PCR primers used were, NGS\_TF-Barcode\_F: ACACCTTTCCCTAACGACGCTCTCCGATCTAGAACTAT-TTCCTGGCTGTTACGCG (SEQ ID NO: 31) and NGS\_TF-Barcode\_R: GACTGGAGTTCAGACGTGTGCTCTCC-GATCTTGCTTCGTTGGAGTGAATTAGC (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  $\mu\text{l}$  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](https://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 glmmnet 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 was 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](https://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, 2x 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 µl 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 3.2×10<sup>5</sup> cells per well in a volume of 500 µl 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 Control	ATGGTGAGCAAGGGCGAGGAGGAT AACATGGCCATCATCAAGGAGTTC ATGCGCTTCAAGGTGCACATGGAG GGCTCCGTGAACGCCACGAGTTTC GAGATCGAGGGCAGGGCGAGGGC CGCCCCCTACGAGGGCACCCAGACC GCCAAGCTGAAGGTGACCAAGGGT GGCCCCCTGCCCTCGCCTGGGACA TCCTGTCCCCTAGTTCATGTACGG CTCCAAGGCCCTAGTGAAAGCACCC CGCCGACATCCCCGACTACTTGAG CTGTCCCTCCCCGAGGGCTTCAAGT GGGAGCGCGTGATGAACTTCGAGG ACGGCGCGTGGTGACCGTGACCC AGGACTCTCCCTGCAGGACGGCG AGTTCATCTACAAGGTGAAGCTGC GCGGCACCAACTTCCCCTCCGACGG CCCCGTAAATGCGAAGAAAGACCAT GGGCTGGGAGGCCTCCTCGAGCG GATGTACCCCGAGGACGGCGCCCT GAAGGGCGAGATCAAGCAGAGGCT GAAGCTGAAGGACGGCGGCCACTA CGACGCTGAGGTCAAGACCCACCTA	33	Non-functional control vector	

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CAAGGGCCAAGAAGGCCGTGCAGCT GCCCGGCCCTACAACGTCACAT CAAGTGGACATCACCTCCCACAAAC GAGGACTACACCACATCGTGGAACAG TACGAACGCGCCGAGGGCCGCCAC TCCACCGGCGGCATGGACGAGCTG TACAAG			
ASCL1	ATGGAGTCTTCTGCTAAAATGGAGT CCGGAGGCGCGGGACAACAACCC AACCGCAACCACAACAACCC GCCGCGGCCCATGTTTTCCG ACCGCTGCTGCTGCTCAGCGCG GCGGCTGCTGCCGCCGCAATCC GCCCAACAGCAACAACAACACAG CAGCAGCAGCAACAAGCGCTCAA CTTCGACCCGCTGCAGACGGGCAG CCCTCAGGGGGAGGGCACAAGAGC GCTCCGAAGCAGGTTAAAAGGCAG AGGAGCAGTAGTCCGAACTGATG CGATGTAAGAGGGCCCTCAATT GCGGTTTGGTTACTCTTGCCCCA GCAAGCAGCCGCTGCCGTAGCTCG CCGAAATGAGGGAAAGGAACCG CGTTAAACTTGTGAATCTCGTTTC GCGACACTTCGAGAGCACGTACCA AATGGGGCAGCTAACAAAGAAAATG AGTAAAGTTGAGACACTGCGTCT GCAGTGGAGTATTTAGAGCTTTC ACAATTGCTTGACGAGCACGATG CCGTATCAGCCGCATTCAAGCCGG GGTGCTGTCCCCAACAAATATCTCCG AACTACAGCAATGATCTTAATAGC ATGGCGGAAGTCCCTTCTCCT ACTCCTCTGATGAGGGCAGCTACG ACCCTCTCAGTCCGAGGAGCAAG AGCTTCTTGACTTCACTAACTGGTT C	34	Involved in neuronal specification and differentiation. Demonstrated to drive neuronal differentiation from hPSCs	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).
ASCL3	ATGATGGACAACAGAGGCAACTCT AGTCTACCTGACAAACTCCTATCT TCCCTGATTCTGCCGCTTGCCT TACCAAGGTCTTCTATCTGGAGCC ATGGTCACTTCCACGTGACCCAG AGGCCCGGGTGTATCTCTTACTC TGAGGAGCTGCCACGGCTGCCTTT CCCAGCGACTCTTATCCTGGAA ATTACAGTGAACCCCTGCCCTTCTC TTTCCCGATGCCCTATCCAATTAC AGAGGGTGCAGTACTCCTACGGG CCAGCCTCACCCGAAAGGAAT GAGCGGGAAAGGCAGCGGGTGA TGTGTCATGAGGGCTACGCCAG CTCCGACATCATCTGCCAGAGGAGT ATTGGAGAAGCGACTCAGCAAAG TGAAACCCCTCAGAGCTGCGATCA AGTACATTAACCTACCTGCAGTCTCT TCTGTACCCCTGATAAAAGCTGAGAC AAGAATAACCCCTGGAAAAGTTC TCCATGATAGCAACCCACCGCCAC CATGCTGACCCCTATGTTAGAAATTG TTGCCCAACTTCTTGTACAAAGT TGTCCCC	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)

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
ASCL4	ATGGAGACGCGTAAACCGGCGAA CGGCTGGCCTGCCATACTCGCTGC GCACCCGCCCTGGCGTTCCGG GGACCTGCCCCGACTCCCGGGA GGGACCCCCTCAGGGTCGCCCTGC GTCTGGACGCCGCGTGTGGAGT GGGCGCGCAGCGGCTGCGCACGGG GATGGCAGTACTTGCCCCGTGCCGCT GGACAGGCCCTTCGAGCCCCTTC CTCCGCAAGCGAACAGCGCGAG CGGCAGCGGGTGCCTGCGTGAAC GAGGGTATGGCGCTCCGAGAC CACCTGCCCCGGGAGCTGGCAGAC AAGCGCTCAGCAAAGTGGAGACG CTCCGCGCTGCCATCGACTACATCA AGCACCTGCAGGAGCTGCTGGAGC GCCAGGCCTGGGGCTCGAGGGCG CGGCCGGCGCGTCCCCCAGCGCA GGGCGGAATGCAACAGCGACGGGG AGTCCAAGGCCTTCGGCGCTTC GCCCAGCAGCGAGGCCGAGGAGGG GGGCAGC	36	Involved in development of skin	Jonsson, M. et al. Hash4, a novel human achaete-scute homologue found in fetal skin. <i>Genomics</i> 84, 859-866 (2004)
ASCL5	ATGCCGATGGGGCAGCAGAAAAGA GGTGCTGGCCCCAATCATCTGCAG CACCATGGCTGGTTCAGAAAAAGG CGGCAAAGAGAGGGCCATAAAAAA GCTGGTACCCAAGAGCTGCTGCATC TGATGTCACGTGCCGACTGGTGGT GATGGAGCTGACCCAAAACCTGGA CCTTTGGAGGGTTAGCTTAG GGCCTGCGCCAGAGGAACAATGA ATAATAATTCTGCAGGGCCCTTGT TGACAGAAGGCCCTTAGGACCCCT TCATGTATGCAATTAGGTGTAATGC CACCGCCAAGACAAGCGCCCTCC CGCCGGCTGAACCCCTGGAAATGT ACCTTCTCCTATAACCTGGCCA GCTGAACCACCATATTATGATGCAT ATGCTGGTGTTCCTCATATGTGCC TTCCCTGGTGTCTTGGTGTATAT GAATACCTTGTGAGCCGGCTTTA TCCAAAAGAGGAATGAAAGAGAGA GACAGAGAGTGAAGTGTGTGAATG AAGGATACGCCAGATTGAGAGGCC ATTGCTGGTGCCCTGGCAGAAAAA GAGATTATCAAAGTTGAAACCT GAGGGCGCAATCAGATATAAAA ATACCTCCAAGAACTCTTTCATCA GCACCTGATGGATCGACACCAACCG GCTTCAAGAGGTTACCTGGAACTG GACCATGCCCTGCACCGCCTGCTAC ACCAAGGCCAGACAGACCTGGAGA TGGAGAAGCAAGAGCACCTCTC CCTTGTCCCTGAATCTCTGAATCA TCATGTTTTCGCCCTCCCCTTTTT AGAAAATGAAGAATCCTGGCA	37	Paralog of ASCL4	Wang, C. et al. Systematic analysis of the achaete-scute complex-like gene signature in clinical cancer patients. <i>Molecular and Clinical Oncology</i> 6, (Spandidos Publications, 2017).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
ATF7	ATGGGAGACGACAGACCGTTTG TGCATGCCCGGGCTGTGGACAG AGATTTACAAACGAGGACCACTG GCAGTTCATAAACACAAGCATGAG ATGACATTGAAATTGGCCAGCCC GAACTGACTCAGTCATTCAGA TCAAACGCCCTACTCCAACTAGATTC CTGAAGAACTGTGAGGAGGTGGGA CTCTTCATGAACCTAGCTAGCTCCT TTGAACATGAATTCAAGAAAGCTG CAGATGAGGATGAGAAAAAGGCAA GAAGCAGGACTGTGCCCCAAC TGGTGGCTGCTGGCCCTTGA CATGTCTCTGCCCTCACACCAGAC ATCAAATCAAAGAAGAAGGCCA GTGGAGGTAGACTCATCCCCACCTG ATAGCCTGCCCTAGTCCCTGTT CCCACCACTGAAGGAGAAGGAGGT TACCCCCAACCCACATTGTACGTC CTGGCCTCCCTGCTCTCCACTGGG CTATGATCCACTCATCCAACCTT CCCTCCCCAACCTCTGTCATCACAC AGGCTCCACCATCCAACAGGCAA TGGGGTCTCCCACTGGCTCCCTCC TCTTGTATGCATCTTGCTAATGGA CAGACCATGCCCTGTGTTGCCAGGG CTCCAGTACAGATGCCGCTGTAT ATCGCTGGCCAGACCTGTGTCATG GTGCCAACATTCTGGTATCCCTG GCCCACCAAGTTAACAGTAGTGCTC CATTCTCCCTGGCACCCCTATA CCATCAGAAGCCAAGATGAGACTG AAAGCCACCCCTAACTCACCAAGTCT CCTCAATCAATGGTGGTTGTGGAAT GGTGGTGGTACTGCCAGCACCAT GGTGACAGCCCCCAGAGCAGAG CCAGATTCTCATCCAGCACCCTGAT GCCCATCCCCCTGCCAGCCACAG GTCTCACCAGCTCAGCCCACCCCTA GTACTGGGGCGACGGCGGCGA CAGTAGATGAAGATCCAGATGAGC GACGGCAGCGCTTCTGGAGCGCA ACCGGGCTGCAGCCTCCGCTGCCG CCAAAAGCGAAAGCTGTGGGTGTC CTCCCTAGAGAAGAAGGCCAGAGA ACTCACTTCTCAGAACATTCACTG AGTAATGAAGTCACTTACTACGC AATGAGGTGGCCAGTTGAAACAG CTACTGTTAGCTCATAAGACTGCC CAGTCACTGCACTACAGAAAAAGA CTCAAGGCTATTAGAAAGCCCA AGGAAAGCTCAGAGCCAACGGGTT CTCCAGCCCCCTGTGATTCACTG CTCAGCAACAGCCCCCTAGCAATGG CCTCAGTGTGCTCTGCAGCTGAA GCTGTGGCCACCTCGGTCTCACTC AGATGGCCAGCCAAAGGACAGAAC TGAGCATGCCGATACAAATCGCATGT AATCATGACCCACAGTCCCAGTCT GCGGGCAGA	38	Involved in early cell signaling, binds cAMP response element	Peters, C. S. et al. ATF-7, a novel bZIP protein, interacts with the PRL-1 protein-tyrosine phosphatase. J. Biol. Chem. 276, 13718-26 (2001). Hamard, P.-J. et al. A functional interaction between ATF7 and TAF12 that is modulated by TAF4. Oncogene 24, 3472-3483 (2005).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
CDX2	ATGTACGTGAGCTACCTCCTGGACA AGGACGTGACCATGTACCCCTAGCT CCGTGGGCCACTCTGGCGGCTCAA CCTGGCGCCAGAACTTCGTCAGC CCCCCGCAGTACCCGGACTACGGC GGTTACACGTGGCGCCGCAGCT GCAGCGGCAGCGAACATTGGACAGC GCGCAGTCCCCGGGCCATCTGG CCGGCAGCGTATGGCGCCCACTCC GGGAGGACTGGAATGGCTACGCGC CCGGAGGGCGCCCGGGCCCGCA ACGCCGTGGCTCACGGCTAACG GTGGCTCCCCGCCCGCAGCCATGG GCTACAGCAGCCCCGAGACTACC ATCCGACCACCAACCCGACATACC ACCCGACCACCCGGCCGCCGCGC CTTCCTGCGTTCTGGCTGTGCA AACGCTCAACCCGGGCCCTCTGGG CCCGCCGCCACCGCTGCCCGAG CAGCTGTCTCCGGCGGCCAGCGG CGGAACCTGTGGAGTGGATGGG AAGCCGGCGCAGCAGTCCCCTCGGC ACCCAAGTAAAACCAAGGACGAAA GACAAATATCGAGTGGTGTACACG GACCACAGCGGCTGGAGGCTGGAG AAGGAGTTTCACTACAGTCGCTACA TCACCATCGGAGGAAAGCCGAGC TAGCCGCCACGCTGGGCTCTCTGA GAGGCAGGTTAAAATCTGGTTCA GAACCCGAGAGCAAAGGAGAGGA AAATCAACAAGAAGAAGTTGCA AGCAACAGCAGCAGGCCAACAC AGCGGCCCTCGCCGCCACACAGC CTCCCCAGCCTCAGCCAGGTCTCT GAGAAGTGTCCAGAGCCCTTGAG TCCGGTGTCTTCCCTGCAAGCTCA GTGTCTGGCTCTGTCCCCGGGTTCA TGGGGCCAACTGGGGGGGTGCTAA ACCCCAACCGTACCCAG	39	Involved in trophectoderm specification and differentiation	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).
CRX	ATGATGGCGTATATGAACCCGGGG CCCCACTATTCTGTAACGCTTGG CCTTAAGTGGCCCAAGCTGGATCT GATGCACCAAGGCTGTGCCCTACCA AGCGCCCCAGGAAGCAGCGGGGG GAGCGCACCAACCTCACCCGGAGC CAACTGGAGGAGCTGGAGGCACTG TTTGCAAGACCCAGTACCCAGAC GTCTATGCCGTGAGGAGGTGGCTC TGAAGATCAATCTGCTGTGAGTCAG GGTCAGGTTGGTCAAGAACCGG AGGGCTAAATGCAAGGAGCAGCGA CAGCAGCAGAACAGCAGCAGCAG CCCCCAGGGGCCAGGCCAAGGCC CGGCCTGCCAAGAGGAAGGGGGC ACGTCCCCAAGACCCCTCACAGAT GTGTGTCAGACCCCTCTGGCATCT CAGATTCTACAGTCCCCCTCTGCC CGGCCCTCAGGCTCCCCAACAC GGCAGTGGCACTGTGCTATCTGG AGCCCAGCCTAGAGTCCCCTTGC CTGAGGGCGAGGGGCTGGGCTGG TGGCCTCAGGGCGTCTCTGACCTC CGCCCCCTATGCCATGACCTAGGCC CGGGCCTCGCTTCTGCTTCCC CCTCCGCCTATGGGTCTCGAGCTC CTATTCAGCGGCCCTAGACCCCTAC CTTCTCCCATGGTGCCCCAGCTAG GGGGCCGGCTCTTAGCCCCCTCTC TGGCCCCCTCCGTGGGACCTCCCTG GCCAGAGCTATGGCGCCTACAGCC CCGTGGATAGCTGGAATTCAAGG	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).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	ACCCCACGGGCACCTGGAAATTCA CTTACAATCCCATGGACCCCTGGGA CTTACAAGGATCAGAGTCGGCTGGAA GTTTCAGATCTTG			
ERG	ATGCCAGCACTATTAAGGAAGCC TTATCAGTTGTGAGTGAGGACCGT CGTTGTTGAGTGTGCCTACGGAAC GCCACACCTGGCTAAGACAGAGAT GACCGCGTCCTCCCTCAGCGACTAT GGACAGACTTCCAAGATGAGCCA CGCGTCCCTCAGCAGGATTGGCTGT CTCAACCCCAAGCCAGGGTCACCAT CAAATGGAATGTAACCCTAGCCA GGTGAATGGCTAAGGAACTCCCT GATGAATGCAGTGTGGCAAAGGC GGGAAGATGGTGGGCAGGCCAGAC ACCGTTGGGATGAACTACGGCAGC TACATGGAGGAGAAGCACATGCCA CCCCAAACATGACCACGAAACGAG CGCAGAGTTATCGGCCAGAGAT CCTACGCTATGGAGTACAGACCAT GTGCGGCAGTGGCTGGAGTGGCG GTGAAAGAATATGGCCTTCCAGAC GTCACACATCTTGTATTCCAGAAC TCGATGGGAAGGAACTGTGCAAGA TGACCAAGGACGACTTCCAGAGGC TCACCCCCAGCTACAATGCCGACAT CCTCTCTCACATCTCCACTACCTC AGAGAGACTCCTCTTCCACATTGA CTTCAGATGATGTTGATAAAGCTT ACAAAACCTCCACGGTTAATGCAT GCTAGAAACACAGGGGGTGCAGCT TTTATTTCCCAAATACTTCAGTAT ATCCTGAAGCTACGCAAGGATTA CAACTAGGCCAGATTACCATATGA GCCCCCCAGGAGATCAGCCTGGAC CGGTACGGCCACCCACGCCCA GTCGAAAGCTGCTCAACCATCTCCT TCCACAGTGCCAAAAGACTGAAGAC CAGCGTCCTCAGTTAGATCCTTATC AGATTCTTGACCAACAAGTAGCC GCCTTGCAAATCCAGGCAGTGGCC AGATCCAGCTTGGCAGTTCTCCCT GGAGCTCTGTCGGCAGGCTCCAA CTCCAGCTGCATCACCTGGGAAGG CACCAACGGGGAGTTCAAGATGAC GGATCCCAGCAGGGTGGCCCGCG CTGGGGAGAGCGGAAGAGCAACC CAACATGAACTACGATAAGCTCAG CCGGCCCTCCGTTACTACTATGAC AAGAACATCATGACCAAGGTCCAT GGGAAGCGCTACGCCCTACAACTTC GACTTCCACGGGATGCCCAAGGCC CTCCAGCCCCACCCCGGAGTCAT CTCTGTACAAGTACCCCTCAGACCT CCCGTACATGGCTCTATCAGGCC CACCCACAGAAGATGAACTTTGTG GGGCCCCACCCCTCCAGCCCTCCCG TGACATCTTCCAGTTTTTGCTGCC CCAAACCCATACTGAAATTACCA ACTGGGGGTATATACCCCAACACT AGGCTCCCCACCAAGCCATATGCCCT CTCATCTGGCAGTTACTAC	41	Involved in endothelial cell specification and differentiation	McLaughlin, F. et al. Combined genomic and antisense analysis reveals that the transcription factor Erg is implicated in endothelial cell differentiation. Blood 98, 3332-3339 (2001).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
ESRRG	ATGTCAAACAAAGATCGACACATT GATTCCAGCTGTCGCTTCATCA AGACGGAACCTTCCAGCCCAGCCT CCCTGACGGACAGCGTCAACCACC ACAGCAGCTGGTGGCTCTTGACAGCG CAGTGGGAGCTACAGTTCAACCAT GAATGCCATCAGAACGGACTTGA CTCGCCACCTCTCACCTTCTGCT CCTATCTGGGAGGTAGTGGGCCTG TCAGGAAAATGTATGATGACTGCTC CAGCACCAATTGTGAAGATCCCCAG ACCAAGTGTGAATACATGCTCAACT CGATGCCAAGAGACTGTGTTAGT GTGTTGGTGAATCGCTTCTGGGTAC CACTATGGGTAGCATGTGAA GCCTGCAAGGCATTCTCAAGAGG ACAATTCAAGGCATATAGAATAC AGCTGCCCTGCCACGAATGAATGT GAATCACAAAGCGCAGACGTAAG TCTGCCAGGCTGCCGCTTCAATG AGTGTAAAAGTGGCATGCTGA AAGAAGGGGTGCGCTTGAACAGAG TACGTGAGGTGGCAGAAGTACA AGCGCAGGATAGATGCCGAGAAC GCCCATACCTGAACCCCTCAGCTGGT TCAGCCAGCAGAAAAAGCCATTGCT CTGGTCTGATCTGCAAGATAACAAAG ATTGTCACATTGTTGGTGGCTG AACCGGAGAAGATCTATGCCATGC CTGACCCCTACTGTCCCCGACAGTGA CATCAAAGCCCTCACTACACTGTGT GACTTGGCCGACCGAGAGTTGGTG GTATCATGGATGGCGAAGCAT ATTCAGGCTTCTCCACAGCTGCTCC TGGCGGACCGAGATGAGCCTTCTGC AGAGTCTTGGATGAAATTGGAT CCTTGGTGTGTATACCGGTCTCTT TCGTTTGAGGATGAACTTGTCTATG CAGACGATTATAATCGACAGAAG ACCAAGTCAAATTAGCAGGCTTCT TGATCTAAATAATGCTATCCTGAG CTGGTAAAGAAATACAAGAGCATG AAGCTGGAAAAAGAAGAATTGTC ACCCCTAAAGCTATAGCTCTGCTA ATTCAAGACTCCATGCACATAGAAG ATGTTGAAGCCGTTCAAGAGCTCA GGATGTCTTACATGAAGCGCTGCA GGATTATGAAGCTGGCCAGCACAT GGAAGACCCCTCGCAGCTGGCAA GATGCTGATGACACTGCCACTCTG AGGCAGACCTCTACCAAGGGCGTG CAGCATTCTACAAACATCAAACTAG AAGGCAAAGTCCAATGCACAAAC TTTTTTGGAAATGTTGGAGGCCAA GGTC	42	Involved in cardiac development	Alaynick, W. A. et al. <i>ERRγ Directs and Maintains the Transition to Oxidative Metabolism in the Postnatal Heart.</i> Cell Metab. 6, 13-24 (2007).
ETV2	ATGGATCTTGGAACTGGGATGAA GCTTCCCTCAAGAAGTCCCG GAAATAAAACTCGCGGGTGGAA GACTCCCTCGCCTTCCGCAACCGT CTGGGGCGGATGCCCTGGTGGAGC CTCAGCGGACCCAAACCCCTTGCT CCAGCGGAGGGGCAAAGTGGGT TTCGCTTCCCGATCTGCTTTC AAGGCATGATACTCCAACGGCAGCG CAGAGACCTGTGGAAAGGCACCA GTAGCTCCCTGGCCAGCTTCCGCA GTCGATTGGGGTCAAGCCCTTCTC CATCCGAAGTCCCTGGGGGGCG GAACCCGACTCCAAGCCCTCCCT GGAGTGGTGAATTGGACAGATATGG CATGCACAGCCTGGGACAGTTGGT CCGGGGCGTCAAGACATTGGGAC	43	Involved in haemato-endothelial specification and differentiation, and in vasculogenesis	Lee, D. et al. ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. Cell Stem Cell 2, 49-507 (2008).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CAGCCCCACTTGGACCGGGGCTAT CCCCGCAGCAGGAAGCGAAGGGAGC TGCTGGTCAGAACCTGTGTGCCCTG GCTGGTGAGGCCTACCAAGTTGCCA GGGCCAGGCAGCAGCAGTAACA CCAGCTGGGATTGGCTCAGTGGGGC CTGACGGGATACTTATTGGGCTC TGTTCTTGGTGGAGAACCGAGAAC GGACTGTACGATAAGTTGGGGCG TCCAGCTGGGCTGATTGTAATACG TCATGGAATCTGGCTTGCACGCCG GGGGCACGACAAGCTTAAGAGAT ATCAAAGTTCAGCCCTACAGTTG CTCAGAACCTTCCCCCAAAGTGAC CGAGCGTCACTGGCGCGATGTCCTA AAACTAATCATCGAGGGCCATCC AGTTGTTGCAGTTTGCTTGAECT CCTTCACGATGGCGCGAGGAGCAG TTGGCATCAGATGGACCGGTAACAG CAGGGAGTTCCAATTGTTGACCCCC AAGGAAGTGGCTCGACTGTGGGGT GAGCGCAAACCGGAAGGCTGGTATG AATTACGAAAAGTTGAGTAGGGGT TTGGCATATTACTATAGGCGGACA TCGTTGAAAGTCGGTGGTCGAA AGTACACATACAGATTGGCGGTC GCGTACCATCTCTGCATACCTGAA TTGGCGCAGGGGGGGTAGGGGTGC GGAAACACAA			
FLI1	ATGGACGGGACTATTAGGAGGCT CTGTCGGTGGTGAGCGACGACCAG TCCCTTTGACTCAGCGTACGGAG CGGCAGCCCACATCCCCAAGGCCG ACATGACTGGCTGGGGAGTCCTG ACTACGGGCAGCCCCAACAGATCA ACCCCTCCACACAGCAGGAGT GGATCAATCAGCAGTGAGGGTCA ACGTCAGCGGGAGTATGACCA TGAAATGGATCAGGGAGTCTCCGG TGGACTGCAGCGTTAGCAAATGCA GCAAGCTGGTGGCGGGAGGT CCAAACCCATGAACACTACAACAGCT ATATGGACGAGAAGAATGGCCCCC CTCCTCCCAACATGACCCACACGA GAGGAGAGTCATCGTCCCCCAGA CCCCACACTGTGGACACAGGAGCA TGTGAGGCAATGGCTGGAGTGGGC CATAAAGGAGTACAGCTTGTGGA GATGACACATCCTTTCCAGAAC ATGGATGGCAAGGAACTGTGTA ATGAACAAGGGAGGACTTCCCTCCGC GCCACACCCCTCTACAAACACCGAA GTGCTTGTCAACACCTCAGTTAC TCAGGGAAAGTTCACTGCTGCCCTA TAATACAAACCTCCACACCGACA ATCCCTCACGATTGAGTGTCAAAGA AGACCCCTCTTATGACTCAGTCAGA AGAGGGAGCTTGGGGCAATAACATG AATTCTGGGCTCAACAAAAGTCTC CCCTTGGAGGGCACAACAGATCA GTAAGAATACAGAGCAACGGCCC AGCCAGATCCGTATCAGATCTGG GCCCGACCAGCAGTCGCCCTAGCCA ACCCCTGGAAAGCAGGGCAGATCCAGC TGTGGCAATTCTCTGGAGCTGCT CTCCGACAGCGCCAACGCCAGCTG TATCACCTGGGAGGGGACCAACGG GGAGTTCAAATGACGGACCCCGA TGAGGTGGCCAGGGCTGGGGCGA GGGGAAAAGCAAGCCAACATGAA TTACGACAAGCTGAGCCGGGCCCT CCGTTATTACTATGATAAAAACATT ATGACCAAAAGTGCACGGCAAAAGA	44	Involved in haemato-endothelial specification and differentiation	Liu, F. et al. <i>Fli1 Acts at the Top of the Transcriptional Network Driving Blood and Endothelial Development.</i> Curr. Biol. 18, 1234-1240 (2008).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TATGCTTACAAATTGACTTCCACG GCATTGCCAGGCTCTGCAGCCACA TCGGACCGAGTCGTCCATGTACAAG TACCCCTCTGACATCTCTACATGC CTTCCTACCATGCCAACAGCAGAA GGTGAACTTGTCCTCCATCCA TCCTCCATGCCTGTCACTTCTCCA GCTTCTTGAGCCGATCACATA CTGGACCTCCCCAACGGGGGAAT CTACCCCAACCCCAACGTCCCCCGC CATCCTAACACCCACGTGCCTTCAC ACTTAGGCAGCTACTAC			
FOXA1	ATGTTGGGCACCGTGAGATGGAG GGGCATGAGACAAGCGACTGGAAT TCCCTACTACCGGATAACCCAAGAA GCGTATTCTTCAGTTCCGTAAGCA ATATGAACTCCGGATTGGGAGCA TGAATAGTATGAACACGTATATGA CAATGAATACGATGACCACCGCG GCAACATGACACCGGCCCTTTAA TATGTCATATGCGAACCTGTTCTT GGCGCTGGCCTCTCACAGGTGCG GTCGCTGGAATGCCCGGGGGAGC GCCGGAGCGATGAACTCCATGACC GCTGCGGGCGTGACGGCATGGGT ACGGCCCTGTACCCAGTGGAAATG GGAGCTATGGGGCCAGCAAGCC GCTCAATGAATGGATTGGGCCCT ATGCCCGGGCATGAAATCCCTGCAT GTCCTATGGCTTATGCCCTCAGC AATTGGGTGGCAGTAGAGCGGGC GGTGGTGGCAGTGCCTTACCTTC AAGCGAAGTTATCCTCATGCGAAG CCTCCTTATTCTATATATCTTGAT TACGATGGCGATACAGCAGGCC GCTCTAAGATGCTGACTCTGAGTGAG ATATACCAAGTGGATCATGGACCTT TTCCTTACTACGGCAAACCAACA GAGATGGCAAACACTAAACGCCA TAGCCTTCCTCAATGATTGCTT GTCAAAGTCGCTCGGAGCCCTGAC AAGCCGGTAAGGGTCTTATTGG ACCCCTCATCCAGATAGCGGCAATA TGGTCGAGAATGGTTTATCTTAG ACGGCAGAAACGATTCAAATGTGA GAAACAGCCAGTGCCGGCGTGG TGGCGGCAGCGGTTCAAGCGGAAG TGGTGCCAAGGGTGGGCTGAGTC TAGAAAAGACCCCAAGGGAGCAAG CAATCCAAGCGGGACTCTCCCTG CACCGCGGTGTTCATGGTAAGACA GGTCAGCTTGGGGCGCTGCT CCAGGCCGGTGGCATCCAGC ACACTGGACCATAGTGGAGCTACA GCGACCGGAGGTGCTCAGAAC AAGACGCCCTGCGTCCCTCACTGCG CTCCGATCTCAGTGGTCCCGGTG ACTTGCCCTGTTCTGCATCTCAT CCAGCACACGGACTCGCGCCAC GAGTCCCAGCTCATTGAAAGGG GACCCACACTACAGCTTAAACCACC CATTCTCTATTAAACAATTGATGTC ATCCTCAGAACAGCAGCATAAACT CGACTTCAAAGGCTATGAACAGGC	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).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CCTGCAGTATTCTCCATATGGCTCT ACACTTCCTGCTCTCTTCCATTGG GGTCTGCAAGTGTGACAACGGCT CCCCAATCGAGCCAAGTGCCCTCG AGCCTGCTTATTATCAAGGAGTATA TTCCCGACCAGTTTGAAATACAAGT			
FOXA2	ATGCTGGGAGCGGTGAAGATGGAA GGGCACGAGCCGTCCGACTGGAGC AGCTACTATGCAGAGCCCGAGGGC TACTCCTCCGTGAGCACATGAACG CCGGCTGGGGATGAAACGGCATGA ACACGTACATGAGCATGTCGGGG CCGCCATGGGCAGCGCTGGGCA ACATGAGCGCGGCTCCATGAACA TGTGTCGTACGTGGCGCTGCGAT GAGCCCGTCCCTGGCGGGATGTC CCCCGGCGGGCGCCATGGCGGG CATGGGCGGCTGGCGGGGGGG TGGCGTGGCGGGCATGGGGCGCA CTTGACTCCCAGCCTGAGCCCGCTC GGGGGGCAGGGCGCCGGGGCATG GGCGGCCTGGCCCCCTACGCCAAC ATGAACCTCATGAGCCCCATGTACG GGCAGGGCGGCCTGAGCCGCGCC GCGACCCCAAGAACCTACAGGGCA GCTACACGCAACGCAAAGCCGCC ACTCGTACATCTGCTCATCACCAT GCCCATCCAGCAGAGCCCAACAA GATGCTGACGCTGAGCGAGATCTA CCAGTGGATCATGGACCTCTTCCCC TTCTACGGCAGAACCCAGCAGCG TGGCAGAACTCCATCCGCCACTCGC TTCCTTCAACGACTGTTCTGAA GGTGGCCCGCTGGCCACAGCC CGGCAAGGGCTCTCTGGACCTG CACCTGACTCGGGCAACATGTCTG AGAACCGCTGCTACCTGCGCCGCC AGAAGCGCTTCAAGTGCAGAGAAGC AGCTGGCGCTGAAGGAGGCCAG GGGCCGCCGGCAGCGGAAGAAGG CGGCCGCCGGGGCCAGGCCCTCAC AGGCTCAACTCGGGAGGCCCG GCCCGCCCTCCGAGACTCCGGGG GCACCGAGTCGCCACTCGAGCG CCTCCCCGTGCCAGGAGCACAAGC GAGGGGGCTGGGAGAGCTGAAGG GGACGCCGGCTGCCAGCGCTGAGCC CCCCAGAGCCGGCGCCCTCTCCG GGCAGCAGCAGCAGGCCGCC ACCTGCTGGGCCGCCACCAACCC GGGCCTGCCGCCCTGAGGCCACCT GAAGCCGAACACCACTACGCC CAACCAACCGTCTCCATCAACAA CTCATGTCCTGGAGCAGCAGCAC ACCACAGCCACCAACCAACCA CCCACAAAATGGACCTCAAGGC AGAACAGGTGATGCACTACCCG GCTACCGTTCCCCCATGCCCTGGCAG CTTGGCCATGGGCCGGTCAAGGAA CAAAACGGGCTGGACGCCCTGCC CCTGGCCCCAGATACTCCTACTAC CAGGGGGTGTACTCCGGCCCATTA TGAACCTCTTTG	46	Involved in branching morphogenesis, development of notochord, lung, liver, prostate, and pancreas.	Friedman, J. R. et al. The Foxa family of transcription factors in development and metabolism. <i>Cell. Mol. Life Sci.</i> 63, 2317-2328 (2006).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
FOXA3	ATGCTGGGCTCAGTGAAGATGGAG GCCCATGACCTGGCCGACTGGAGC TACTACCCGGAGGCAGGGCGAGGTC TACTCGCCGGTGACCCCAGTGCCCA CCATGGCCCCCTCAACTCTACAT GACCCTGAATCTCTAAGCTCTCCC TATCCCCCTGGGGGTCCCTGGCT CCCCACTGCCCTCAGGACCCCTGGC ACCCCAAGCAGCACCTGAGCCCCCTG GGGCCACTTTCCAGGCCTGGGTG TCAGCGGTGGCAGCAGCTCCG GGTACGGGGCCCCGGGTCTGGGC TGGTGCACGGGAAGGGAGATGCCGA AGGGGTATCGGGGCCCCCTGGCAC ACGCCAAGCCACCGTATTCTATAT CTCACTCATCACCATGGCCATCCAG CAGGGCCGGGCAAGATGCTGACC TTGAGTGAAATCTACCAAGTGGATCA TGGACCTCTTCCCTACTACGGGA GAATCAGCAGCGCTGGCAGAAC CATTGCCACTCGCTGTCTTCAC GACTGTTCTGTAAGGTGGCGCGTT CCCCAGACAAGCTGGCAAGGGCT CCTACTGGGCTTACACCCCCAGCTC AGGGAACATGTTGAGAATGGCTG CTACCTGCGCCGCCAGAAAAGCTTC AAGCTGGAGGAGAAGGTGAAAAAA GGGGGAGCGGGGCTGCCACCA ACCAGGAACGGGACAGGGTCTGCT GCCTCGACCAACCCCGCGGCC ACAGTCACCTCCCCGCCCCAGCCCC CGCCTCCAGCCCCGAGCCTGAGGC CCAGGGCGGGGAAGATGTGGGGC TCTGGACTGTGCGTCAACCGCTTCC TCCACACCCCTATTCTACTGGCTGG AGCTCCCAGGGGAGCTGAAGCTGG ACGCGCCCTACAACCTAACCA TTCTCCATCAACAACCTAATGTCA GAACAGACACCAAGCACCTCCAAA CTGGACGTGGGGTTGGGGCTAC GGGGCTGAAGGTGGGGAGCCTGGA GCTCTACTACCAAGGGCTCTATTCCC GCTCTTGCTTAATGCATCC	47	Involved in cell glucose homeostasis	Friedman, J. R. et al. The Foxa family of transcription factors in development and metabolism. <i>Cell. Mol. Life Sci.</i> 63, 2317-2328 (2006).
FOXP1	ATGATGCAAGAACCTGGGACTGAG ACAAAAGTAACGGGTTCAAGCCATC CAGATGGGTGCGGCCAGCAAC CACTTACTAGAGTGCAGCGGTTCTC GGGAGGGGCGGTCCAACGGAGAGA CGCCGGCCCGTGACATCGGGGCAG CTGACCTCGCCACGGCCAGCAGC AGCAGCAACAGTGGCATCTCATAA ACCATCAGGCCCTTAGGAGTCCCAG CAGTTGGCTTAAGAGACTAATTCA AGCCCTTGGGAGTTGGAAGTCTG AGGTCCCCCTTGTGGGGAGCAGTTG TGAGACGAAGATGAGTGGACCTGT GTGTCAGCCTAACCTCCCCATT	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. <i>Nat. Immunol.</i> 7, 819-826 (2006). Shu, W. et al. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. <i>Development</i> 134, 1991-2000 (2007). Bacon, C. et al. Brain-specific Foxp1 deletion impairs neuronal

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
GATA1	ATGGAGTCCCTGGCCTGGGGTCCC TGGGGACCTCAGAGCCCCCTCCCCA GTTTGATGGATCTGCTCTGGTGTCC TCCACACCAGAACATCAGGGGTTTCT TCCCCTCTGGGCCTGAGGGCTTGGG TGCAGCAGCTTCTCCACTGCCCG AGCACAGCCACCGCTGCAGCTGCG GCACTGGCCTACTACAGGGACGCT GAGGCTTACAGACACTCCCCAGTCT TTTCAAGGTGACCCATTGCTCAACTG TATGGAGGGATCCCAAGGGGCTC ACCATATGCCGGCTGGGCCTACGG CAAGACGGGGCTAACCTGCTCA ACTGTGTGTCACCCGGCGAGGACT CTCCCTCCCCAGGCCGTGAAGATCT GGATGAAAAGGCAGCACCGACTT CCTGGAGACTTTGAAGACAGAGCG GCTGAGCCCAGACCTCTGACCCCTG GGACCTGCACTGCCCTCATCACTCC CTGTCCCCAATAGTGTCTATGGGG CCCTGACTTTCCAGTACCTCTTT CTCCCAACGGGAGGCCCTCAATT AGCAGCCTATTCTCTCCAAAGCTT CGTGGAACTCTCCCCCTGCTCCCT GTGAGGCAGGGAGTGTGTGAACCT GCGGAGCAACAGCCACTCCACTGT GGCGGAGGGACAGGACAGGGCACT ACCTATGCAACGCCCTGGGCTCTA TCACAAGATGAATGGCAGAACAG GCCCTCATCCGGCCAAGAACGCG CCTGATTGTCACTAACGGGCAAG TACTCAGTGCACCAACTGCCAGAC GACCACACGACACTGTGGCGGAG AAATGCCAGTGGGGATCCCGTGTG CAATGCTGCGCCCTACTACAAG CTACACCACAGCACTACTGTGGT GCTCCGCTCAGCTCATGAGGGAC AGAGCATGGCTCCAGAGGAGGG TGGTGTCTCTCCCTTGTAGCCA GAATTCTGGACAACCCAAAGTCTG GGCCCCAGGCACCCCTGGCT	49	Involved in erythroid development	Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. <i>PNAS</i> 93, 12355-12358 (1996).
GATA2	ATGGAGGTGGCCGGAGCAGCCG CGCTGGATGGCGCACCCGGCGTG CTGAATGCGCAGCACCCGACTCA CACCAACCCGGGCTGGCGCACAC TACATGGAACCCGGCGAGCTGCTG CCTCCAGACGAGGTGGACGTCTTCT TCAATCACCTCGACTCGCAGGGCA ACCCCTACTATGCCAACCCCGCTCA CGCGCGGGCGCGCGTCTCTACAG CCCCGGCACGCCCGCCTGACGGG AGGCCAGATGTGGCCGCCACACTT GTGACAGCCGGGTTTGGCTTGG CTGGACGGGGGCAAAGCAGCCCTC TCTGCCGCTGCCGCCACCCACA ACCCCTGGACCGTGAGCCCCCTCTC CAAGACGCCACTGCACCCCTCAGCT GCTGGAGGGCCCTGGAGGGCCACTC TCTGTGTACCCAGGGCTGGGGT GGGAGGGGGGGAGGCAGCGGGAG CTCAGTGGCCTCCCTACCCCTACA GCAACCCACTCTGGCTCCCACCTT TCCGCTTCCCACCCACGCCACCAA AGAAGTGTCTCTGACCCCTAGCACC	50	Involved in haematopoietic development	Pimanda, J. E. et al. Gata2, Flil, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 104, 17692-7 (2007). Lugus, J. J. et al. GATA2 functions at multiple steps in hemangioblast development

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	ACGGGGGCTGCGTCTCCAGCCTCAT CTTCGGGGGGGTAGTGCAGCCC GAGGAGGGACAAGGACGGCTCA AGTACCAAGGTGTCACTGACGGAGA GCATGAAGATGGAAAGTGGCAGTC CCCTGGGCCAGGCCAGCTACTAT GGGCACCCAGCTGCTACACCCA CCCCATCCCCACTAACCCCTCTTAT GTGCCGGCGCTGCCACGACTAC AGCAGGGACTCTTCCACCCCGGA GGCTTCTGGGGGACGGGCCCTCC AGCTTCACCCCTAAGCAGCGAGC AAGGCTCGTTCTGTTAGAAGGCC GGGAGTGTGTCAGTGACGGCCA CAGCCACCCCTCTGGCGGG ACGGCACCGGCCACTACCTGTGCA ATGCCGTGGCTCTACACAAAGAT GAATGGGCAGAACCGGACACTCAT CAAGCCAAGGAAGACTGTGGC CGCCAGAAGAGCCGGCACCTGTG TGCAAATTGTCAGACGACAACCAC CACCTTATGGGCCGAAACGCCAA CGGGGACCCCTGTCTGCAACGCCCTGT GCCCTTAACAAAGCTGCAAAATG TTAACAGGCCACTGACCATGAAGA AGGAAGGGATCCAGACTCGGAACC GGAAAGATGTCCAACAAGTCCAAGA AGAGCAAGAAAGGGCGGAGTGCT TCGAGGAGCTGTCAGTGATGC AGGAGAAAGTCATCCCCCTCAGTGC AGCTGCCCTGGCTGGACACATGGC ACCTGTGGGCCACCTCCGCCCTTC AGCCACTCGGACACATCCCTGCCCA CTCCGACGCCATCCACCCCTCTC CAGCCTCTCTTCGGCCACCCAC CCGTCAGCATGGTGACCGCCATG GGC			and differentiation. Development 134, 393-405 (2007).
GATA4	ATGTACCAGAGCTGGTATGGCTG CTAATCATGGACCTCCCCCTGGAGC CTATGAAGCCGGAGGACCTGGCGC TTTATGCACTGGAGCTGGCGCCGCT TCTCTCCCGTGTATGTGCTTACAC CTAGAGTGGCCAGCAGCTGCTGG GCCCTTATCTTCAGGGAGGAG AGCAGGATCTGCTCTGGCGAGCT TCAGGGGATCTCTGGAGGGCGCTG CTTCAGGTGCTGGACCTGGAAACTCA ACAGGGATCTCTGGATGGTCACA GGCAGGAGCTGATGGAGCCGCTTA TACCCCTCTCTGTGAGGCCCGAGG TTAGCTTCTGGCACAAACAGGCT CTTAGCTGCCGCTGCTGCTGCCAGC CGCAGCTAGAGAACGAGCTGCTCA TTCTAGTGGCGAGGGAGCTGCTGG AGCCGGCTTAGCTGGAAGAGAGCA GTACGGAAGAGCCGATTGCGGG AAGCTATAGCAGCCCTTACCTGCC TATATGCCGATGTTGGCGATCTT GGCAGCCGCCAGCAGCTCTG CAGGACCTTGTACTCACCTGTGCT TCACTCTGCTGGCAGAGCTAAT CCTGCCGCCAGACATCCAACTTGG ACATGTTGACGACTTCAGCGAGG GCAGAGAATGCCGACTGCCGAG CCATGAGCACCCCCCTTGGAGAA GAGACGGCACCGGCCACTACCTT GCAATGCCCTGTGGCTGTACCAAA GATGAACGGCATCAACAGACCCCT GATCAAGCCCCAGAGAACGACTGAG CGCTAGCAGAACGAGTGCCCTGTC CTGCGCCAATTGCCAGACCAAC CACCACACTGTGGAGGAGAAATGC CGAGGGCGAGCTGTGTAAACGC	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

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CTGTGGACTGTACATGAAGCTGCAC GGCGTCCCCAGACCTCTGGCCATG AGAAAGGAGGGCATCCAGACAGA AAGAGAAAGCCCAGAACCTGAAC AAGAGCAAGACCCCGCTGCTCCTT CTGGAAGCGAGAGCCTGCCCTCAG CCTCTGGAGGCCAGCAATAGCT CTAACGCCACCCACATTTCTCTGA GGAGATGAGGCCATCAAACCGA GCCAGGGCTGAGCAGCCACTACGG CCACAGCTCTAGCGTGAGGCCAGAC TTTAGCGTGTGCGCATGTCAGGC CACGGACCTAGCATTCAACCTGTGC TGAGGCCCTGAAGTTGAGGCCAC AGGGCTATGCTCTCCTGTGTCCTCA GAGCCCTCAGACCTCCAGCAAGCA GGACTCTGGAATTCTCTGGTGTGCT GCCGACAGGCCACGGCGATATCATC ACCGCC			
GATA6	ATGGCCCTGACCACGGCGGATGG TGCTCCCTAAAGATCTGGCGCCG CTGGCGCTGATGCTCTGACAGCAG AGCCTTCCCCCTAGGGAACCCAG CACACACCTAGCCCCATCAGCAG CTCAAGCTCTAGCTGTAGCAGGG CGGAGAGAGAGGACCTGGAGGCG TCTTAACCTGGCGCACACCTCAGCT GATAAGAAGCCGCCGGACCA CCAGCCAGATCTCTTTACTTAGCA GCTACGCCAGGCCACCTTGGCGC TCCATGGACCCCTGCTCTGGT GTGGCGGACCTGGCGGAAACCTG AGCTCTGGAGGAGACCTCTGTG TTACCGACCTGAGCAGGCTGCAC CGCTAGCAAGCTCTGTGAGCAG CAGGGCGCTAAGCTGAGCCCTTT GCCCTGAGCAGCCCGAGGAGATG TACCAAGACCTGGCTCTTAAAGCT CTCAGGGACCTGCCGTTATGACGG AGCCCTGGTGGATTGTTACTCA GCGGCAGCAGCCGAGCTGCTGCA GCCGCTGCCAGCTCACCTGTGTATG TGCCCTACCAAAAGGTGGCGAGCA TGTACCTGGACTCTTACCATCT GCAGGGCAGCGGAAGCGGCCCTGC TAACCATGCCGGAGGAGCTGGAGC TCACCCCGGATGGCCTCAGGCTTCT GCAGATTCTCTCCCTTATGGATCTG GAGGAGGAGCAGCTGGAGGGGGA GCTGAGGACCAAGGTGGAGCGGA AGCGCAGCAGCACATGTGTCTGCC AGATTTCCCTATAGCCCTAGCCCT CTATGCCAATGGCGTGTAGAG AACCCGGAGGATATGCTGCCAG GCTCTGGGGCGCTGGGGAGTTTC TGGAGGGATCTTCACTGGCGCT ATGGGAGGAAGAGAGCCTCAGTAC TCTTCTGTAGCGCCCTAGACAC TGAACGGCACCTATCATCACCACCA CCATCACCATCATCACCCCCAGC CCTTACTCCCCCTATGTGGGAGCC CCCTTACACCCGCTTGGCCTGCCCG CCCTTCTGAGACACCCCTGTGCTG AGCCCTTCAGCTAGACTGGCGCAC CTTTACCACTGCTAGAGGCCCTC TGCCGACTTGCTGGAGGATCTGAGC GAGAGCAGAGAGTGCCTGAACCTG GGCAGCATCCAGACACCCCTGTGG AGAAGAGACGCCACCGGCACTAC CTGTGCAACGCTTGCGCCCTGTACA GCAAGATGAATGGGCTGAGCAGAC CCCTGATCAAGCCCCAGAAGAGGG TGCCCCAGCAGCAGACGGCTGGGAC	52	Involved in cardiac, lung, endoderm and extraembryonic 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). Morrisey, 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 and airway regeneration.

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TGAGCTCGCCAATGTACACC AACACCACACTGTGGCGGAGAAA CGCCGAGGGCAGCCCGTGTAA CGCCTCGGCTTTACATGAAGCTG CACGGCTGCCAGACCTCTGGCC ATGAAGAAGGAGGGAAATCCAGACC AGAAAGAGAAAGCCCAAGAACATC AACAGAGCAAGACCTGCAGGGC AACAGCAACAACAGCATCCCCATG ACCCCACAGCACATCTAGCAAC AGCGACGACTGTAGCAAGAACACA TCACCTACCCACAGCCCACAGCTA GCGGAGCGGCCGCCCCGTGATGA CAGGCGCCGGAGAGTCCACAAATC CCGAGAATAGCGAACTGAAGTACT CTGGACAGGACGGACTGTATATCG GCGTGAGCCTGGCTCTCCGCGCA GGTGACCAAGCTCTGTAGACCTGAC TCTTGGTGTGCCCTCGCCCCGGCC			Nat. Genet. 40, 862-870 (2008).
GLI1	ATGTTCAACTCGATGACCCCACAC CAATCAGTAGCTATGGCAGCCCT GCTGTCTCGGCCCTCCCGTCA GGGGCCCCCAGTGTGGGGAGA AGGACTGTCTGCCCGCCCTCTGC CACCAAGCTAACCTCATGTCGGCC CCCAAGCTTATGGCCAGAG AGACCAACAGTGCACCGAGGGCC CACTTTTCTTCTCCCCGGAGTGC AGTCAAGTTGACCAAGAGCGGC ACTGTCCATCTCACCTCTGTGGAT GCCAGCTGGACCTGCGACGGTT ATCCGCACCTCACCCAGCTCCCTG TAGCTTCATCAACTCGCGATGCAC ATCTCAGGAGGCTCTACGGCAT CTCTCCATTGGCACCATGAGCCCAT CTCTGGGATTCCAGGCCAGATGAA TCACCAAAAAGGGCCCTCGCTCC TTTGGGGTCCAGCCTTGTGGTCCCC ATGACTCTGGGGGGTGGGATGAA TCCCACATCTCAGTCCGGGAGCC CTTCCAACCTGCCAGCTGAAGTCT GAGCTGGACATGCTGGTGGCAAG TGCCGGGAGGAACCCCTTGGAGGT GATATGTCAGCCCCAACCTCACAG GCATAACAGGATCCCCTGTTGGGAT GCTGGATGGGGGGAGGACCTCGA GAGAGAGGAGAAGCGTGAGCGTGA ATCTGTGTATGAACACTGACTCCGT TGGGATGGCTGCAGCCAGGAATT GACTCCAAAGAGCAGCTGGTCCAC CACATCAACAGCGAGCACATCCAC GGGGACGGGAAGGAGTTCTGTGTC CACTGGGGGGCTGCTCCAGGGAG CTGAGGCCCTTCAAAGCCAGTAC ATGCTGGTGGTTACATGGCCAGAC ACACTGGCGAGAGCACACAAGT GCACGTTGAAGGGTGCCTGGAAAGT CATACTCACGCCCTGAAACCTGAA AGACGCACCTGCCGTACACACGG GTGAGAAGCCATACATGTGTGAGC ACGAGGGCTGCAGTAAGCCTTA GCAATGCCAGTGACCGAGCCAAGC ACCAGAATCGGACCCATTCCAATG AGAAGCCGTATGTATGTAAGCTCCC TGGCTGCACCAAACGTATAACAGA TCCTAGCTCGCTGGAAAATGTC AAGACAGTGATGGTCTGACGCC CATGTGACCAAACGGCACCGTGGG GATGGCCCCCTGCCCTGGGACCCAT CCATTCTACAGTGGAGGCCAAGA GGGAGCGGGAAAGGAGGTCCCATCA GGGAGGAAAGCAGACTGACTGTGC CAGAGGGTGCCATGAAGCCACAGC	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).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CAAGCCCTGGGGCCCAGTCATCCGT CAGCAGTGACCACTCCCCGGCAGG GAGTGCAGCCATACAGACAGTGG TGTGGAATGACTGGCAATGCAGG GGGCAGCACTGAAGACCTCTCCAG CTTGGACGAGGGACCTTGCATTGCT GGCACTGGTCTGCCACTCTCGCC GCCTTGAGAACCTCAGGCTGGACC AGCTACATCAACTCCGGCCAATAG GGACCCGGGGTCTCAAACGTGCCA GCTTGTCCCACACCGGTACCATGTT GTCCCCCGCGTGGGCCCGGCGTC TCTCTTGAAACGCCAGCAGCAGCT CCAGCAGCATCAGCTCTGCCATAC TGTCAAGCCGCGCTCCCTGGCC TCTCCTTCCCCCTGGCTCCCCAC CAGAGAATGGAGCATCCTCCCTGC CTGGCCTTATGCCCTGCCAGCACTA CTTGCTTCGGGCAAGATATGCTTCA GCCAGAGGGGGTGGTACTTCGCC ACTGCAGCATCAGCCTGGATCGG ATAGGTGGCTTCCCATGCCCTCCTT GGAGAAGCCGAGCCGAGTATCCAG GATAAACCCCATGCAAGGGTCA CCCGGAGGGCAGTGAACCGAGCC AGGCTGCTGACCGTCTGCTCCAGC TAGAGTCCAGAGGTTCAAGAGGCT GGGCTGTGTCATAACCCCAACCACT GTGGCAGGGGGAGGACAGAATT GATCCTTACCTCCAACCTCTGCT ACTCACACAGCCCCCAGCATCA CTGAGAATGCTGCCATGGATGCTA GAGGGCTACAGGAAGAGCCAGAAG TTGGGACCTCCATGGTGGCAGTG GTCCTGAAACCCCTATATGGACTTCC ACCTACTGATACTCTGGGATATGGG GGACCTGAAGGGGCAGCAGCTGAG CCTTATGGAGCGAGGGTCCAGGC TCTCTGCCCTTGGGCTGGTCAC CCACCAACTATGCCCAACCCCTG TCCCCAGCAGGCCATATCTGAC CCCACCAAGAACATGGGTGAG TTCCTTCCCACTCTGGCTGTAC CAGGGCCCAGGCTCTAGGTGAA CCTACAGCCAGTGTCTCGACTTGA ACATTATGGACAAGTGAAGTCAA GCCAGAACAGGGGTGCCAGTGGG GTCTGACTCCACAGGACTGGCACCC TGCCCTAATGCCAACCCAGTGAG GGCCCCCACATCCACAGCCTCTT TTCCCATTAACCCCAAGCCCTCTCT CCCCAATATCTCCAGTCAGGCCCT ATACCCAGCCACCCCTGATTATCT TCCTTCAGAACCCAGGCCCTGCTG GACTTTGATTCCCCACCCATTCCA CAGGGCAGCTCAAGGCTCAGCTTG TGTGTAATTATGTTCAATCTCAACA GGAGCTACTGTGGGAGGGTGGGG CAGGGAAAGATGCCCGCCCGAGGA ACCTTCCCTACCAAGTCCCAAGTT CTGGGGGGTCCAGGTTAGCCA AGCCGTGCTAAAGCTCCAGTGAA ACATATGGACCTGGCTTGACCCA ACTTGCCCAATCACAAGTCAGGTT CTATCCCCACCCCTCACCATGCCAT GAAAATTGTAGTGGGGCAAT AGGGCTTCACATAGGGCAGCAGCA CCACCTCGACTTCTGCCCTCATGTC CCACTTGCTATGGCCCTCTCAAAGT GGGAGGCCAACACCCAGCTGG TCATCTGAGGTGGGAGGCTAGG AGGGGGTCTGCCTTGTACCCCTCCT CCCGAAGGACAGGTATGTAACCCC CTGGACTCTTGTATCTTGACAAACA			

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CTCAGCTGGACTTTGTGGCTATTCT GGATGAGCCCCAGGGCTGAGTC TCCCTTTCCCATGATCAGGGGGC AGCTCTGGACATACCCACCTCCCT CTGGGCCCCCAACATGGCTGTGG GCAACATGAGTGCTTACTGAGATC CCTACCTGGGAAACAGAATTCTC AACTCTAGTGCC			
HAND2	ATGAGTCTGGTAGGTGGTTTCCC ACCACCGGGTGGTGACCCAGAGG GCTACCCGTTGCCGCCGCCGC CGCCAGCGCTGAGCCATGAGGA GAACCCCTACTTCCATGGCTGGCTC ATCGGCCACCCCGAGATGTCGCC CCGACTACAGCATGGCCCTGCTTA CAGCCCCGAGTATGCCAGCGCAC CGCCAACCGAAGGAGCGGCCAG GACTCAGAGCATCAACAGCGCTT CGCCGAACCTGCCGAGTGCATCCC CAACGTACCCGCCGACACCAA CTCCAAAATCAAGAACCTTGCGCTG GCCACCGACTACATGCCCTACCTCA TGGACCTGCTGCCAAGGACGACC AGAATGGCGAGCGGAGGCC AGGCAGAGATCAAGAAGACCGACG TGAAAGAGGAGAAGAGGAGAAG GAGCTAACGAAATCTGAAAGC ACAGTGAGCAGCAACGACAAGAAA ACCAAAGGCCGAGCGGCTGGCG CAGCACGCTCTGGCCCTGGAC AAGCAG	54	Involved in cardiac development	Srivastava, D. et al. Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. <i>Nat. Genet.</i> 16, 154-160 (1997).
HNF1A	ATGGTTCTAAACTGAGCCAGCTGC AGACGGAGCTCTGGGCCCTGC TGGAGTCAGGGCTGAGCAAAGGG CACTGCTCCAGGCACCTGGGTAGC CGGGGCCCTACCTCTGGCTGGAG AAGGCCCTCTGGACAAGGGGAGT CCTGGCCGGCGCTGAGGGAGC TGGCTGAGCTGCCCAATGGGTGG GGGAGACTCGGGCTCCGAGGACG AGACGGACGACGATGGGAAGACT TCACGCCACCCATCCTCAAAGAGCT GGAGAACCTCAGGCCCTGAGGAGC GGCCCACAGAAAGCCGTGGTGA GACCCCTCTGAGGGAGGACCCGTG GCGTGTGGCAAGATGGTCAGTC CTACCTGCAGCAGCAACATC ACAGCGGGAGGTGGTCGATA TGGCCTCAACCAAGTCCCACCTGT CAACACCTCAACAAAGGGC ATGAAGACGAGAAGCGGCC CTGTACACCTGGTACGTCCG AGCGAGGGTGGCGCAGCAGTCA CCCAGTGCAGGGCAGGGAGGCTGA TTGAAGGCCACAGGTGATGAGC TACCAACCAAGAAGGGCGGAGGA ACCGTTCAAGTGGGCCAGCATC CCAGCAGATCCTGTTCCAGGCC GAGAGGGAGAAGAACCTAGCAAG GAGGAGCAGAGAGACGCTAGTGGAG GAGTGAATAGGGCGGAATGCATC CAGAGAGGGTGTCCCCATCACAG GCAACAGGGCTGGCTCAACCTC GTCACGGAGGTGGTGTCTACA GGTTTGCACACGGGCCAAAGAAG AAGCCTTCCGGACAAGCTGGC TGGACACGTACAGCGGGCCCC CAGGGCCAGGCCGGGACCTG TGCCCCGCTCACAGCT GCCTCCACCTGCCCTCTCCCC AAGGTCCACGGTGTGCGCTATG GAGCCTGCGACAGTGAGACTGCA	55	Involved in liver, kidney, pancreatic and gut development	D' Angelo, A. et al. Hepatocyte nuclear factor 1alpha and beta control terminal differentiation and cell fate commitment in the gut epithelium. Development 137, 1573-82 (2010). Servitj a, J.-M. 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. Organogenesis and Development of the Liver.

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GAA GTAC CCT CAAG CAG CGG CGGT CCCT TAGT GAC AGT GTCT ACAC CCC TC CACCAAGT GTCCCCCAGGGCGT GGAGCCCAGCACAGCCTGCTGAG TACAGAAGCCAAGCTGGTCTCAGC AGCTGGGGGCCCTCCCCCTGTC AGCACCC TGACAGCACTGCACAGC TTGGACCA GAGACATCCCCAGGCC AAC CAGCAGCCCCAGAACCTCATC ATGGCCTCACTTCCCTGGGGTCATGA CCATCGGGCTGGTGAGCCCTGCC CCTGGGTCTCACGTTACCAACACA GGTG CCTCCACCTGGTATCGGCC TGGCCTCCACCGCAGGACAGATG TGCCTGGTCA TACAGCAGCATGGCA GCAGCCTGACCAACCCCTGCAGCC CCAGTTCTCCCAGCCGCTGCACCC TCCTACCCAGCAGCCGCTATGCCAC CTGTGCAAGAGCCATGTGACCCAGA GCCCTTCATGCCACCATGGCTCA GCTGCAGAGCCCCACGCCCTCTAC AGCCACAAGCCCGAGGTGGCCAG TACACCCACACAGGCCATGCC AGACTATGCTCATCACCGAACAC CAACCTGAGGCCCTGGCCAGCCT AGGCCCAACCAAGCAGGTCTTACCT CAGACACTGAGGCCCTCCAGTGAGT CGGGGCTTCACACGCCGCTCTCA GGCCACCACCCCTCCACGTCCCCAG CAGGACCTGCGGGCATCCAGCAC CTGCAGCGGCCACCGGCTCAGC GCCAGCCCCACAGTGTCTCCAGCA GCCTGGTGTACAGAGCTCAG ACTCCAGCAATGCCAGAGCCACC TGCTGCCATCCACACAGCGCTCAT CGAGACCTTCATCTCACCCAGATG GCCCTCTCCCTCCAGTT			Dev. Cell 18, 175-189 (2010). Martovetsky, G., Tee, J. B. & Nigam, S. K. Hepatocyte nuclear factors 4α and 1α regulate kidney developmental expression of drug- metabolizing enzymes and drug transporters. Mol. Pharmacol. 84, 808-23 (2013).
HNF1B	ATGGTTAGCAA ACTGACATCCCTCC ACAGGA ACTCTTTCTGCCCTCCT CTCCAGGGT TAACAAAGAGGT ACTGGTCCAGGCTTGGAGGAGTTG CTCCCCCTCACGGAAATTGGTGTAA AGTTGGAGACTCTCCCCCTCTCCC TGGTTCTGGAGCAGAGCGGATAC TAAACCGT ATTTCATACGTTACA AACGGACACGCAAGGGTCGGCTT TCAGGTGACGAAGGGTCTGAGGAC GGCGATGATTATGACACCCGCC ATCCTCAAAGAACATGCAGGCCCTTA ATACAGAGGAAGCGGGGAGCAGC GAGCTGAAGTTGACAGAAATGCTCT CAGAAGATCCGGAGAGCTCGGA AAATGATTAAGGGATATATGAGC AACATAACATTCCCCAGAGAGGG TAGTTGATGTACCGGCCCTAACCA GAGCCACCTGTCAGCATCTCAAT AAGGGTACTCCTATGAAAACACAG AAGCGAGCGGCCCTTACACATGG TACGTGCGGAAGCAACGAGAAATT CTCCGACAGTTCAATCAGACAGTAC AATCTCAGGGACATGACGGATA AAAGCTCACAGGATCAGCTTGTGTT TCTCTCCCCGAGGTTCAAGGAAACAG TCCCACGGTCCAGGTCAATCTGATG ATGCTTGCAGTGAACCTACAAACA AAAAAATGAGGGAGGAACAGGTTA AATGGGGACCGGCCCTCAGCAGA TACTGTACCAAGCGTACGATCGGC AGAAAAACCCAAGCAAGAGGAGGC GCGAGGCATTGCTGAGGAGTGTA ATCGGGCGAGTGCTTGCAACGGG GTGTAAGTCTAGCAAAGCCCATG GTCTCGGCTCAAACCTGGTCACGG	56	Involved in liver, kidney, pancreatic and gut development	D' Angelo, A. et al. Hepatocyte nuclear factor 1alpha 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. & Bingham, C. HNF1B- associated renal and extra-renal

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GGTGAGGGTATAATTGGTTGCC AACAGGCGGAAGGAGGAAGGCATC CGGCAAAGCTGGCATGGATGCC TACTCAAGCAACAGACACATAGC CTCAACCTCTGGTCACACGGGT CCCTCATACCAACCTTCTTCCTC TCCACCCAACAACATTCTGGTGT CGATATTCCCAGCAGGGAAACAC GAGATAACATCTCCTACTATAA GTCATCACGGAATTCTGCAATGGT AACATCACAGAGTGTGCAACA GGTATCACCCGGTCTTGATCCA GGCCACAATCTGGTAGGCCGTGAC GAAAGATGATCTGTTCTGGTGG CGGACTCCCGCCGGTCTCCACATT ACCAACATACATAGTCTCAGTCATC ATAATCTCAGCAGAGCCAAACC TGATTATGACTCTCTTAGCGGAGT GATGGCTATTGGCAATCTTGAAAC ACTCACAAAGCACAAATGTACCCG TCATAAACAGCGTAGCGGGTCATT GGCGGGCTCCACCAACAGTGCAGTT CTCCAGCAGCTCCATTACCCCCAT CACAGCCTCTGATGAGCAGAGC CCTGGTAGTCACATGGCTAACAGC CGTCATGGCAGCTGCACTCAGCT CCAGAACTCCCATATGTATGCCAC AAGCAAGAACCAACCAATACAGT CACACATCAAGATTCCCCAGTGCTA TGTTGTTACTGACACATCCTCTAT CTCAACTCTGAGCAACATGTCCAGT AGTAAACAATGCTCTGCAAGCAT GG			disease-an expanding clinical spectrum. <i>Nat. Rev. Nephrol.</i> 11, 102-112 (2014). De Vas, M. G. et al. Hnf1b controls pancreas morphogenesis and the generation of Ngn3+ endocrine progenitors. <i>Development</i> 142, 871-82 (2015). El-Khairi, R. & Vallier, L. The role of hepatocyte nuclear factor 1 $\beta$ in disease and development. <i>Diabetes, Obes. Metab.</i> 18, 23-32 (2016).
HNF4A	ATGCGACTCTCCAAAACCCCTCGCG ACATGGACATGGCGACTACAGTG CTGCACTGGACCCAGCCTACACCC CTGGGATTGAGAATGTGCAAGGT GTGACCATGGCAATGACAGCTC CCCATCAGAAGGCACCAACCTCAA CGCGCCAAACAGCCTGGGTGCG CGCCCTGTGTCGCACTGCGGGGAC CGGGCCACGGGAAACACTACGGT GCCCTGAGCTGTGACGGCTGCAAG GGCTTCTCCGGAGGAGCGTGG AAGAACACATGTACTCTGAGA TTAGCGGCAGTGCCTGGTGGAC AAAGACAAGAGGAACCAAGTGC TACTGCAGGCTCAAGAAATGCTCC GGGCTGGCATGAAGAAGGAACCG TCAGAATGAGCGGGACCGATCA GCACTCGAAGGTCAGCTATGAGG ACAGCAGCCTGCCCTCCATCAATGC GCTCTGCAAGGGAGGTCTGTCC CGACAGATCACCTCCCCGGTCTCCG GGATCAACGGGACATTCGGGGA AGAAGATTGCCAGCATCGCAGATG TGTGAGTCCATGAAGGAGCAGC TGCTGGTCTCGTTGAGTGGGCCAA GTACATCCCAGCTTCTGAGCTC CCCTGGACGACCAAGGTGGCCCTG CTCAGAGCCATGCTGGCGACAC CTGCTGCTGGAGGCCACCAAGAGA TCCATGGTGTCAAGGACGTGCTGC TCCTAGGCAATGACTACATTGTCC TCGGCACTGCCGGAGCTGGCGGA GATGAGCGGGGTGTCATACGCAT CCTTGACGAGCTGGTGTGCTGCCCT CAGGAGCTGCAGATCGATGACAT GAGTATGCCCTACCTAAAGCCATCA TCTCTTGACCCAGATGCCAAGGG GCTGAGCGATCCAGGGAAAGATCAA GGGGCTGCGTCCAGGTGCAAGGT	57	Involved in liver, kidney, pancreatic and gut development	Si-Tayeb, K.; Lemaigre, F. P.; Duncan, S. A. Organogenesis and Development of the Liver. <i>Dev. Cell</i> 18, 175-189 (2010). Martovetsky, G., Tee, J. B. & Nigam, S. K. Hepatocyte nuclear factors 4 $\alpha$ and 1 $\alpha$ regulate kidney development 1 expression of drug-metabolizing enzymes and drug transporters. <i>Mol. Pharmacol.</i> 84, 808-23 (2013). Maestro, M. A. et al. Distinct roles of HNF1 $\alpha$ , HNF1 $\alpha$ lpha, and HNF4 $\alpha$ lpha in regulating pancreas

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GAGCTTGGAGGACTACATCAACGA CCGCCAGTATGACTCGCGTGGCCGC TTGGAGAGCTGCTGCTGCTGCTGC CCACCTTGAGAGCATCACCTGGCA GATGATCGAGCAGATCCAGTCATC AAGCTTCTGGCATGGCCAAGATTG ACAACCTGTTGAGGAGATGCTGCT GGGAGGTCCGTGCCAACGCCAGGA GGGGGGGGTTGGAGTGGGACTC CCCAGGAGACAGGCCCTCACACAGT GAGCTCACCCCTCAGCTCCTGGCT TCCCCACTGTGCCGCTTGGCAAG TTGCT			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	ATGGACAACGCCGGATGAATTCC TTCCTGAGTACCCAAATTGCTTA GTGGAGACAGTGGCACTTGAGTG CCCGAGCTATCCATCAGACCA GAATTACAACATCCAAGCTGTC GGTGTAGCCACAGTTGGCGGG AGACGACCGCTTCTGGTCGGAAG AGGGGTTCAAATTGGATCACCTCAC CATCACCATCACCAACCCATCAC ACCCCCAACCGGGACTTACAAA CCAGCGCAATTGGCGTGA ATAGCCATTCCATGTGGACCTTC CTATGGTCTCAGAATTCTCGCC CCTTATAGCCCATAACCCCTGAACC AAGAGGCCGATGTATCAGGAGCT ATCCCCAGTGCAGCGCAGCGTTA CTCAGGTAATCTTCTAGCCCGATG GTCAGCACCAACCATCACCATCAA GTTATGCGGGTGCAGTCGGA TCCCCACAATAACATAACCATAGTT ACGGCCAAGAGCACAATCCCTGG CCCTCGCTACATATAACAACACT GCTCTCGCTTCAATGCTTCCACCAA GAAGCTTGTGGAGTCCCGCTCAG AAACTTCTCTCCAGCTCAGACTTT TGATTGGATGAAGGTCAAGCGGAA TCCGCTAAACGGGAAAGTAGG TGAATATGGCTATTGGGACAGCT AATGCTGTCCGACCAATTTCACAA AAAAACAGCTTACTGAACCTCGAGA AGGAATTCAATTAAAGTATT GAATCGAGCGAGACGAGTCGA CGCCGCTAGTCTCAACTAACGAG ACCCAGGTTAAGATAATGGTCCAG AACAGAAGAATGAAACAAAAAA GCGGGAGAAGGAAGGACTCCCTCC TATATCACCAGCCACACCCCCCAGGT AACGACGAGAACGGCGGAGGAATCT TCAGAGAAGACTCCAGCTCCCTT GTGTTCTCTCTGGTAGCTAAC CAGCGATAACCTCACGACGAGTC C	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 GAAAACGGCAGCCAACTGGCTCACG GCAAAGAGTGGTGGAGAAGCGC TGCCTCTACAGAAGCACAGACA CTGGAGCTGGAGAAGGAGTTCTG TTCAATATGTAACCTACTCGAGAGC GGCGCTAGAGATTAGCCGAGCG TCCACCTCACGGACAGACAAGTGA AAATCTGGTTTCAAGACCGCAGGA	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

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TGAAACTGAAGAAAATGAATCGAG AAAACGGATCCGGGAGCTCACAG CCAACTTAATTTC		vivo. Blood 97, 2286- 2292 (2001). Satokata, I., Benson, G. & Maas, R. Sexually dimorphic sterility phenotypes in Hoxa10- deficient mice. <i>Nature</i> 374, 460-463 (1995).	
HOXA11	ATGGATTTGATGAGCGTGGTCCCT GCTCCTCTAACATGTATTGCCAAG TTGTACTTAACGTCTCGGTCCA GATTTCTCCAGCTCCCTCTTTCT GCCCGAGACCCCGTCTCGGCCCA ATGACATACTCTACTCTCCAACC TGCCCCAGGTCCAACCCGTGCGCG AAGTACCTTCAGAGAGTACGCCA TTGAGCCCAGCAACTATGGCACCC CGCGGCAATCTGGCCACTGCTAC TCGCAGGAGGAGCTCGTGACAGA GACTGCTGCAAGGCCAGGGCG GCCGGCGTGCCTGGCAGCTGCTG GCCAAGAGCTGGCAACGTCTAC CACCACCCACCCCGCAGTCCTG CCAATTCTATAGCACCGTGGCAG GAACGGCGTCTGCCACAGGTTTC GACCAGTTTCAGAGACAGCTACG GCACCCGGAAAACCTCGCTCTC CGACTACCCGGGACAAGAGCGC CGAGAAAGGGCCCCGGCGGGCAC GGCGACCTCCGGCGGGCGGGCG GGCTGCAACGGCGCGCCGGCAAC TTCAAGTCCGACAGCGGGCG CGGGCGCTGCCGGAGATGGGGC GGCAGCAGAGGAGAAAGAGCGC GGCGGGCCCGAGAGCAGCAGCA GCCCGAGTCGTTCCGGCACAC TGAGGACAAGGCCGGCGCTCAG TGGCAACGCAACCGCAAAAGCG CTGCCCTATAACCAAGTACCAAGATC CGAGAGCTGGAACCGGGAGTTCTC TTCAGGGTCTACATTAACAAAGAG AAGGCCCTGCAACTGTCCGCATGC TCAACCTCACTGATCGTCAAGTC AATCTGGTTTCAGAACAGGAGAAT GAAGGAAAAAAATTAACAGAGA CCGTTACAGTACTACTCAGCAAAT CCACTCTCTTG	60	Involved in kidney development	Patterson, L. T., Pembaur, M. & Potter, S. S. Hoxa11 and Hoxd11 regulate branching morphogenesis of the ureteric bud in the developing kidney. Development 2153-2161 (2001).
HOXB6	ATGAGTTCTATTCGTGAACCTCA CCTTCCCCGTCACTCGGCCAGCGG GCAGGAGTCCTTCCCTGGGCCAGCTA CCGCTCTATTCGTCGGGTATCGGG ACCCGCTGAGACATTACCCCGGCC CTACGGGCCAGGGCGGGGCCAGGA CAAGGGCTTGCACCTCCCTCTAT TACCCGCCGGCGGGCGGTGGCTAC GCCGAGCGGCCCTGCGACTAC GGGCCGGCGGCCCTACCGC GAGAAAGAGTCGGCTGCGCACTC TCCGGCGCCGACGAGCAGCCCCCG TTCACCCCGAGCCGCGGAAGTC GACTGGCGCAGAGACAAGAGCGTG TTCGGCGAGACAGAAGAGCAAG TGCTCCACTCCGGTCTACCGTGG TGCAGCGGATGAATTCTGTGAAACA GTTCCCTCTTGGGCCAGCGGGCG	61	Involved in lung and epidermal development	1. Patterson, L. T., Pembaur, M. & Potter, S. S. Hoxa11 and Hoxd11 regulate branching morphogenesis of the ureteric bud in the developing kidney. Development 2153-2161 (2001). Komuves, L.

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GCGAGGCCGAGACATACACAG TTACCAAGACGCTGGAGCTGGAGAA GGAGTTTCACTACAATCGCTACTG ACGCAGCCGGCGCATCGAGATC GCGCACGCCCTGTGCCTGACGGAG AGGCAGATCAAAGATAATGGTTCCAG AACCGACGCATGAAGTGGAAAAAG GAGAGCAAACCTGCTCAGCGCTCT CAGCTCAGTGCCGAGGAGGAGGAA GAAAAACAGGCCAG			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. & Williams, M. C. Retinoic acid alters the expression of pattern- related genes in the developing rat lung. Dev. Dyn. 207, 47- 59 (1996).
KLF4	ATGGCTGTCAGCAGCGCTGCTCC CATCTTCTCCACGTTGCGCTCTGG CCCGCGGAAAGGGAGAACACT GCGTAAGCAGGTGCCGAAATAA CCGCTGGCGGGAGGAGCTCTCCA CATGAAGCGACTTCCCCCAGTGCTT CCCGGCCGCCCTATGACCTGGCGG CGGCAGCGTGGCACAGACCTGG AGACCGCCGGAGCCGTGGCGCTT GCGGCCTAGCAACCTGGGCC TACCTCGGAGAGAGACCGAGGAGT TCAACGATCTCTGGACCTGGACTT TATTCTCTCCAAATTGCGTGACCCAT CCTCCGGAGTCAGTGCCGCC GTGTCTCGTCAGCGTCAGCCTCT CTTCGTCGTCGCCGTCGAGCAGCG CCCTGCCAGCGGCCCTCCACCTGC AGCTTACACCTATCGATCCGGCC GGAACGACCCGGCGTGGCGCC GCGGACGGGGAGTCCCGTACATTCC ATGGCAGGGAGTCCCGTACATTCC GACGGCTCCCTCAACCTGGGGAC ATCAACGACGTGAGCCCCCTGGGC GGCTTCTGGCGAGCTCTGGCGC CAGAAATTGGACCCGGTGTACATTCC GCCGCAGCAGCGCAGCGCAGG TGGCGGGCTGATGGGCAAGTCGT GCTGAAGGCAGTCGCTGAGCGCCC TGGCAGCGAGTACGGCAGGCCGTC GGTCATCAGCGTCAGCAAAGGCAG CCCTGACGGCAGCCACCGGTGGT GGTGGCGCCCTACAACGGCGGGCC GCCGCAGCACGTGCCCAAGATCAA GCAAGGAGGCAGTCTTCGTGCA CACTTGGCGCTGGACCCCTCTCA GCAATGGCCACGGCCGGCTGCA ACGACTTCCCCCTGGGGCGCAGCT CCCCAGCAGGACTACCCCGACCC GGGTCTTGAGGAAGTGCTGAGCAG CAGGGACTGTACCCCTGCCCTGCC CTTCCTCCGGCTTCCATCCCCACC CGGGGCCAATTACCCATCCTTCCT GCCCGATCAGATGCAGCCGCAAGT	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 stem cells

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CCCGCCGCTCCATTACCAAGAGCTC ATGCCACCCGGTCTGCATGCCAG AGGAGCCAAGCCAAGAGGGAA GACGATCGTGGCCCCGGAAAGGA CCGCCACCCACACTTGTGATTACGC GGGCTGGGCAAACCTACACAAA GAGTTCCCACATCTCAAGGCACACTG CGAACCCACACAGGTGAGAAACT TACCACTGTGACTGGGACGGCTGTG GATGGAATTGCCCCGTCAGATG AACTGACCAGGACTACCGTAAAC ACACGGGCACCGCCGTTCCAGT GCCAAAAATGGGACCGAGCATTT CCAGGTGGACACCTCGCCTTACA CATGAAGAGGCATTT			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).
LHX3	ATGGAGGCAGCGGGGGAGCTGGGC CCGGCCGGGAGTCGGGGAGGC GACCTGCTGCTAGCACTGCTGGC GGAGGGCGGACTGGCCGAGAGA TCCCCTGTGCGCTGGCTGTGACCA GCACATCTGGACCGCCTCATCCTC AAGGCTCTGGACCGCCTACTGGCAC AGCAAGTGTCTCAAGTGCAGGGAC TGCCACAGGCCACTGGCGAGCGC TGCTTCAGCCGAGGGGAGAGGGTT TACTGCAAGGACGACTTTTCAAGC GCTTCGGGACCAAGTGCAGCCGCGT GCCAGCTGGGATCCGGCCACGCG AGGTGGTGGCCGGCCAGGACT TCGTGTACCACTGCACTGCTTTG CTGCGTCGTGTGCAAGGGCAGCT GGCCACGGGCACGAGTTCTACCT CATGGAGGACAGCCGGCTCGTGTG CAAGGGGACTACGAACCGCCAA GCAAGCGAGAGGCCACGGCACGGC CAAGCGCCGCGCACGACCATCAC CGCCAAGCAGCTGGAGACGCTGAA GAGCGCTTACAACACCTCGCCCAA GCCGGGGCGCCACGTGGCGAGCA GCTCTCGTCCGAGACGGGCTGG CATGCGCTGTGCAAGGTTGGTTC CAGAACGCCGGCCAAGGAGAAG AGGCTGAAGAAGGACGCCGGGG CAGCGCTGGGGCAGTATTTCCGC AACATGAAGCGCTCCCGCGGGC TCCAAGTCGGACAAGGACAGCGTT CAGGAGGGCAGGACAGCGACGCT GAGGTCTCTTCCCGATGACCTT CCTTGGCGGAAATGGGCCGGCCA ATGGCCTCTACGGGAGCTGGGG AACCCACCCAGGCCCTGGCGGCC CCTCGGGAGCCCTGGCAACTCTC CCTGGAGCATGGAGGCCCTGGCAGG CCCAAGAGCAGTACCGAGAGCTGCG TCCCAGCAGCCCCCTACGGTGTCCCC CCATCCCCGCCGCCGGCAGAGC CTCCCTGGCCCCCAGCCCCCTCT CCAGCTGGTGTACCCAGACACCA GCTTGGGCCCTGTGCCCCCTGGAGC CCCCGGCGGGCCCCACCCATGAG GGTGTGGCAGGGAACGGACCCAG TTCTGACCTATCACGGGGAGCAGC GGGGGTACCCGACTTCCCTGCCA GCCCGCCTCTGGCTGGATGAGGT AGACCACGCTCAGTTCTCAGGCC ATGGGCCAGCTTCTTGAC	63	Involved in pituitary gland development	Sheng, H. Z. et al. Multistep Control of Pituitary Organogenesis. Science (80-). 278, 1809-1812 (1997).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
LMX1A	ATGGAAGGAATCATGAACCCCTAC ACGGCTCTGCCAACCCACAGCAG CTCCCTGGCCATCGAGCAGAGTGTCT ACAGCTCAGATCCCTCCGACAGG GTCTCACCCCACCCAGATGCCCTGG AGACCACATGCACCCCTATGGTGCC GAGCCCCCTTTCCATGACCTGGATA GCGACGACACCTCCCTCAGTAACCT GGGTGACTGTTCCCTAGCAACCTCA GAAGCTGGGCTCTGCAGTCCAGA GTGGAAACCCCATTGACCATCTGT ACTCCATGCAGAATTCTTACCTCAC ATCT	64	Involved in neuronal development	Lin, W. et al. <i>Foxal and Foxa2 function both upstream of and cooperatively with Lmx1a and Lmx1b in a feedforward loop promoting meso-diencephalic dopaminergic neuron development.</i> <i>Dev. Biol.</i> 333, 386-396 (2009). Qiaolin, D. et al. <i>Specific and integrated roles of Lmx1a, Lmx1b and Phox2a in ventral midbrain development.</i> <i>Development</i> 138, 3399-3408 (2011).
MEF2C	ATGGGGAGAAAAAGATTCAAGATT ACGAGGATTATGGATGAACCTAAC AGACAGGTGACATTACAAAGAGG AAATTGGGTGATGAAGAAGGCT TATGACCTGACCGTGTGTGACT GTGAGATTGCGCTGATCATCTCAA CAGCACCAACAAGCTGTTCCAGTAT GCCAGCACCGACATGGACAAAGTG CTTCTCAAGTACACGGAGTACAC GAGCCCATGAGAGCGGGACAAAC TCAGACATCGTGGAGACGTTGAGA AAGAAGGGCCTTAATGGCTGTGAC AGCCCCAGACCCGATGCGGACGAT TCCGTAGGTACAGCCCTGACTCTG AGGACAAGTACAGGAAAATTACG AAGATATTGATCTAATGATCAGCA GGCAAAGATTGTTGCTGTTCCACC TCCCAACTTCGAGATGCCAGTCCTCC ATCCCAGTGTCCAGCCACAACAGTT TGGTGTACAGCAACCTGTGAGCTC ACTGGGAAACCCCAACCTATTGCC ACTGGCTCACCCCTCTGAGAGG AATAGTATGTCCTCTGGTGTAAACAC ATCGACCTCCAAGGTGAGGTAAACA CAGGTGGCTGTGATGGTGGAGACC TCACGTCGTTGCAGGCACCGATG AGGGAACGGGTATGGCAATCCCG AAACTCACCAAGGTCTGCTGGCTC CTGGTAACCTGAACAAGGAATATG CAAGCAAATCTCCTCCCCAATGA ATTAGGAATGAATAACCGTAAAC CAGATCTCGAGTTCTTATTCCACC AGGCAGCAAGAATACGATGCCATC AGTGTCTGAGGATGTCGACCTGCTT TTGAATCAAAGGATAATAACTCC CAGTCGGCTCAGTCATGGTACCC CAGTGGTTCCGTAGCAACTCC TTTACCAAGGACAAGGAATGGGAGG ATATCCATCAGCCATTCAACACAA	65	Involved in cardiac development	Lin, Q. et al. <i>Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C.</i> <i>Science</i> 276, 1404-7 (1997).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TATGGTACCGAGATACTCTCTGAGTA GTGCAGACCTGTCATCTCTGTCCTGG GTTAACACCGCCAGGGCTTTCAC CTTGGTCAGTAACTGGCTGCCAAC AGCAACACCTACATAACATGCCAC CATCTGCCCTCAGTCAGTTGGGAGC TTGCACTAGCACTCATTATCTCAG AGTTCAAAATCTCCCTGCCCTCTA CTCAAAGCCTCAACATCAAGTCAG AACCTGTTTCTCTCTTAGAGACCCG TACCACCAACCCCTCGAGATACCCA CAACACACGCCAACGAGGGCGGG AGATCTCTGTTGACAGCTTGAGCA GCTGTAGCAGTTCGTACGACGGGA GCGACCGAGAGGATCACCGGAACG AATTCCACTCCCCATTGGACTCAC CAGACCTTCGCGGGACGAAAGGGA AAGTCCCTCAGTCAAGCGCATGCG ACTTCTGAAGGATGGGCAACA			
MESP1	ATGGCCCAGCCCCGTGCCCCGCCG TCTCCGAGTCTGGATGCTCTCTGC GCCGGGGCCCCAACTCGGCCGCC GCCGCCCTCCGACAAGGACTGCCG CCGCTCCCTCGTCTGTCGGGAGAC TCATGGGCAGCACCCAGCGAC AGCCCCGGTGGGAGGCCCGCGCG CCAGGCACCCCTCGGGACCCCGC GCCCCCTCCGTAGGTAGGCTGGCG GGCGCAGCAGCCGCTGGGCAGC GGGCAGAGGCGAGAGGCCAGTGAG CGGGAGAAAATGCGCATGCGCACG CTGGCCCGGCCCTGACGAGCTGC GCCGTTTCTACCGCCGTCCGTGGC GCCCGGGGGCAGAGGCCGACCAA GATCGAGACGCTGCGCCTGGCTATC CGCTATATCGGCCACCTGTCGCCG TGCTAGGCCTCAGCGAGGAGAGTC TCCAGGGCCGGTGGCCGGCAGGGCG GTGACCCGGGCTCCCTCGGGCT GCCCGTGTGCCCGACGACTGCC CGCGCAGATGCAAGACACGGACGA GGCTGAGGGCAGGGGGCAGGGCG CGGGCTGGGCTGGTATCCGCCGTC CGCGCCGGGGCGTCCCTGGGATCC CCGCCTGCCGTCCCCGGAGCCGA GCTGCAACCGAGCCGGCGACCCG CCTGCGCTGTTGCCGAGGCCGGT GCCCGGAAGGGCAGGGATGGAGC CAAGCCCACCGTCCCCGCTCTTCC GGGCAGCTGCTGGCTCTGTGGA GACCTGGATGCCCTCTCGCCTCTG GAGTGGCTGCCAGGGAGCCCAAG TTG	66	Involved in cardiac development	Bondu, A. et al. Mesp1 Acts as a Master Regulator of Multipotent Cardiovascular Progenitor Specification. Cell Stem Cell 3, 69-84 (2008).
MITF	ATGGCTGGAAATGCTAGAATAATAAT CACTATCAGGTGCAAGACCCACCTCG AAAACCCCACCAAGTACCAACATAC AGCAAGCCCAACGGCAGCAGGTTAA AGCAGTACCTTTCTACCTTCTAGC AAATAAACATGCCAACCAAGTCCT GAGCTGCCATGCCAACCAAGCCT GGCGATCATGTCATGCCACCGGTGC CGGGGAGCAGCGCACCCAAACGCC CCATGGCTATGCTTACGCTTAACCTC CAACTGTAAAAAGAGGGATTTA TAAGTTGAAGAGCAAAACAGGGC AGAGAGCGAGTGCCAGGCATGAA CACACATTACAGAGCGTCTGTATG CAGATGGATGATGTAATCGATGAC ATCATTAGCCTAGAATCAAGTATA ATGAGGAAATCTGGGCTTGATGG ATCCTGCTTGCAGGAAATGGCAAATAC GTTGCCCTGTCGGGAAACTTGATT	67	Involved in pigment cell and melanocyte differentiation	Widlund, H. R. & Fisher, D. E. Microphthalmia-associated transcription factor: a critical regulator of pigment cell development and survival. Oncogene 22, 3035-3041 (2003).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GATCTTATGAAAACCAAGGTCTGC CCCCACCAGGCCCTCACCACAGCA ACTCCTGTCAGCCAACCTTCCCAA CATAAAAAGGGAGCTCACAGAGTC TGAAGCAAGAGCACTGGCCAAAAGA GAGGCAGAAAAAAGGACAATCACAA CCTGATTGAACGAAGAAGAAGATT TAACATAAATGACCGATTAAAGA ACTAGGTACTTGTATCCCAAGTCA AATGATCCAGACATGCGCTGGAAC AAGGGAACCCTTAAAGCATCC GTGGACTATATCCGAAAGTTGCAA CGAGAACAGCAACGCGAAAAGAA CTTGAAAACCGACAGAAGAAACTG GAGCACGCCAACCGGCATTGTTGC TCAGAAATACAGGAACCTGAAATGC AGGCTCGAGCTCATGGACTTCCCT TATTCCATCCACGGGTTCTCTGCTCT CCAGATTTGGTGAATCGGATCATCA AGCAAGAACCCGTTCTGAGAACT GCAGCCAAGACCTCCTCAGCATCA TGCAGACCTAACCTGTACAACAACT CTCGATCTCACGGATGGCACCATCA CCTTCACAAACAAACCTGGAACTG GGACTGAGGCCAACAAAGCTATA GTGTCCCCACAAAAATGGGATCCA AACTGGAAGACATCTGTATGGACG ACACCCTTCTCCGTCGGTGTAC TGATCCACTCCTTCCTCAGTGTCC CCCGGAGCTTCCAAAACAAGCAGC CGGAGGAGCACTATGAGCATGGAA GAGACGGAGCACACTTGT			
MYC	ATGCCCTCAACGTTAGCTTCACCA ACAGGAACTATGACCTCGACTACG ACTCGGTGCAAGCGTATTCTCTAG CGACGAGGAGGAGAATTCTACCA GCAGCAGCAGCAGAGCGAGCTGCA GCCCGCCGCCAGCGAGGATAT CTGGAAGAAATTGCACTGCTGCC CACCCGGCCCTGTCCTGACCGC CGCTCCGGCTCTGTCGCCCTCT ACGTTGCGGTACACCCCTCTCCCT TCGGGGAGACAACGACGGCGGTGG CGGGAGCTTCTCACCGGCCACCA GCTGGAGATGGTACCGAGCTGCT GGGAGGAGACATGGTGAACCAAGAG TTTCATCTGCGACCCGGACGACGAG ACCTTCATCAAAACATCATCATCC AGGACTGTATGGAGCGGCTTCTC GGCGCCGCCAACGCTCGTCTCAGA GAAGCTGGCCTCTACCGGCTG GGCAGGAGACAGCGGAGCGCGAA CCCCAGCCGCCAGCGCTG CTCCACCTCCAGCTTGTACCTGAG GATCTGAGGCCGCCCTCAGAG TGCATGACCCCTCGTGGCTTCC CCTACCCCTCTCAACGACAGCAGCTC GCCCAAGTCTGCGCTCGAAGA CTCCAGCGCCTCTCCGCTCTCG GATTCTCTGCTCTCCCTGACGGAGT CCTCCCGCAGGGCAGCCCCGAGC CCCTGGTGCCTCATGAGGAGACAC CGCCCACCCAGCAGCGACTCTG AGGAGGAACAAGAAGATGAGGAA GAAATCGATGTTGTTCTGTGGAAA AGAGGCAGGCTCTGGCAAAGGT CAGAGTCTGGATCACCTTCTGCTGG AGGCCACAGCAAACCTCTCACAG CCCACTGGCTCTCAAGAGGTGCAC GTCTCACACATCAGCACAACTACG CAGCGCCTCCCTCACTCGGAAGG ACTATCTGCTGCCAAGAGGGTCA AGTTGGACAGTGTAGGTCTGA	68	Involved in cell proliferation, differentiation and apoptosis. Reprogramming factor for induction of pluripotency.	Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. <i>Nat. Rev. Cancer</i> 2, 764-776 (2002). Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. <i>Cell</i> 126, 663-76 (2006). Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. <i>Cell</i> 131, 861-72 (2007). Yu, J. et al. Induced Pluripotent Stem Cell

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GACAGATCAGCAACAACCGAAAAAT GCACCAAGCCCCCAGGTCTCGGACA CCGAGGAGAAATGTCAGAGGGCAA CACACAACGTTGGAGCGCAGA GGAGGAACGAGCTAACCGGAGCT TTTTTGCCTGCGTGACCAGATCCC GGAGTTGGAAAACAATGAAAAGGC CCCCAAGGTAGTTATCCTTAAAAAA GCCACAGCATACATCCTGTCCGTC AAGCAGAGGAGCAAAAGCTCATTT CTGAAGAGGACTTGTGCGGAAAC GACGAGAACAGTGAACACAAAC TTGAACAGCTACCGAACTCTTGTC G			Lines Derived from Human Somatic Cells. <i>Science</i> (80-). 318, 1917-1920 (2007).
MYCL	ATGGACTACGACTCGTACCGCACT ATTCTACGACTATGACTGCGGGGA GGATTCTACCGCTCACGGGCC AGCGAGGACATCTGGAAGAAATC GAGCTGGTGCATCGCCCCAACCGT CGCCGCCCTGGGCTGGGTCCCG CGCAGGGGACCCGGCCCCCGGAT TGGTCCCCCGGAGCGGTGGCCCG AGGGTGCACCGAGACGAAGCGGA ATCCCCGGGCACTCGAAAGGCTG GGGCAGGAACTA CGCTTCCATCAT ACGCCGTGACTGCATGTGGAGCGG CTCTCGGCCGGGAACGGCTGGA GAGAGCTGTGAGCGAACGGCTGC TCCCTGGCGGCCCGGGGAAACCC GCCCAAGGCGTCCGCCGCCCGGA CTGCACTCCCAGCCTCGAAGCGGC AACCCGGCGCCGCCGCCCTGTC CGCTGGCGAACCCAAGACCCAGG CTGCTCGGGTCCCGAGAGCCCAA GCGACTCGGTAAGGACCTCCCG AGCCATCAAGAGGGGGCCACCCC ATGGGTGCCAAAGCTCTGCCCTG CTGAGGTCAAGCATTGGCTCTCT CAAGCTTGGCCATCTCCGCTC TCTTTGCG	69	Involved in cell proliferation, differentiation and apoptosis.	Hatton, K. S. et al. Expression and activity of L-Myc in normal mouse development. <i>Mol. Cell.</i> <i>Biol.</i> 16, 1794-804 (1996).
MYCN	ATGCCGAGTTGTTCCACGTCTACGA TGCAGGAATGATATGCAAGAAC CCGACTTGGAGTTGACTCTTGCA ACCATGTTTATCCGGATGAAGAC GACTTTTATTCGGCGGGCGACA GCACCCCTCCTGGAGAGGACATCT GGAAAAAAATTGAACTTTGCTTAC ACCCCACTCAGCCCTCTGAGGA TTGCGGAACACAGCAGTGAACCG CCGTCTTGGGTGACAGAGATGCTCC TCGAGAACGAAATTGTGGGAAGCC CTGCGGAGGAAGACGGCTTCGGGC TCGGTGGAACCTGGAGGTCTCACGCC GAACCCAGTCAACTGCAGGATG CATGTGGTCTGGATTCTCAGCTCGG GAGAAGCTGGAACAGGGCAGTTCT GAGAAACTCCAACATGGCCGGGC CCTCCAACAGCGGGTTCTACCGCAC AGTCCCCCTGGTGTGAGGCCCTAG TCCCGGGGAGAGGCCATGGGG CGCGGAGGAGGGTAGGGCGG CGCTGCGTTGCCGTGAGCTTGCG CACCCGGCCGCTGAATGTGAGATC CGCGGTAGTGTTCGTTCCCGT TAATAAGCGAGAACCGGCACCGT GCCAGCCGCTCTGCGTCTGCACCC GGCGCAGGTCTGCTGCGCTCAG GAGCAGGTATTGCGCTCTGCG GGGCACCAAGGAGTAGCCCTCCAA GGCCCGCGGTAGGCAAACCTCG GCGCGACCAAAAGCACTCTCAA CGAGCGGAGAGGATACACTGTCCG	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. <i>Genes Dev.</i> 14, 1390-9 (2000). Sawai, S. et al. Defects of embryonic organogenesis resulting from targeted disruption of the N-myc gene in the mouse. <i>Development</i> 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
	ATAGTGATGACGGAGGACGACGAAG AGGAGGACGAGGAGGAGGAGATA CATGTGTCACGGTCGAGAACGCA AGGAGTCTTCAAATACAAAAGCG GTAACGACATTACGATAAACAGTA AGACCTAAGAACGCAGCCCTGGT CCAGGGCGGGGCCAGTCCAGTGAG CTTATACCTAACCGCTGCCCTGCCA TTCACCAGCAGCATAACTACGCGG CCCCTAGTCCCTACGTTGAGAGCGA GGATGCCCCCCCACAAAAAAAT AAAGTCTGAAGCGTCCCCCGCCCC CTGAAATCCGTAATCCCCCAAAG GCGAAGTCACTCAGTCCCAGGAAT TCAGATTCCGGAGCTCCGAACGG CGGCAGGAACTATAACATACTTGAG AGACAACGACGCAATGACCTGAGG TCTCTTTTTGACCCCTCGAGATC ACGTCCCGAGCTGGTTAAGAATG AGAAAGCTGCGAAGGGTAGTCATAC TGAAAAGGCCACCGAGTATGTC ATAGTTGCAAGCTGAGGAGCAC AGCTTCTCCTGAAAAGGAGAAC TTCAGGCACGACAACAGCAATTGC TGAAAAGATTGAGCATGCACGCA CTTGT			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 ACGGCTCTCTCTGCTCTTTCAC AACGGACGACTTCTATGACGCC GTGTTTCGACTCCCGGACCTCGC TTCTCGAAGACCTGGACCCCGGCC TGATGCACGTGGCGCGCTCTGA AACCGAAGAGCACTCGCACTTC CCGGCGGGTGCACCCGGCCCG GCGCACGTGAGGACGAGCATGTC GCGCGCCCAGCGGGCACCCACCG CGGGCCGCTGCTACTGTGGGGCTG CAAGGCGTGAAGCGAAGACCAAC CAACGCCGACCGCCGCAAGGCC CACCATGCGCAGCGGCCGCGCT GAGCAAAGTAAATGAGGCCATTGA GACACTCAAGCGCTGACGTCGAG CAATCCAACCAGCGGTTGCCCCAA GGTGGAGATCTTGCACGCGCAT CCGCTATATCGAGGGCTGAGGCT CTGCTGCGCAGCAGGACGCCCG CCCCCTGGCGCCCGACGCCCTCT ATGCGCCGGGCCGACTACAGCG GCGCGCGCGGAGCAGCAGCG GCGACTCCGACCGTCCAGCCCG GCTCCAACGCTCCGACGGCATGAT GGACTACAGCGCCCCCGAGCGG CGCCCGGGCGGGAAGTCTACGA AGGCGCTACTACAACGAGGCC CAGCGAACCCAGGCCGGAAAGAG TGCAGGGGTGTCGAGCCTAGACTG CCTGTCAGCATCGTGGAGCGCATC TCCACCGAGAGCCCTGGGGCCCC GCCCTCTGCTGGCGACGTGCCTT CTGAGTCGCCTCCGCGCAGGCAAG AGGCTGCCGCCAGCGAGGGAG AGAGCAGCGGCCAGCCACCCAGT CACCGGAGCGCCGCCAGTGCC CTGCGGGTGCAGAACCCAAACCGA TATACCAGGTGCTC	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).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
MYOG	ATGGAGCTGTATGAGACATCCCCCT ACTTCTACCAGGAACCCCGTTCTA TGATGGGAAACACTACCTGCGTGTG CACCTCCAGGGCTTCGAACCACCA GGCTACGAGCGACGGAGCTCACC CTGAGCCCCGAGGGCTGGGACCC CCCGAGCACTGTCAGGCCAGTGC CTGCCGTGGCGTGTAAAGGTGTGA AGAGGAAGTCGGTGTCCGTGGACC GGCGGGGGCGCCACACTGAGGG AGAACGCAAGGCTCAAGAAGGTGA ATGAGGCCCTTGAGGGCCCTGAAGA GAAGCACCCCTGCTCAACCCCCAAC AGCGGTGCCCAGGTGGAGATCC TGCGCAGTGCATCCAGTACATCGA GCCCTCCAGGCCCTGCTCAGCTCC CTCAACCGAGGAGGAGCGTGACCTC CGCTACGGGGGGGGGGGGGGGGCC CAGCCAGGGGTGCCAGCGAATGC AGCTCTCACAGGCCCTCTGCAGTC CAGAGTGGGGCACTGTGCACTGGAGT TAGCGCCAACCCAGGGGATCATC TGCTCACGGCTGACCTACAGATGC CCACAACCTGCACTCCCTCACCTCC ATCGTGGACAGCATCACAGTGGAA GATGTGTCTGTGGCCTCCAGATG AAACCATGCCAAC	72	Involved in skeletal muscle specification and differentiation	Pownall, M. E., Gustafsson, M. K. & Emerson, C. P. Myogenic Regulatory Factors and the Specification of Muscle Progenitors in Vertebrate Embryos. <i>Annu. Rev. Cell Dev. Biol.</i> 18, 747-783 (2002). Shi, X. & Garry, D. J. Muscle stem cells in development, regeneration, and disease. <i>Genes Dev.</i> 20, 1692-708 (2006).
NEUROD1	ATGACCAAATCGTACAGCGAGAGT GGGCTGTAGGGCGAGCTCGACCCC CAAGGTCTCTCAAGCTGGACAGAC GAGTGTCTCAGTCTCAGGACGAG GAGCAGGAGGAGACAAGGAAGGA GGACGACCTCGAAGGACGAGC AGAGGGAGACTCACTGAGGAACGG GGGAGAGGAGGGAGGAGCAAGATG AGGACCTGGAAGAGGAGGAGGA GAGGAAGAGGAGGATGACGATCAA AAGCCAAGAGACGCCGGCCAAA AAGAAGAAGATGACTAAGGCTGC CTGGAGCGTTTTAAATTGAGACGCA TGAAGGCTAACGCCGGAGCGGA ACCGCATGCACGGACTGAACCGGG CGCTAGACAACCTGCGCAAGGTGG TGCCCTGCTATCTAAAGACGCGAA GCTGTCAAATCGAGACTCTGCGC TTGGCCAAGAACATACATCTGGGCTC TGTGGAGATCTGCGCTCAGGCA AAAGCCCAGACTGGCTCTTCCTCGT TCAGACGCTTGTCAAGGGCTTATCC CAACCCACCAACCTGGTGTGCG GGCTGCCGTGCAACTCAATCCTCGGA CTTTCTGCTTGAGCAGAACCGGA CATGCCCTCCACCTGCCGACGGCC AGCGCTTCCCTCCCTGTACACCCCT ACTCTTACCACTGCGCTGGCTGCC CAGTCGGCTTACGGTACCATGGAC AGCTCCCATGTCTTCACTGGTTAAGC CTCCGCCGCACGCCAACGGCAG CGCTGGAGCCCTTGTAAAGGCC TCTGACTGATTCGACCAAGGCCCTTCC TTTGATGGACCCCTCAGCCCGCGC TCAGCATCAATGGCAACTTCTCTTT CAAACACGAACCGTCCGGAGATT TGAGAAAAATTATGCTTTACCATG CACTATCTGCAAGCGACACTGGCA GGGGCCCAAAGCCACCGATCAATC TTCTCAGGCACCGCTGCCCTCGCT GGCAGATCCCCATAGACAATATTAT GTCCTTCGATAGCCATTACATCAT GAGCGAGTCATGAGTGCCCCAGCTC ATGCCATATTCATGAT	73	Involved in neuronal specification and differentiation Demonstrated to induce neuronal factor differentiation in hPSCs	Pataskar, A. et al. NeuroD1 reprograms chromatin and transcription landscapes to induce the neuronal program. <i>EMBO J.</i> 35, 24-45 (2016). Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. <i>Neuron</i> 78, 785-98 (2013).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
NEURO G1	ATGCCAGCCCGCTTGAGACCTGCA TCTCCGACCTGACTGGCCAGCAG CAGCGCAGTCAGCTATCCGCCCTC CTCACCGACGAGGAAGACTGTGCC AGACTCCAACAGGCAGCTCCGCC CGGGGCCGCCCAATACTCCCGGG CGTCTGAGGTTCCAGGGGCACAGG ACGACGAGCAGGAGAGGGCGGCGC GCCGCGGCCGGACGCCGGTCCGCT CCGAGCCGCTCTGACTCGCTGCG CAGGAGCCGGCGCTCAAGGCCAA CGATCCGAGGCCAACCGCATGCA CAACTGAAACGCCGCCCTGGACGC ACTGCCAGCGCTGCTGCCCTGTT CCCGACGACACCAAGCTACCAAA ATCGAGACGCTGCGCTTCGCTAC ACTACATCTGGGCTCTGGCGAGAC ACTGCCGCTGGGGATCAAGGGCT GCCCGGAGGCAGGTGCCGGGAGCG CCTCCCTGCCGCCAGTGCCTCCCC TGCCCTGCCGGTCCCCCAAGCCCC CCAGCGACGCCGAGTCTGGGCGT CAGGTGCCGCCCGCCTCCCCGCT CTCTGACCCCAGTAGGCCAGGCC TCCTGAAGACTTACCTACCGCCCC GGACCCCTGTTTCTCCCTCCAAAG CCTGCCAAAGACTTGCTCCACACA ACGCCCTGTTCATTCCTTACACAC	74	Involved in neuronal specification and differentiation	Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. <i>Nat. Rev. Neurosci.</i> 3, 517-530 (2002).
NEURO G3	ATGACACCACAAACCATCTGGGCTC CCACAGTCAGGTGACCGAGAGA CTGAAAGATCATCCACGCGCCTC CGAGGATGAGGTGACATGTCAAAC TAGCGCACCCCCCTCTCACCCGG ACCCGGGAATTGTGCTGAGGCC GAAGAGGGAGGATGCAAGAGGAGC ACCAAGGAAACTCGAGGCCGACG GGGTGAAAGAACGCCGCCAACGTC TGAGCTGCCCTTAGCAAGCAGCG CCGCAGTCGGAGGAAAAGGCAA CGACCGGAAAGGAATTAGGATGCA TAATCTTAATTCTGCTCTGGACGCT CTGGCAGGGTACTTCTACTTTCC CGGATGACGCGAAATTGACCAAGA TAGAGAGCTCCGGTTTGCAATAAA TTACATCTGGGCTCTACACAAACA CTGAGATTGCCATCACAGCTTT ACGCTTGTGAGCCACCGGCCCGCA CTGTGGCAGCTGGGTAGCCCCGG CGGCTCTCTGGAGACTGGGGTCT TTGTATTCTCTGTAGCCAAGCGG GATCTTGAGTCGGCTGCCAGTCT CGAAGAAAGACCCGGACTCTTGG AGCGACTTTTCAGCATGCTGTCC CCTGGCTATTGGCTTCTCAGACT TTTTG	75	Involved in pancreatic development, and neuronal specification and differentiation	Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. <i>Nat. Rev. Neurosci.</i> 3, 517-530 (2002). Arda, H. E. et al. Gene Regulatory Networks Governing Pancreas Development. <i>Dev. Cell</i> 25, 5-13 (2013).
NRL	ATGGCCCTGCCCTCCAGCCGCTGG CCATGGAATATGTCATGACTTTGA CTTGATGAAGTTGAGGTAAAGCG GGAAACCTCTGAGGGGCCAGCTGG CCCACCTACAGCTCAGTGCCTCCT ACACCTTACAGCTCAGTGCACAGGCT CACCCACCTTCAGTGAACCGAGGCT GCTAGGGCAACCGAGGGTACACG ACCAGGTTGGAGGAGCTGTACTG GCTTGTACCCCTGCAGCAGCTT GGGGCTGGGAGGCATTGGGACTG AGTCCCTGAAGAGGCCATGGAGCTA CTGCAAGGTCAAGGCCAGTCCCT GTTGATGGACCCCATGGTTACTACC CAGGGAGCCAGAGGAGACAGGAG	76	Involved in photoreceptor development	Mears, A. J. et al. Nrl is required for rod photoreceptor development. <i>Nat. Genet.</i> 29, 447-452 (2001).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CCCAGCACGTTAGTTGGCAGAGC GGTTTTCGACGCCGGCTTGTCTC GATGTCGTGGAGAACTAAACCG GCAGCTGCCGGATGCCGGAGAGA CGAGGCTCTACGACTGAAGCAGAG GGGTCGAAACGCTGAAGAACCGTGG CTATGGCAAGCATGTCGTTCCAAG AGGCTCAAACAGAGGCCAGGTCTT GAGGCCGAGGCCGCCGTCTGCA GCCAGCTAGATGCGCTACGAGCT GAAGTAGCACCTTGCAAGAGAG CGAGATCTTACAAGGCTCGTGTG ACCGGTAACCTCGACTGCCCG GTCGGGGATCCCTCCACCTTT CCTCTGCCAACCTTCTGTACAAA GTTGTCCCC			
ONECU T1	ATGAACCGCAGCTGACCAGTGGAA GGCATGGCAGCTGACGGGTG AGCCATGAGCCGGTGCCCGCCCT GCCGACCTGCTGGCGGGCAGCCCC CACGCCGCAGCTCCGTGGCAC CGGGCAAGGCCACCTGCCCG CACCCGGCTCATGGCATGGCGT CCCTGCTGGACGGCGGAGCGGG GGGGAGATTACCAACCACCA GGGCCCTGAGCACAGCTGGCCG GCCCTGGCATCCCACCATGACCAT GGCCTGGAGACTCCCCCAGGTAT GAGCATGCCAACACCTACACAC CTTGACCCCTCTGCAGCGTGCT CCCATCTCACAGTCTGGCAAGT TCCCCCACCATACCAACCA TCACCAACCAACCCGACCA CCAGCGCTGGGGCAACGTGAG CGTAGCTTCAGGCTCATGCGGGAT GAGCGGGGTGGCTCCATGAAT AACCTTATAACCCCTACCAAGG AGCTGGCGGCATGGGCAGAGCC TCTCGCCCTTCCAGCTCCGGTCT GGGCAGCATCCACAACCTCCAGCA AGGGCTCCCCACTATGCCAACCCG GGGGCCGCATGCCACCGACAG ATGCTCACCCCAACGGCTCGAAG CCCACACCCGGCATGCTCGGG GCCACGGGAGCAGCACCTCACGC CCACCTGGCGGCATGGTGCCAT CAACGGCTTCTCGCACATTCCC CACGCCAACCTGAACGCCAGGG CACGGCAACTCTGGCACAGCC CGGGAGCCAACCCCTCGGTGACC GGCGCCAGGTAGCAATGGAAGT AATTCAAGGGCAGATGGAAGAGATC AATACAAAAGAGGTGGCGAGCGT ATCACCAACCGAGCTAACGCTAC AGCATCCCACAGGCCATTTGCC AGAGGGTGGCTTGCCGCTCCAGG GGACCCCTCTGGACCTGCTGCCCAA CCCCAAACCCCTGGAGCAAACCTCAA ATCCGGCCGGGGAGACCTTCCGGAG GATGTTGAAGTGGCTGCAGGAGCC GGAGTTCCAGGCCATGTCGGCTC CGCTTAGCAGCATGCAAAAGGAAA GAACAAGAACATGGGAAGGATAGA GCCAACACACCCAAAAGGCCAGG TTGGTCTTCACAGATGTCAGCGTC GAACTCTACATGCAATTTCAGG AAAATAAGCGTCATCCAAAGAAT TGCAAATCACCATTTCCAGCAGCT GGGGTGGAGCTGAGCACTGTCAG	77	Involved in retinal, liver, gallbladder and pancreatic development	Chakrabarti, S. K., et al. Transcription factors direct the development and function of pancreatic $\beta$ cells. Trends Endocrinol. Metab. 14, 78-84 (2003). Clotman, F. et al. The onecut transcription factor HNF6 is required for normal development of the biliary tract. Development 129, 1819-1828 (2002). Sapkota, D. et al. Onecut1 and Onecut2 redundantly 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 CGAGGGCAGCTCCAATTCAAGGCAA CTCATTTCTTCATCAAGCACTTGT ACCAAAAGCA			
OTX2	ATGATGTCTTCTTAAGCAACCGC CTTACCCAGTCATGGCTGAGTCT GACCACCTCGGGTATGGACTGGCTG CACCCCTCCGTTGGGTACCCGGGC CCTGGGCTTCTTGTCCCGAGCAC CCCCCGAAACAGCGCCGGGAGAC GACGACGTTACTCGGGCGAGCT AGATGTGCTGGAAGCTGTGGC AAGACCCGGTACAGACATCTC ATGGCAGAGGAGGTGGCACTGA ATCAACTTGGCCCAGTCAGGGTG CAGGTATGGTTAAAGAATCGAAGA GCTAAGTGGCCCAACAAACAGCAA CAACAGCAGAAATGGAGGTCAA AAAGTGAAGACCTGCCAAAAGAAG ACATCTCAGCTCGGGAAAGTGAGTT CAGAGAGTGGAAACAAGTGGCCAT TCACTCCCCCTCTAGCACCTCAGT CCCGACCATTGCCAGCAGCAGTGCT CCTGTGTCTATCTGGAGGCCAGCTT CCATCTCCCACTGTCAAGTCCCTT GTCCACCTCCCTCTCTGCAATGCAG AGGTCTATCCATGACCTATACTC AGGCTTCAGGTATAGTCAGGAT ATGCTGGCTCAACTTCTACTTGG GGGCATGGACTGTGGATCATATTG ACCCCTATGCATCACAGCTTCCC GACCAGGGCCACACTCAGTCCC TGGGTACCAATGCAGTCACCGAGC ATCTCAATCAGTCCCCAGCTCTCT TTCCACCCAGGGATAATGGAGCTTC AGCTTGGTTTTAACTCAACCACTG ATGCTGGGATTATAAGGACCAAAC TGCCTCTGGAAAGCTTAACCTCAAT GCTGACTGCTTGGATTATAAGATC AGACATCCTCGTGGAAATTCCAGGT TTG	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	ATGGCGGCCCTCCCGCACGGTAC CGAGAACATGCGGCCGGCTCCGG GGCAGAACTACCCCGCAGGGAT TCCCTTGGAAAGTGTCCACCCGGCT TGGCCAAGGCCGGTCAATCAGCT GGGAGGGGTCTTCATCAATGGCG ACCCCTGCCATAACCATCCGCCAC AAGATAGTGGAGATGGGACCAT GGCATCGGCCCTGTGTCACTCCC GACAGCTGCGTGTCTCCACGGCTG CGTCTCCAAGATCTTGCCTCTAC CAGGAGACCGGGTCCATCGGCCCT GGGGCCATCGGGCAGCAAGCCC AGACAGGTGGCAGTCCGGATGTA GAGAAAAGATGGAGGATCAAG AGGGAAAACCCAGGCATGTTCA TGGGAGATCCGGGACAGGCTGCTG AAGGATGGGACTGTGACCGAAGC ACTGTGCCCTCAGTGAAGTTCGATTA GCCGCGTGTCTCAGAATCAAGTTC GGAAGAAAGAGGAGGAGGATGAA GCGGACAAGAAGGAGGACGAGCG GAAAAGAAGGCCAAACACAGCATC GACGGCATCTGGCGACAAAGGG AACCGGCTGGAGGAGGCTGGAT GTGGAGTCGGAACCTGACCTCCA CTGAAGCGCAAGCAGCGACGAGT CGGACCACTCACGGCGAGCAG CTGGAGGAGCTGGAGAAGGCCTT GAGAGGACCCACTACCCAGACATA	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).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TACACCCGCGAGGGAGCTGGCGAG AGGACCAAGCTGACAGAGGGCGT GTGCAGGTCTGGTCAGTAACCGCC GCGCCCGTTGGCGTAAGCAGGCAG GAGCCAACCAGCTGGCGGCC ACCCACCTCTGCCAGGAGGCTTCCC GCCAACCGGCATGCCAACGCTGCC CCCCCTACCACTGGCCGACTCCACC TACCCCACCACCAACATCTCCAAAG ATGGGGGCAGCAGCTGTGCACCGGC CTCAGCCCCCTGCCACCGTCCACCAT GCACCAAGGGCGGCTGGCTGCAGC GGCTGAGGCCGCGAACACAGCTC TGCCCTACGGAGCCCAGCCACAGCTC TCCAGCTACTCTGACAGCTTCATGA ATCCGGCGGCCCTCCAACCAT GAACCCGGTCAGCAAGGCCTGTC TCCTCAGGTATGAGCATCTTGGC AACCCCGAGTGGGGTGCCTCCCGAG CCACAGGCTGACTTCTCATCTCCC CGCTGATGGCGGCCCTGGACTCGG CCACCTTCATCTCAGCCAGCTGCAG CCAGCGGGCCGACTCCATCAAGCC AGGAGACAGCTGCCAACCTCCA GGCCTACTGCCAACCCACCTACAGC ACCACCGGCTACAGCGTGGACCC GTGGCCGGCTATCAGTACGGCCAG TACGGCCAGAGTGAGTGCGCTGTG CCCTGGCGTCCCCCGTCCCCATTG CTTCTCCACCCCCAGGGCCTCTG CTTGTATTGAGAGCTACAAGGTG GTGTCAAGGTGGGAATGTCCATT CACAGATGAAAATTGAAGTCCA GCCAGATGGAACAGTTCAAC			
POU1F1	ATGAGTTGCCAAGCTTTACTTCGG CTGATACTTTATACTCTGAATTCTG TGACGCCCTCTGCAACTCTGCCCTCTG ATATATGCATCACAGTCTGCCAGT GTCTACCAAGCTCTCAACCATGCCAC CAATGTGATGTCACAGCAACAGG ACTTCATTATTCTGTTCTCTCTGTC ATTATGGAAACCAAGCCATCAACCT ATGGACTGATGCCAGGTAGTTAA CCCCTTGTCTTATAAATTCTGTA COACACCTTGAGTCATGGATTCCCT CCTATACACCAAGCCTCTCTGCCAG AGGACCCCACAGCTGCTGATTCAA GCAGGAACCTAGGGGAAAGTAA ATGGTGGAAAGAGCCAATAGACAT GGATTCTCCAGAAATCAGAGAACT TGAAAAGTTGCCAATGAATTAA GTGAGACGAATTAAATTAGGATAC ACCCAGACAAATGTTGGGAGGCC CTGGCAGCTGTGCATGGCTCTGAAAT TCAGTCAAACAAACAATCTGCCGATT TGAAAATCTGCAGCTCAGCTTAA AATGCATGCAAACATGAAAGCAATA TTATCCAAATGCTGGAGGAAGCT GAGCAAGTAGGGAGCTTGACAT GAAAAGTGGGAGCAAATGAAAGG AAAAGAAAACGAAGAACAACTATA AGCATTGCTGCTAAAGATGCTCTGG AGAGACACTTTGGAGAACAGAATA AACCTTCTCTCAAGAGATCATGAG GATGGCTGAAGAACTGAATCTGGA GAAAGAAAGTAGTAAAGAGTTTGTGTT TTGCAACCGGAGGCAGAGAGAAAA ACGGGTGAAAACAAGTCTGAATCA CATCTTGAGTGAGATCAGGCCCTCA TGGGCCAGCTTCTTGAC	80	Involved in pituitary gland development	Turton, J. P. G. et al. Novel Mutations within the POU1F1 Gene Associated with Variable Combined Pituitary Hormone Deficiency. J. Clin. Endocrinol. Metab. 90, 4762-4770 (2005).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
POU5F1	ATGGCGGGACACCTGGCTTCAGATT TTGCCCTTCTCGCCCCCTCCAGTGG TGGAGGTGATGGGCCAGGGCC GGAGCGGGCTGGGTTGATCCTCG CACCTGGCTAACGCTTCCAAGGCCCT CCTGGAGGGCAGGAATCGGGCCG GGGGTTGGGCCAGGCTCTGAGGTG TGGGGGATTCCCCATGCCCGCCG CGTATGAGTTCTGTGGGGGGATGG CGTACTGTGGGCCAGGTTGGAGT GGGGCTAGTGCCCCAAGCGGCC GGAGACCTCTCAGCTGAGGGCGA AGCAGGAGTCGGGTGGAGGCAA CTCCGATGGGCCTCCCCGGAGCCC TGCACCGTCACCCCTGGTGCCTG AGCTGGAGAAGGAGAAAGCTGGAGC AAAACCGGAGGAGTCCCAGGACA TCAAAGCTCTGAGAAAGAACTCG AGCAATTGCCAAGCTCTGAAGC AGAAGAGGATCACCCCTGGGATA CACAGGCCGATGTGGGCTACCC TGGGGTTCTATTGGAAAGGTATT CAGCCAACAGCACCATCTGCCGCTT GAGGCTCTGCAGCTTAGCTTCAAGA ACATGTGTAAGCTGCGGCCCTGCT GCAGAAGTGGTGGAGGAAGCTGA CAACAATGAAAATCTCAGGAGAT ATGCAAGCAGAAACCCCTCGTGA GGCCCGAAAGAGAAAGCGAACCAG TATCGAGAACCGAGTGAAGAGC CTTGGAGAATTGTTCCTGCAGTGC CCGAAACCCACACTGCAGCAGATC AGCCACATCGCCAGCAGCAGCTGG CTCGAGAAGGATGTGGTCCGAGTG TGGTTCTGTAACCGGGCCAGAAG GGCAAGCGATCAAGCAGCAGTAT GCACAACGAGAGGATTTGAGGCT GCTGGTCTCTTCTCAGGGGAC CAGTGTCTTCTGCCCCAGG GCCCATTTGGTACCCAGGCTAT GGGAGCCCTACTCACTGCACTGT ACTCCTCGGTCCTTCCCTGAGGG GGAAGCCTTCCCCCTGTCTGTG ACCACCTGGCTCTCCATGCATT CAAAC	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, 1917-1920 (2007).
RUNX1	ATGGCTTCAGACAGCATATTGAGT CATTTCTTCTGTAACCCACAGTGC CATGAGAGAATGCATACTTGGAT GAATCCTCTAGAGACGTCCACGAT GCCAGCACGAGCCGCGCTTCACG CCGCCTTCCACCGCGTGGAGCCAG GCAAGATGAGCGAGGGCTGGCG TGGCGCCCGGAGCGCCGCGCTG CCCTGGCGGCAAGCTGAGGAGCG GCGACCGCAGCATGGTGGAGGTGC TGGCCGACCACCGGGGAGCTGG TGCAGCAGCAGGCCAACTTCC CTGCTCCGTGCTGCCTACGCACTGG CGCTGCAACAAAGACCTGGCCATC GCTTCAAGGTGGTGGCCCTAGGG GATGTTCCAGATGGCACTCTGGTCA	82	Involved in haematopoietic cell development	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

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CTGTGATGGCTGGCAATGATGAAA ACTACTCGGCTGAGCTGAGAAATG CTACCGCAGCCATGAAGAACCGG TTGCAAGATTAAATGACCTCAGGTT TGTGGTCGAAGTGGAAAGAGGGAA AAGCTTCACTCTGACCATCAGTGT TTCACAAACCCACCGCAAGTCGCC ACCTACACAGGCCATCAAATC ACAGTGGATGGGCCCCGAGAACCT CGAAGACATCGGCAGAAACTAGAT GATCAGACCAAGCCGGGAGCTG TCCTTTCCGAGCGGCTCAGTGAAAC TGAGCAGCTGCGGCGCACAGCCA TGAGGGTCAGGCCACACCACCCAG CCCCCACGCCAACCTCGTCGCTC CTGAAACCACTCCACTGCCCTTAAC CCTCAGCCTCAGAGTCAGATGCAG GATAAAGGCAGATCCAACCATCC CCACCGTGGCTCATCGATCAGTCCT ACCAATACCTGGATCCATTGCTCTC TCCTCTGTGCACCCAGCAACGCC ATTACCTGGACGTGCCAGGGCA TGACAACCCCTCTGCGAGAACTTTC CAGTCAGCTCAACGCCACCCGA CCTGACAGCGTTAGCGACCCCG CCAGTCCCCGGCTGCCCTCCATC TCAGACCCCCGGCATGCACTATCCAG GGCCCTTCACCTACTCCCCGAGGCC GGTCACCTCGGGCATGGCATCGG CATGTCGCCATGGGCTCGGCCAC GGCTACCAACACCTACCTGCCGCCG CCCTACCCGGCTCGCGAACCGC AGGGAGGCCGTCCAAGCCAGCT CGCCCTCCTACACCTGTACTACGG CGCCCTGGCCGGCTCTACCAAGTTC TCCATGGTGGGGGGCGAGCGCTCG CGGCCGCCATCTGCCGCCCTGCA CCAACGCCCTCACCGGCTCCGCC GCTCAACCCCGACCTCCGAACCA GAGCGACGTGGTGGAGGCCGAGGG CAGCCACAGCAACTCCCCAACCAA CATGGCGCCCTCCGCGGCCCTGGA GGAGGCCGTGTGGAGGCCCTAC		hematopoietic commitment at the hemangioblast stage of development in vitro.	Blood 100, 458-66 (2002).
SIX1	ATGTCGATGCTGCCGTCGTTGGCT TTACGAGGAGCAAGTGGCGTGC TGTGGAGGTCTGAGCAAGCG GAAACCTGGAGGCCAGGT TCTCTGGTCACTGCCGCTGCC CCACCTGACAAGAACGAGAGCGT ACTCAAGGCCAAGGGCGGTGGTGC CTTCACCGCGGCAACTTCCGTGAG CTCTACAAGATCTGGAGAGCCAC CAGTTCTCGCCTACAACCAACCCA AACTGAGCAACTGTGGCTGAAGG CGCATTACGTGGAGGCCGAGAAC TGTGGCCGACCCCCTGGGCCCGT GGGCAAATATCGGGTGCGCCAAA ATTTCACCTGCCGCCACCATCTGG GACGGCGAGGAGACCGACTACTGC TTCAAGGAGAACGTGAGGGGTGTC CTGGGGAGTGGTACGCCAACAT CCTACCCATGCCGCCGTGAGAG CGGGAGCTGGCCGAGGCCACCGGC CTCACCAACCCAGGTCAAGCAACT GGTTAAAGAACCGGAGGCAAAGAG ACCGGGCCGCCGAGGCCAAGGAAA GGGAGAACACCGAAACAAATAACT CCTCCTCCAACAAGCAGAACCAAC TCTCTCTCTGGAAGGGGGCAAGCC GTCATGTCCAGCTCAGAACAGGA ATTCACCTCCCCAAAGTCCAGAC CAGAACTCGGTCTTCTGCTGCAGG GCAATATGGGCCACGCCAGGAGCT	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.

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	CAAACATTCTCTCCGGGCTAAC AGCCTCGCAGCCCAGTCACGCC GCAGACCCACCAAGCATCAGCTCA AGACTCTGCTCGGCCCTCAC TCCAGTCTGGTGGACTTGGGTTCC			311, 53-68 (2007).
SIX2	ATGTCCATGCTGCCAACCTTCGGCT TCACGAGGACAAGTGGCGTGC TGTGCAGGTGCTGCAGCAGGGCG GCACACATCGAGCGGTGGGCC TCTGTGGTCGCTGCCCGCT GCACCTTCACAAAGAATGAAACG GCTCAAGGCCAAGGCCGTGGTGG CTTCCACCGCGCGCAACTCCCGAG CTCTACAAGATCTGGAGAGCCAC CAGTTCTCGCCGCAACACCAGCCA AGCTGCAGCAGCTGTGGCTCAAG CACACTACATCGAGGCCGGAGAAC TGC CGCGGCCGACCCCTGGGCGCG TGGGCAAATACCGCGTGC CGCG AATTCCCGCTGCCCGCGCTCCATCTG GGACGGCGAGGGAGACCAGCTACTG CTTCAAGGAAAAGAGTCGCAGCGT GCTGCCGAGTGTA CGGCCACAA CCCCCTACCCCTCACCCCGAGAAC CGTGAGCTGACGGAGGCCACGGGC CTCACCAACACAGGT CAGCAAC TGGTTCAAGAACCGCGCGAGCG GACCGGGCGCCGAGGCCAAGGAA AGGGAGAACACAGAGAACCTCAA TCTAACAGCCACAAACCGCTGAAT GGCAGCGGCAAGTCGGTGTAGGC AGCTCGGAGGATGAGAACGACTCCA TCGGGGACGCCAGACCACTCATCA TC CAGCCCCGCACTGCTCCTCAGCC CGCCGCCCTGGGCTGCCGTCC GCACAGCTGGCCACCCCTCCGG CCCACCGCAGTGCCAGTGCGGT GCCAGGGCGAGGTGGAGGCCACCC ACTGCAACACCAACATGGCCTGCA GGACTCCATCTCAACCCATGTCA GCCAACCTCGTGGACCTGGGCTCC	84	Involved in kidney development	Kobayashi, A. et al. <i>Six2</i> Defines and Regulates a Multipotent Self-Renewing Nephron Progenitor Population throughout Mammalian Kidney Development. Cell Stem Cell 3, 169-181 (2008).
SNAI2	ATGCCGGCTCTTCTGGTCAAGA AGCATTCAACGCCCTCCAAAAGC CAAACATCAGCGAACTGGACACAC ATACAGTGATTATTCCTCGTATCT CTATGAGAGTTACTCCATGCCGTG ATACCCACAACCCAGAGATCCTCAGC TCAGGAGCATACGCCCATCACT GTGTGACTACGCCGTCTCCATCC ACGCCAGCTACCAATGGCCTCTC TCCTTTCCGATACTCCTCATCTT TGGGGAGTGAGTCCCCCTCTCC ATCTGACACCTCTCCAAAGGACAC AGTGGCTCAGAAAGCCCCATTAGT GATGAAGAGGAAAGACTACAGTCC AAGCTTCAGACCCCCATGCCATTG AAGCTGAAAAGTTCACTGCAATT ATGCAATAAGACCTTATTCAACTTT TCTGGCTGGCAAACATAAGCAG CTGCACTGCGATGCCAGTCAGAA AATCTTCAGCTGTAATAACTGTGA CAAGGAATATGAGGCTGGGCG CCTGAAGATGCATATTGGACCCAC ACATTACCTTGTTGCAAGATCT GCGGCAAGGCCGTTCCAGACCTG GTGCTTCAGGACACATTAGAAGT CACACGGGGAGAAGCCTTTCTT GCCCTCACTGCAACAGAGCATTG AGACAGTCAAATCTGAGGGCTCA	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 SNAI2 in Cancer and Development. Annu. Rev. Genet. 41, 41-61 (2007).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TCTGCAGACCCATTCTGATGTAAAG AAATACCAAGTGCAAAAAGTGTCC AAAACCTTCTCAGAATGTCTCTCC TCGCACAAACATGAGGAATCTGGCT GCTGTGTAGCACAC			
SOX10	ATGGCGGAGGAGCAGGACCTATCG GAGGTGGAGCTGAGCCCCGTGGC TCGGAGGAGGCCCGTGCCTGTCCC CGGGGAGCGCCCCCTCGCTAGGGC CCGACGGCGGGCGGGATCGG GCCTGGAGGAGGCCGGGGCCAG GCGAGCTGGCAAGGTCAAGAAGG AGCAGCAGGACGGCGAGGCAGCG ATGACAAGTTCCCGTGTGACATCG CGAGGCCGTCAAGCCAGGTGTCAG CGGCTACGACTGGACGCTGGTGC CATGCCGTGCGCGTCAACGGCGC CAGCAAAGCAAGCCGACGTC GGGGCCCATGAACGCCATTG TGGGCTCAGGCAGCGCGCAGGAAG CTCGCGGAGGAGTACCCGCACCTGC ACAACGCTGAGCTCAGCAAGACGC TGGCAAGCTTGGAGGCTGTGA ACGAAAGTGACAAGGCCCTCA TCGAGGAGGCTGAGCGGCTCGTA TGCAGCACAAGAAAGACCACCCGG ACTACAGTACCAAGGCCAGGGCG GGAAGAACGGGAAGGCCGCCAGG GCGAGGGGGAGTGGCCCGGTGGGG AGGCGGAGCAAGGTGGGACCCCG CCATCCAGGCCACTACAAGAGCG CCCACTGGACCACCGGCACCCAG GAGAGGGCTCCCCATGTCAGATG GGAACCCCGAGCACCCCTCAGGCC AGAGCCATGGCCCACCCACCCCTC CAACCAACCCGAAGACAGAGCTGC AGTCGGCAAGGCAGACCCGAAGC GGGACGGCGCTCCATGGGGAGG GCCGGAAAGCTCACATCGACTTC GCAACGTGGCATTGTTGAGATCA GCCACGAGGTAATGTCAACATGG AGACCTTTGATGTGGCTGAGTTGGA CCAGTACCTGGCCGCCAACATGGCA CCCAGGCCATGTGAGCAGCTACTC AGCAGCCGGCTATGGCTGGCAG TGCCTGGCCCTGGCCAGTGGACA CTCCGCTGGATCTCCAAGCCACCA GGCGTGGCTCTGCCAACGGTCTCAC CACCTGGTGTGGATGCCAACGCC AGGTGAAGACAGAGACGGCGGGC CCCAGGGGCCAACACTAACACCG ACCAAGCCATCCACCTCACAGATCGC CTACACCTCCCTCAGCTGCCAAC TATGGCTCAGCCTTCCCTCCATCT CCGGCCCCCAGTTGACTACTCTGA CCATCAGCCCTCAGGACCCCTATTAT GGCCACTCGGGCTCAGGCCCTCTGGC CTCTACTCGGCCCTCTCTATATGG GGCCCTCGCAGGGCCCCCTACAC GGCCATCTCAGCCCCAGGCCCTCA GGGCCAGTCCCACAGCCCCACA CACTGGGAGCAGCCAGTATAACG ACACTGTCCCCGGCC	86	Involved in neural crest and neuronal development	Southard-Smith, E. M., Kos, L. & Pavan, W. J. SOX10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. <i>Nat. Genet.</i> 18, 60-64 (1998). Britsch, S. et al. The transcription factor Sox10 is a key regulator of peripheral glial development. <i>Genes Dev.</i> 15, 66-78 (2001).
SOX2	ATGTACAACATGATGGAGACGGAG CTGAAGCCGGCGGGCCGCAGCAA ACTTCGGGGGGCGGGCGGCAC TCACCGCGGCCGGCGCCGGCGC AACCAAGAAAACAGCCGGACCGC GTCAAGGGCCCATGATGCTTC TGGTGTGGTCCCGCGGGCAGCGGC GCAAGATGGCCAGGAGAACCCCA AGATGCACAACCTGGAGATCAGCA	87	Involved in regulation of pluripotency and embryogenesis, and in neuronal development. Reprogramming factor for	Boyer, L. A., et al. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. <i>Cell</i> 122,

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	AGCGCCTGGGCCGAGTGGAAAC TTTGTCGGAGACGGAGAACGGC CGTCATCGACGAGGCTAACGGC TGCAGCGCTGCACATGAAGGGC ACCCGGATTATAATACCGGCC GGCGAAAACCAAGACGCTCATGA AGAAGGATAAGTACACGCTGCCG GCGGCTGCTGCCCGGCCGCA ATAGCATGGCGAGCGGGGTGGGG TGGCGCCCGCTGGCGGGCG TGAACCAGCGCATGGACAGTTACG CGCACATGAACGGCTGGAGAACG GCAGCTACAGCATGATGCAGGACC AGCTGGGCTACCCGAGCACCCGG GCCTCAATGCGCACGGCGAGCGC AGATGCAGCCCCATGACCGCTACG ACGTGAGCGCCCTGCAGTACAAC CCATGACCACTCGCAGACCTACAT GAACGGCTGCCACCTACAGCAT GTCTACTCGCAGCAGGGCACCCCT GGCATGGCTTGGCTCCATGGGTT CGGTGGTCAAGTCCGAGGCCAGCT CCAGCCCCCTGTGGTTACCTCTTC CTCCCACCTCAGGGCGCCCTGCCAG GCCGGGACCTCGGGACATGATC AGCATSTATCTCCCGGGCGAG GTGCGGAACCCGGCGCCCCAGC AGACTTCACATGTCAGCACTACC AGAGCGGCCGGTGCCGGCACGG CCATTAACGGCACACTGCCCTCTC ACACATG		induction of pluripotency. 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 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).	947-956 (2005).
SOX3	ATGCGACCTGTCGAGAGAACTCAT CAGGTGCGAGAACGCCGGGGTTC CTGCTGATTTGGCGCGGAGCATTT GATAAGCCTACCCCTCCGCCGAC TCGCTGGCCACAGGCCCAAGCT CCGCTCCGACGGAGTCCCAGGGCC	88	Involved in neuronal and pituitary development	Rizzoti, K. et al. SOX3 is required during the formation of the

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	TTTCACCGTGGCCGCTCCAGCCCC GGGAGGCCCTCTCCCTCCCACG CTGGCCACCTTCTTCCGCCCG CAATGTACAGCCTCTGGAGACTGA ACTCAAGAACCCGTAGGGACACC CACACAAGCGGGGGGACCGGGGG CCCCGAGGCCGGGGAGGCGCAGG CAAGACTAGTGGAAACCGAGCGG CGGCAGAAGCTGGGGCGGCAG CAGCGGTGGTGGAGCGGAGGTTG CGGGGTACAGACAGGACCGTGT GAAACGGCCATGAACGCCCTCAT GGTATGGTCCCGCGGGCAGCGCG CAAATGGCCCTGGAGAACCCCAA GATGCACAATTCTGAGATCAGCAA GGCCTGGGCCGACTGGAAACT GCTGACCGACGCCGAGAAGCGACC ATTCATCGACGGGCAAGCGACT TCGGCCGCTGACATGAAGGAGTA TCGGAACTACAAGTACCGACCGG CCGCAAGACCAAGACGCTGCTCAA GAAAGATAAGTACTCCCTGCCAG CGGCCTCTGCTTCCGGTGCGCG GCGCGCCCGCCGCTGCCGCC GCAAGCGCTGGCCAGCAGTCG GTGGCGTGGGCCAGCGCCTGGAC ACGTACACGCACTGAAACGGCTGG GCAACCGCGCTACTCGCTGTG CAGGAGCAGCTGGCTACGCGCAG CCCCGAGCATGAGCAGCCCGCG CCGCCGCCGCCGCTGCCGCCATG CACCGCTACGACATGGCCGGCTG CAGTACAGCCCATGATGCCGCC GGCGCTCAGAGCTACATGAACGTC GCTGCCGCCGCCGCCGCTCG GGCTACGGGGCATGGCGCCCTCA GCCACAGCAGCCGCCGCC TACGGGCAGCAGCCGCCACGCC GGGGCCGAGCTGGCCGCCAGGCC GCCATGAGCTGGCCCCATGGC TCGGTAGTGAAGTCTGAGCCCAGCT CGCCGCCGCCGCCATCGCATCG ACTCTCAGCGCGCTGCCCTCGCGA CTCGCGGACATGATCAGCATGTAC CTGCCACCGGGGGGACGCCGCC GACGCCGCTCTCGCTGCCGCC GTCGCTGCAACGGCGTGCACCA ACTACCAGGGGCCGGGACTGCAG TCAACCGAACGGTGCCGCTGACCC ACATC			hypothalamo-pituitary axis. Nat. Genet. 36, 247-255 (2004).
SPI1	ATGTTACAGGGCTGAAAAATGGAA GGGTTTCCCTCGTCCCCCTCAGC CATCAGAACCTGGTGCCTATG ACACGGATCTATACCAACGCCAA CGCACAGATATTACCCCTATCTAG CAGTGTGGGGAGAGCCATAGCGA CCATTACTGGGACTTCCACCCCCAC CACGTGCACAGCGAGTTCGAGAGC TTCGGCGAGAACAACTTCACGGAG CTCCAGAGCGTGCAGCCCCCGAG CTGCAGCAGCTTACGCCACATGG AGCTGGAGCAGATGCACTGCTCTCG ATACCCCCCATGGTGCCACCCCATCC CAGTCCTGGCACCCAGGTCTCTAC CTGCCCGGGATGTGCCCTCAGTACC CATCCCTGTCCCCAGGCCAGCCAG CTCAGATGAGGAGGAGGGCGAGCG GCAGAGCCCCCACTGGAGGTGTC TGACGGCAGGGGGATGGCTCTGG GCCCGGGGCTGGCTCTGGCTGG GAGACAGGCAGCAAGAAGAAGATC CGCTGTACCGATTCTGTTGGACC TGCTCCGAGCGCGACATGAAGG	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	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	ACAGCATCTGGTGGGTGACAAGG ACAAGGGCACCTCCAGTTCTGTC CAAGCACAAGGAGGCCGTGGCGA CCGCTGGGCATCCAGAAGGCCAA CCGCAAGAAGATGACCTACCAGAA GATGGGGCGCCGCTGCCAACATA CGGCAAGACGGCGAGGTCAAGAA GGTGAAGAAGAGCTCACCTACCA GTTCAGCGCGAAGTGTCTGGCCG CGGGGGCTCTGCCGAGCGGCCCA CCCGCCCCAC			with transformation. <i>Nat. Rev.</i> <i>Immunol.</i> 7, 105-117 (2007).
SPIB	ATGCTGCCCTGGAGGCTGCACAG CTCGACGGCACACTTCAGCTGTC TGTAACCCAGATGGCTCTCTATGA CTTGGACAGCTGCAAGCATTCCAG CTACCCCTGATTCAAGAGGGGCTCCT GACTCCCTGTGGACTGGACTGTG CCCCACTGTCCCAGGCCACCCCTA TGAAGCCTTCGACCCGGCAGCAGC CGCTTTAGCCACCCCCAGGCTGCC CAGCTCTGCTACGAACCCCCCACCT ACAGCCCTGCAAGGGAACCTCGAAC TGGCCCCCAGCTGGAGGCCCCGG GGCCTGGCCTCCCGCATACCCAC GGAGAACTTCGCTAGCCAGACCT GGTTCCCCCGGCCATATGCCGTCAC CCAGCCCTGTGCTATCAGAGGAG GAAGACTTACCGTTGGACAGCCCT GCCCTGGAGGTCTGGACAGCGAG TCGGATGAGGCCCTCTGGCTGCC CCGAGGGGAAGGGATCGAGGAG GGACTCGCAAGAAGCTGCCCTGT ACCAGTTCCCTGCTGGGGCTACTGAC GCCGGGGACATGGCTGAGTGC GTGGTGGGTGGAGGCCAGGCCGG CGTCTTCAGTTCTCTCCAAGCAC AAGGAACTCCTGGCGCGCGCTGG GCCAGCAGAAAGGGAACCGCAAG CCCATGACCTACCAAAGCTGCC CCGCCCTCCGAACACTACGCCAG ACCGCGAGATCCGAAGGTCAAG CGCAAGCTCACCTACCAAGTTCGACA GGCGCTGCTGCCCTGCAGTCCCG GCCCTTG	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. <i>Oncogene</i> 19, 6432-6442 (2000).
SPIC	ATGACGTGTGTTGAACAAGACAAG CTGGGTCAAGCATTTGAAGATGCTT TTGAGGTTCTGAGGAACATTCAC TGGAGATCTTCAGTACTCGCAGAT TACAGAAATTACCTGCTTTAATCA ACCATGTCCTCATGTCAAAGGAA ATTCAGCTGTATGGAGTGTGGCC TACAGAGGAGCTGTCTATAATTGG AGAACCGTAATTAAACAGTGTGCG GACTTCTATTGAGGAATATTCA ATCAATCTCTGAGAACATAACTGA AAACCAGCTGGTACAACCCACTTT CTCCAGCAAAAGGGGGAAAAGGC AGGAAGAAGCTCGACTGTTGAA TACCTTCACGAACTCCCTGTTAATC CGGAGATGGCATTTGTTACTG GGTAGATAAAACCAAGGCATCTT TCAGTTGTATCAAAAAACAAAGA AAAACCTGCCGAGCTTGGGGAA AAGAAAAGGCCAACAGGAAGACAT GACTTACCCAGAAAATGCCAGGGC ACTCAGAAAATTACGGAAGGAAGTGG GGAAATTACCAAATCCGGAGGAA GCTGACTTACCAAGCTCTCCATCCT ATTCTCCAAAGACTCTCTCCATCCT ATTTCCTGGGAAAGAGATCTTCTA TTCACAGTGTGTTCAACCTGATCAA GAATATCTCAGTTAAATAACTGGA	91	Involved in macrophage development	Kohyama, M. et al. Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. <i>Nature</i> 457, 318-321 (2009).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	ATGCAAATTATAATTATACATATGC CAATTACCATGAGCTAAATCACCAT GATTGC			
SRY	ATGCAATCATATGCTTCTGCTATGT TAAGCGTATTCAACAGCGATGTTA CAGTCAGCTGTGCAAGAGAATAT TCCCCTCTCCGGAGAACGCTCTCC TTCCCTTGCACTGAAAGCTGTAACT CTAAGTATCAGTGTGAACGGGAG AAAACAGTAAGGCAACGCTCAGG ATAGAGTGAAGGCCATAACG CATTACATGTGGTCTCGCAGTCA GAGGCCAAGATGGCTCTAGAGAA TCCCAGAATGCGAAACTCAGAGAT CAGCAAGCAGCTGGGATACCGATG GAAAATGTTACTGAAGCCAAAA ATGGCCATTCTTCCAGGAGGCACA GAAATTACAGGCCATGCAAGAGA GAATAACCGAATTATAAGTATCG ACCTCGTCGGAAGGCGAAGATGCT GCCGAAGAATTGCAAGTGTCTCCC GCAGATCCCGCTTCGGTACTCTGCA GCGAAGTGCACGGCAACAGGT TGTACAGGGATGACTGTACGAAAG CCACACACTCAAGAATGGAGCAC AGCTAGGCCACTTACCGCCCATCAA CGCAGCCAGCTCACCGCAGAACG GGACCGCTACAGCCACTGGACAAA GCTG	92	Involved in sex determination and spermatogenesis	Polanco, J. C. & Koopman, P. Sry and the beginnings of male development. <i>Dev. Biol.</i> 302, 13-24 (2007). Koopman, P. et al. Male development of chromosomally female mice transgenic for Sry. <i>Nature</i> 351, 117-121 (1991).
TBX5	ATGGCCGACGCAGACGAGGGCTTT GCCCTGGCGCACAGCCTCTGGAG CCTGACGAAAGACCTGCCCTGC GATTGAAACCCGAGAGCGCCCTC GGGGCCCCCAGCAAGTCCCCGTC TCCCCGCAGGCCCTCACCCAGC AGGGCATGGAGGAATCAAAGTGT TTCTCCATGAAAGAGAACTGTGGCT AAATTCCACGAAACTGGCACCGA AATGATCATAACCAAGGCTGGAG GCGGATGTTCCCAAGTTACAAAGTG AAGGTGACGGGCCCTTAATCCCCAA ACGAAGTACATTCTCATGGACA TTGTACCTGCCGACGATCACAGATA CAAATTGCAGATAATAATGGTCT GTGACGGGCAAGGCTGAGCCGCC ATGCCCTGGCCGCTGTACGTGCA CAGACTCCCCGCCACGGGGCGC ATTGGATGAGGCAGCTCGTCTCTT CCAGAAACTCAAGCTACCAACAA CCACCTGGACCCATTGGGCAATT ATTCTAAATTCCATGCACAAATACC AGCCTAGATTACACATCGTAAAG CGGATGAAAATAATGGATTGGCT CAAATAATCAGCGTTCTGCACTC ACGTCTTCTGAGACTGCGTTTAT AGCAGTGACTTCTTACCAAGCCA CAAGATCACGCAATTAAAGATTGA GAATAATCCCTTGCCAAAGGATT CGGGGCAAGTGTGACATGGAGCTG CACAGAATGTCAGAATGCAAGT AAAGAAATATCCCGTGGTCCCCAGG AGCACCGTGAGGCCAAAAGTGGCC TCCAACCCAGTCCCTTCAGCAGCG AGTCTCGAGCTCTCCACCTCATC CAATTGGGTCCCAATACCAAGTGT GAGAATGGTGTTCGGCCCTCCC AGGACCTCCTGCCCTCACCCACCC ATACCCACTGCCCCAGGAGCATAG CCAAATTACCATGTAAGGAGAGG AAAGAGGAAGAATGTTCCACCA GACCATCCCTATAAGAAGCCCTAC ATGGAGACATCACCCAGTGAAGAA	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. <i>Cell</i> 106, 709-721 (2001).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	GATTCCCTCTACCGCTCTAGCTATC CACAGCAGCAGGGCTGGGTGCT CCTACAGGACAGACTGGCACAGC GGCAAGCTTGCATGTATGCCAGCTC TGCGCCCCCAGCGAGCTGTGCC AGCCTAGAGGACATCAGCTGCAAC ACGTGGCCAAGCATGCCTTCCTACA GCAGCTGCACCGTCACCCACCTGC AGCCCATGGACAGGCTACCCCTACC AGCACTTCTCCGCTCACTTCACCTC GGGGCCCCCTGGTCCCTGGCTGGCT GGCATGGCAACCATGGCTCCCCA CAGCTGGGAGAGGGAATGTTCCAG CACCAGACCTCCGTGGCCACCAAG CCTGTGGTCAGGCAGTGTGGGCTC AGACTGGCCTGCACTGCCCCTGGCAC CCTTCAGCCCCCTGAGTTCTCTAC TCTCATGGCGTGCCAAGGACTCTAT CCCCCTCATCAGTACCACTCTGTGCA CGGAGTTGGCATGGTGCCAGAGTG GAGCGACAATAGCTTG			
TFAP2 C	ATGTTGTGAAAATAACCGATAAT GTCAACTACGAAGGAGACTGGAG GATGCCACGACGGAGCAGCAAT GGGAATCCGGCGGTCCCCAACCTCT CCTCCGGCGGGCAGCACCTACAG CCCCGGGCCACCCCTCTCCACACT GGAGTCGCCGAATATCAGGCCCA CCCTACTTCCCCCTCCCTACCAGC AGCTGGCCTACTCCAGTCGGCCGA CCCCCTACTCGCATCTGGGGGAAGC GTACGCCGCCCATCAACCCCCCTG CACCAGCGGCCACAGGCAGC CAGCAGCAGGCTGGCCGGCC CAGAGCCAGGAGGGAGCGGGGCTG CCCTCGCACCAACGGGGCGCCCGGCC GGCCTACTGCCACCTCTCCGGGC TGGAGGGGGCGCGGTGAGGGCCC GCAGGGATGCCAACCCCGCTCCG ACCTGCTGCTGCCAACGCAACCGC CCTGGATGCCGGGGCTGGCGA GAACCTGGGGCTCCACGACATGCC TCACCAAGATGGACGAGGTGCAAGA TGTGACGACCAAGCACCTGTC CACGATCAGACAGTCATTGCAA GGTCCCATTTCCATGACCAAGAAC CTCTGAACCTCCCTGTCAGAAGGA GCTGGTGGGGCCCGTAATGACCC CACTGAGGTCTTGCTCAGTCC GGAAAGATTGTCGCTCTCAGCTCA CGTCTAAATACAAGTGCAGTGG CTGAAGTACAGAGGCGACTGCCC CACCTGAATGCTAAATGCCTCGT ACTGGGAGGTGTCAGAAGAGC CAAATGAAAAATGGAGGCCGTC CTTGGCGGAGAAGTTGGACAGAT TGGGTGAATCTCCGGCGGGAG GGGAAAGCCGCTCATGTGACTCTC CTGACATCCTTAGTAGAAAGGTGAA GCTGTTCATTTGGCTAGGGATTG CCTATGTCGTGAAGCGAATTCC TAGTAAACCACTGGCAGAATTATT AACCAGACCTCATCTGGAGGACG AAATGAGATGGCAGCTAGGAAGAA CATGCTATTGGCGGCCAGCAACTG TGTAAAGAATTCAACAGAACCTCTCA GCCAAGACCGGACACCCCATGGGA CCAGCAGGCTGCCCTGAGCTTGGGA GACGAACATACAGAAGTGCCTGTCT CATTTAGCTGATTACCCAGGGT TTGGCAGCCAGGCCATCTGTGCGC GGTGTCTGCCCTGCGAAACTACATC AAAGAAGCCCTGATTGTCATAGAC	94	Involved in trophectoderm development	Cao, Z. et al. Transcription factor AP-2 $\gamma$ induces early Cdx2 expression and represses HIPPO signaling to specify the trophectoderm lineage. Development 142, 1606-15 (2015).

TABLE 1-continued

GENE	SEQUENCE	SEQ ID NO:	ROLE	REFERENCES
	AAATCCTACATGAACCTGGAGAC CAGAGTCCAGCTGATTCTAACAA ACCTGGAGAAAATGGAGAAACAC AGGAAA			

TABLE 2

Sample_ID	Description	Media Condition	Estimated Number of Cells	Mean Reads per Cell	Median Genes per Cell
UP_TF_1	HighMOI, (-) TRA-1-60 MACS sorted	Pluripotent stem cell medium	3,640	45,983	3,317
UP_TF_2	HighMOI, Unsorted	Pluripotent stem cell medium	3,505	49,750	3,843
UP_TF_3	HighMOI, Unsorted	Pluripotent stem cell medium	4,223	45,403	3,972
UP_TF_4	HighMOI, (-) TRA-1-60 MACS sorted	Pluripotent stem cell medium	3,461	56,290	4,475
UP_TF_5	LowMOI, (-) TRA-1-60 MACS sorted	Pluripotent stem cell medium	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 medium	2,129	70,519	5,605
UP_TF_11	Library, Endothelial	Endothelial growth medium	6,574	23,250	3,105
UP_TF_12	Library, Multilineage	Multilineage differentiation medium	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	Sequencing Saturation	Fraction Reads in 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_TF_12	141,934,669	98.20%	70.00%	11.00%	88.10%	14,526
UP_TF_13	200,756,922	98.00%	66.20%	15.50%	78.70%	14,286

TABLE 3

Genotype	Number of Genotyped Cells		
	Stem cell media	Endothelial media	Multilineage media
ASCL1	186	78	21
ASCL3	471	150	89
ASCL4	286	90	75

TABLE 3-continued

Genotype	Number of Genotyped Cells		
	Stem cell media	Endothelial media	Multilineage media
ASCL5	140	64	51
ATF7	97	49	45
CDX2	267	192	103

TABLE 3-continued

Genotype	Number of Genotyped Cells		
	Stem cell media	Endothelial media	Multilineage media
CRX	292	107	54
ERG	62	30	7
ESRRG	169	98	64
ETV2	60	22	21
FLI1	55	27	18
FOXA1	53	27	14
FOXA2	89	46	37
FOXA3	255	90	61
FOXP1	413	112	94
GATA1	288	111	72
GATA2	62	81	60
GATA4	71	101	58
GATA6	44	44	35
GLI1	27	11	16
HAND2	310	113	81
HNF1A	88	45	39
HNF1B	53	30	41
HOXA1	166	67	57
HOXA10	344	111	66
HOXA11	237	82	47
HOXB6	166	95	44
KLF4	298	259	145
LHX3	175	76	45
LMX1A	458	155	82
mCherry	1689	689	495
MEF2C	87	49	51
MESP1	227	70	55
MITF	73	63	45

TABLE 3-continued

Genotype	Number of Genotyped Cells		
	Stem cell media	Endothelial media	Multilineage media
MYC	291	113	36
MYCL	356	112	75
MYCN	50	33	12
MYOD1	197	68	40
MYOG	284	122	81
NEUROD1	83	46	10
NEUROG1	154	103	23
NEUROG3	158	138	41
NRL	249	75	49
ONECUT1	159	109	58
OTX2	293	95	47
PAX7	86	56	28
POU1F1	126	61	50
POU5F1	78	30	24
RUNX1	139	47	43
SIX1	260	119	66
SIX2	295	103	84
SNAI2	485	96	50
SOX10	83	54	30
SOX2	137	53	27
SOX3	137	56	31
SPI1	264	142	67
SPIB	199	70	47
SPIC	147	80	35
SRY	166	61	65
TBX5	149	112	35
TFAP2C	90	58	34

TABLE 4

	Enrichment p-value for each genotype in clusters using Fisher's exact test						
	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	0.6444816	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	0.912434	

TABLE 4-continued

	Enrichment p-value for each genotype in clusters using Fisher's exact test						
	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 5

Module	Description	n_genes
GM1	Cytoskeleton and polarity	444
GM2	Ion transport	973
GM3	Chromatin accessibility	1568
GM4	Signaling pathways	873
GM5	Neuron differentiation	444
GM6	Notch pathway	859

TABLE 5-continued

Module	Description	n_genes
GM7	Embryonic development	509
GM8	Mitochondrial metabolism and translation	2242
GM9	Ribosome biogenesis	190
GM10	Growth factor response	492
GM11	Pluripotent state	234

TABLE 6

Gene	Forward Primer (5'→3')	SEQ ID NO:	Reverse Primer (5'→3')	SEQ ID NO:
CDH5	AGACCACGCCTCTGTATGTACCAAATC	95	CACGATCTCATACCTGGCCTGCTTC	113
PECAM1	GGTCAGCAGCATCGTGGTCAACATAAC	96	TGGAGCAGGACAGGTTCACTTTCA	114
VWF	TCTCCGTGGTCCTGAAGCAGACATA	97	AGGTTGCTGCTGGTGAGGTCATT	115
KDR	AGCCATGGTCTCTGGTTGTATG	98	GTTTGAGTGGTGCCGTACTGGTAGGA	116
NANOG	TTTGTGGCCTGAAGAAAAACT	99	AGGGCTGTCTGAATAAGCAG	117
POU5F1	CTTGAATCCGAATGGAAAGGG	100	GTGTATATCCCAGGGTATCCTC	118
SOX2	TACAGCATGTCCTACTCGCAG	101	GAGGAAGAGGTAAACCACAGGG	119
DNMT3B	GAGTCCATTGCTGTTGAAACCG	102	ATGTCCTCTGTCGCCAACCT	120
SALL2	CAGCGGAAACCCCAACAGTTA	103	GAGGGTCAGTAGAACATGCGT	121
DPPA4	GACCTCCACAGAGAAGTCGAG	104	TGCCTTTCTTAGGGCAGAG	122
VIM	AGTCCACTGAGTACCGGAGAC	105	CATTTCACGCATCTGGCGTTC	123
CDH1	CGAGAGCTACACGTTACGG	106	GGGTGTCGAGGGAAAAATAGG	124
CDH2	AGCCAACCTTAACGTGAGGAGT	107	GGCAAGTTGATTGGAGGGATG	125

TABLE 6 -continued

Gene	Forward Primer (5'→3')	SEQ ID NO:	Reverse Primer (5'→3')	SEQ ID NO:
EPCAM	TGATCCTGACTGGCATGAGAG	108	CTTGTCTGTTCTCTGACCCC	126
LAMC1	GGCAACGTGGCTTTCTAC	109	AGTGGCAGTTACCCATTCCCTG	127
SPP1	GAAGTTTCGCAGACCTGACAT	110	GTATGCACCATTCAACTCCTCG	128
THY1	ATCGCTCTCTGCTAACAGTC	111	CTCGTACTGGATGGGTGAACT	129
TPM2	CTGAGACCCGAGCAGAGTTG	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. High-throughput 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 Genome-wide 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 Biol* 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 RNA-seq. *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

ttcctgttgc agataagccc agcttagccc agctgacccc agaccctctc ccctcactcc      60
ccccatgtcg caggatcgag accctgaggg agacagcccg ttccacaagc ccccccggccc      120
gcccccatca ccccgtaaac ttctcccagc ctccgcctg ccctcaccca gcccgtgtt      180
ccccaaagct cgctccaagc ccacgcacc cctgcagcag ggcagcccca gagggcagca      240
cctatccccg aggctggggt cgaggctcgg ccccgccct gcctctgcaa cttgagcctg      300
gctgcgaccc ctgtctgac gtctcgaaa attccccctt gcccaggccc ttgggggagg      360
gggtgcatgg tatgaaatgg ggctgagacc cccggctggg ggcaaggaa cccgcagag      420
aaggagccaa attaggctc tgttccctg atctggact ccaaggggac acgcccacag      480
cgacagcaga gacatgctgg aaaggtacaa gtcatccct ggcaagctc ccacagctgg      540
actggggctc cgcgttactg cacccagaag ttccatgggg ggccggagccc gactctcagg      600
ctcttccgtg gtccggggac tggacagaca tggcgtgcac agcctgggac tcttgagcgc      660
gcgcctcga gaccctgggc cccgcccccc tcggccccggg ccccatcccc gccgcggct      720
ccgaaggcgc cgccggccag aactgcgtcc ccgtgggggg agaggccacc tctgtggcgc      780
gcgcccaggc cgccggggagc aacaccagct gggactgttc tgtggggccc gacggcgata      840
cctactgggg cagtggcctg ggccggggagc cgccgcacgga ctgtaccatt tcgtggggcg      900

```

-continued

---

ggcccgggg cccgactgt accacctct ggaacccggg gtcgtatcg ggtggcacca	960
ccttttcaa gcggtaccag agctcagctc tcaccgttg ctccgaaccg agccgcagt	1020
cgaccgtgc cagttggct cgatgccccaa aactaacca ccgaggccc attcagctgt	1080
ggcagttct cctggagctg ctccacgacg gggcgcgtag cagctgcac cgttggactg	1140
gcaacagccg cgagttccag ctgtgcgacc ccaaaggggt ggctggctg tggggcgagc	1200
gaaagagaaa gcccggcatg aattacgaga agctgagccg gggccttcg tactactatc	1260
gcccgcacat cgtgegcaag agcggggggc gaaaatcac acgttgcgttcc gggggccgcg	1320
tgcggcactt agcctatccg gactgtgcgg gaggcggacg gggagcagag acacaataaa	1380
aattcccggt caaacctcaa aaaaaaaaaaaa aaa	1413

<210> SEQ ID NO 2  
<211> LENGTH: 1275  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 2	
atgccccctca acgttagctt caccaacagg aactatgacc tcgactacga ctgggtgcag	60
ccgtatttct actgcgcacga ggaggagaac ttctaccaggc agcagcgcac gagcgagctg	120
cagccccccgg cgggatcagg tagcggtagc cgccgcgtcc ggctctgcgc gcccctctac	180
gttgcggtaa cacccttctc ctttcggggaa gacaacgcac ggggtggccgg gagttctcc	240
acggccgacc agctggagat ggtgaccgag ctgctggag gagacatggt gaaccagagt	300
ttcatctgcg accccggacga cgagaccttc atcaaaaaca tcatcatcca ggactgtatg	360
tggagcggct tctcgccgcg cgccaagctc gtctcagaga agctggctc ctaccaggct	420
gccccggaaag acagcggcag cccgaacccc gcccgcggcc acagcgtctg ctccaccc	480
agcttgtacc tgcaggatct gagcgcgcgc gcctcagagt gcatgcaccc ctgggtggc	540
ttccccctacc ctctcaacga cagcagctcg cccaaatctc ggcctcgca agactccac	600
gccttctctc cgtcctcgaa ttctctgcgc tcctcgcacgg agtcctcccc gcaggccac	660
cccgagcccc tggtgctcca tgaggagaca ccgccccacca ccagcagcga ctctgaggag	720
gaacaagaag atgagaaaga aatcgatgtt gtttctgtgg aaaagggca ggctctggc	780
aaaagggtcag agtctggatc accttctgtct ggaggccaca gcaaacctcc tcacagccca	840
ctgggtccctca agaggtgccca cgtctccaca catcagcaca actacgcacgc gcctccctcc	900
actcggaaagg actatctgc tgccaagagg gtcaagttgg acagtgtcag agtctgtgaga	960
cagatcagca acaaccgaaa atgcaccgcg cccaggtctt cggacaccga ggagaatgtc	1020
aaggaggcga cacacaacgt ctggagcgc cagaggagga acgagctaa acggagctt	1080
ttggccctgc gtgaccagat cccggagttg gaaaacaatg aaaaggcccc caaggtatgtt	1140
atccctaaaa aagccacagc atacatcctg tccgtccaag cagaggagca aaagctcatt	1200
tctgaagagg acttggcgcg gaaacgcacga gaacagttga aacacaaact tgaacagcta	1260
cggaactctt gtgcg	1275

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

---

-continued

---

```

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

<400> SEQUENCE: 3

atgccccctca acgttagctt caccaacagg aactatgacc tcgactacga ctgggtcag 60
ccgtattttct actgcgcacga ggaggagaac ttctaccagg agcagcagca gagegagctg 120
cagccccccgg cgcccagcga ggatatctgg aagaaattcg agctgctgcc caccggccc 180
ctgtcccccta gccgcgcgtc cgggctctgc tcgccccctt acgttgccgt cacacccttc 240
tcccttcggg gagacaacga cggcggtggc gggagcttct ccacggccga ccagctggag 300
atggtgaccg agctgctggg aggagacatg gtgaaccaga gtttcatctg cgaccggac 360
gacgagacct tcatcaaaaa catcgatca ggtagcggtc tctgtctcaga gaagctggcc 420
tccattaccagg ctgcgcgcaa agacagcggc agcccgaaacc ccggccggcc ccacagcgctc 480
tgcgtccacccat ccagttgtta cctgcaggat ctgagcgcggc ccgcctcaga gtgcatacgac 540
ccctcgggtgg tttttccctca ccctctcaac gacagcagct ccgcctcaagtc ctgcgcctcg 600
caagactcca gcgccttc tccgtctcg gattctctgc tctcctcgac ggagtccctcc 660
ccgcaggggca gcccgagcc cctgggtgtc catgaggaga caccgeccac caccagcagc 720
gactctgagg aggaacaaga agatgaggaa gaaatcgatg ttgtttctgt ggaaaagagg 780
caggctctg gcaaaagggtc agagtctgga tcaccttctg ctggaggcca cagcaaacct 840
cctcacacgcc cactggtc tcaagagggtc cacgtctcca cacatcaga caactacgca 900
ggccctccct ccactcgaa ggactatctt gctgccaaga gggtaagggtt ggacagtgtc 960
agagtcttga gacagatcg caacaacgg aaatgcacca gcccagggtc ctggacacc 1020
gaggagaatg tcaagaggcg aacacacaac gtcttggagc gccagaggag gaacgagacta 1080
aaacggagct tttttccct gcgtgaccag atcccgaggat tgaaaacaa tgaaaaggcc 1140
cccaaggtag ttatccttaa aaaagccaca gcatacatcc tgcgtccgtccca agcagaggag 1200
caaaagctca tttctgaaga ggacttggtg cggaaacgc gagaacagtt gaaacacaaa 1260
cttgaacagc tacggaaactc ttgtgcg 1287

```

```

<210> SEQ_ID NO 4
<211> LENGTH: 1305
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

```

<400> SEQUENCE: 4

atgccccctca acgttagctt caccaacagg aactatgacc tcgactacga ctgggtcag 60
ccgtattttct actgcgcacga ggaggagaac ttctaccagg agcagcagca gagegagctg 120
cagccccccgg cgcccagcga ggatatctgg aagaaattcg agctgctgcc caccggccc 180
ctgtcccccta gccgcgcgtc cgggctctgc tcgccccctt acgttgccgt cacacccttc 240
tcccttcggg gagacaacga cggcggtggc gggagcttct ccacggccga ccagctggag 300
atggtgaccg agctgctggg aggagacatg gtgaaccaga gtttcatctg cgaccggac 360
gacgagacct tcatcaaaaa catcatcatc caggactgtt tgcgtccgtccca 420

```

---

-continued

---

gcccggcaagg tcgttcaga gaagctggcc tcctaccagg ctgcggcggaa agacagcgcc	480
agccccgaacc cccggccggcc acacagcgtc tgctccacct ccagcttgc cctgcaggat	540
ctgagcgccg cccgcctcaga gtgcacatcgac ccctcggtgg tcttccctta ccctctcaac	600
gacagcagct cggcccaagtc ctgcgcctcg caagactcca ggcgccttc tccgtcctcg	660
gattctctgc tctccctcgc acggatccctcc cccgcaggggca gccccgagcc cctgggtgctc	720
catgaggaga caccgcccac caccaggcgc gactctgagg aggaacaaga agatgaggaa	780
gaaatcgatg ttgtttctgt ggaaaagagg caggctctcg gaaaaagggtc agagtctgga	840
tcacccctctg ctggaggcca cagcaaacct cctcacacgc cactggtc tcaagagggtc	900
cacgtctcca cacatcagca caactacgc ggcgcctccct ccactcgaa ggactatgga	960
tcaggtagcg gtagtgtcag agtcctgaga cagatcagca acaaccgaaa atgcaccacg	1020
cccgaggctt cggacacccga ggagaatgtc aagaggcgaa cacacaacgt ctggagcgcc	1080
cagaggagga acgagctaaa acggagcttt ttggccctgc gtgaccagat cccggagttg	1140
gaaaacaatg aaaaggcccc caaggtagtt atccctaaaa aagccacagc atacatcctg	1200
tccgtccaag cagaggagca aaagctcatt tctgaagagg acttgggtcg gaaacgcacga	1260
gaacagttga aacacaaact tgaacagcta cggaactctt gtgcg	1305

<210> SEQ\_ID NO 5  
<211> LENGTH: 1290  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 5	
atgccccctca acgttagctt caccacagg aactatgacc tcgactacga ctgggtcg	60
ccgttattct actgcgcacga ggaggagaac ttctaccagg agcagcagca gagcgagctg	120
cagccccgg cggccacgcg ggatatctgg aagaaattcg agctgtcgcc caccggccc	180
ctgtccctca ggcgcgcgc cgggcctcg tccgcctcc acgttgccgtt caccggccc	240
tcccttcggg gagacaacga cggcggtggc gggagcttcc caccggccgca ccagctggag	300
atggtgaccc agctgtggg aggagacatg gtgaaccaga gtttcatctg cgaccggac	360
gacgagaccc tcatcaaaaa catcatcatc caggactgta tgtggagccg cttctcgcc	420
ggccggcaagg tcgttcaga gaagctggcc tcctaccagg ctgcggcggaa agacagcgcc	480
agccccgaacc cccggccggcc ccacagcgtc tgctccacct ccagcttgc cctgcaggat	540
ctgagcgccg cccgcctcaga gtgcacatcgac ccctcggtgg tcttccctta ccctctcaac	600
gacagcagct cggcccaagtc ctgcgcctcg caagactcca ggcgccttc tccgtcctcg	660
gattctctgc tctccctcgc acggatccctcc cccgcaggggca gccccgagcc cctgggtgctc	720
catgaggaga caccgcccac caccaggcgc gactctgagg aggaacaaga agatgaggaa	780
gaaatcgatg ttgtttctgt ggaaaagagg caggctctcg gaaaaagggtc agagtctgga	840
tcacccctctg ctggaggcca cagcaaacct cctcacacgc cactggtc tcaagagggtc	900
cacgtctcca cacatcagca caactacgc ggcgcctccct ccactcgaa ggactatcct	960
gctggcaaga gggtaagtt ggacagtgtc agagtctga gacagatcag caacaacgc	1020

---

-continued

---

aaatgcacca gccccaggc ctcggacacc gaggagaatg tcggatcagg tagcggtgag	1080
ctaaaacgga gctttttgc cctgcgtgac cagatcccg agttggaaaa caatgaaaag	1140
gcccccaagg tagttatcct taaaaaaagcc acagcataca tcctgtccgt ccaagcagag	1200
gagcaaaagc tcatttctga agaggacttg ttgcggaaac gacgagaaca gttgaaacac	1260
aaacttgaac agctacggaa ctcttgcg	1290

```

<210> SEQ ID NO 6
<211> LENGTH: 1206
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

<400> SEQUENCE: 6	
atgccccta acgttagctt caccaacagg aactatgacc tcgactacga ctcggcag	60
ccgtatttct actgcgacga ggaggagaac ttctaccaggc agcagcagca gagcgagctg	120
cagccccccgg cgccccaggc gaatatctgg aagaaattcg agctgtgcc caccggccc	180
ctgtccctta gccgcccgtc cgggctctgc tcgccccttc acgttgcggt cacaccctc	240
tcccttcggg gagacaacga cggcggtggc gggagcttcc ccacggccga ccagctggag	300
atggtgaccg agctgtggg aggagacatg gtgaaccaga gtttcatctg cgaccggac	360
gacgagacct tcataaaaa catcatcatc caggactgta tgtggagccg cttctcgcc	420
gccgccaagc tcgtctcaga gaagctggcc tcctaccagg ctgcgcgcaa agacagccg	480
agcccgaaacc cccggccggg ccacagcgtc tgctccacct ccagctgtc cctgcaggat	540
ctgagcgccg ccgcctcaga gtgcattcgac ccctcggtgg tttccctta ccctctcaac	600
gacagcagct cgcccaagtc ctgcgcctcg caagactcca gcgccttc tccgtctcg	660
gattctctgc tctctcgac ggagtctcc cccgaggcga gcccggagcc cctgggtctc	720
catgaggaga caccggccac caccaggcgc gactctgagg aggaacaaga agatgaggaa	780
gaaatcgatg ttgttctgt gggaaagagg caggctctg gaaaaaggc agagtctgga	840
tcaccttctg ctggaggcca cagcaaacct cctcacagcc cactggctt caagaggc	900
cacgtctcca cacatcagca caactacgca gcgcctccct ccactcgaa ggactatct	960
gctgccaaga gggtaagtt ggacagtgtc agagtctgaa gacagatcga caacaaccga	1020
aaatgcacca gccccaggc ctcggacacc gaggagaatg tcaagaggcg aacacacaac	1080
gtcttggaggc gccagaggag gaacggatca ggttagcggtc aaaagctcat ttctgaagag	1140
gacttggc ggaaacgacg agaacagttt aaacacaaac ttgaacagct acgaaactct	1200
tgtgcg	1206

```

<210> SEQ ID NO 7
<211> LENGTH: 1230
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

<400> SEQUENCE: 7	
atgccccta acgttagctt caccaacagg aactatgacc tcgactacga ctcggcag	60

-continued

---

cctgtatttct actgcgacga ggaggagaac ttctaccagc agcagcagca gagcgagctg	120
cagccccccgg cgcccagcga ggatatctgg aagaaattcg agctgtgcc caccggccc	180
ctgtcccccta gcccggcgtc cggggctctgc tcggccctct acgttgccgt cacacccttc	240
tcccttcggg gagacaacga cggcggtggc gggagcttct ccacggccga ccagctggag	300
atggtgacgg agctgtggg aggagacatg gtgaaccaga gtttcatctg cgaccggac	360
gacgagacct tcataaaaaa catcatcatc caggactgta tgtggagccg cttdtcggcc	420
gcccggccaa cccggccggcc ccacagcgcc tggccacccct ccagcttgta cctgcaggat	480
ctggccggcc cccggccctcaga gtgcatacgac ccctcggtgg tttttccctt cccttcacac	540
gacagcagct cggccaaatgc ctggccctcg caagactcca gggcccttc tccgtccctcg	600
gatttcttcgc ttccttcgac ggagtccctcc cccggggcc gccccggcc cctgggtgtc	660
catgaggaga caccggccac caccagcagc gactctgagg aggaacaaga agatgaggaa	720
gaaatcgatg ttgtttctgt ggaaaaggagg caggctctg gaaaaagggtc agagtctgga	780
tcaccttcgt ctggaggcca cagcaaacct cctcacacgc cactggctt caagagggtc	840
cacgttcacca cacatcagca caactacgc gggccctccctt ccactggaa ggactatcc	900
gtgtccaaaga gggtaagtt ggacagtgtc agagtctgta gacagatcag caacaaccga	960
aaatgcacca gccccaggcc ctcggacacc gaggagaatg tcaagaggcc aacacacaa	1020
gtcttggagc gccagaggag gaacgagctt aaacgggact ttttgcctt gcgtgaccag	1080
atcccgagtt tgaaaacaa tgaaaaggcc cccaaaggtag ttatccttaa aaaaggccaca	1140
gcatacatcc tgcgtccca agcagaggag	1200
	1230

<210> SEQ ID NO 8  
<211> LENGTH: 906  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 8

atgggatcag gtageggctc cgttcagag aagctggccctt accaccaggc tgcgcgc	60
gacagcggca gccccaaatcc cggccggccgc cacagcgctc gttccacccctc cagttgtac	120
ctgcaggatc tgagcgccgc cgcctcagag tgcatacgacc cctcggtggt tttttccctac	180
cctctcaacg acagcagctc gccaaatgtcc tggccctcgcc aagactccag cgccttc	240
ccgtccctcg attctctgtc ttcctcgacg gatgtccccc cgcaggggccag ccccgagccc	300
ctgggtgtcc atgaggagac accggccacc accagcagcag actctgaggaa ggaacaagaa	360
gtatggaaag aaatcgatgt ttgtttctgt gaaaaggaggc aggctctgg caaaaagggtca	420
gatgtctggat caccttcgtc tggaggccac agcaaacctc ctcacagccc actggcttc	480
aagagggtgcc acgttcacca acatcagcac aactacgcag cgcctccctc cactcgaaag	540
gactatcctg ctggcaagag ggtcaagttt gacagtgtca gatgtccgtac acagatcagc	600
aacaaccgaa aatgcaccaat ccccgaggcc tccggacaccg aggagaatgt caagaggccg	660
acacacaacg tcttggagcc ccaggaggagg aacgagctaa aacggagctt ttttgcctg	720
cgtgaccaga tcccgagtt gaaaacaat gaaaaggccc ccaaggtagt tatccttaaa	780

---

-continued

---

aaagccacag catacatcct gtccgtccaa gcagaggagc aaaagtcat ttctgaagag	840
gacttggc ggaaacgcgc agaacatgg aaacacaac 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: Synthetic polynucleotide

<400> SEQUENCE: 9

atgcccctca acgttagctt caccaacagg aactatgacc tcgactacga ctgggtgcag	60
ccgtatttct actgcgcacga ggaggagaac ttctaccaggc agcagcagca gagcgagctg	120
cagccccggc cgcccgacga ggatatctgg aagaaattcg agctgctgcc caccggccc	180
ctgtccctcta ggcgcgcgctc cgggctctgc tcgccttc acgttgccgtt cacacccttc	240
tcccttcggg gagacaacga cggcggtggc gggagcttcc caacggccga ccagctggag	300
atggtgaccg agctgctggg aggagacatg gtgaaccaga gtttcatctg cgaccggac	360
gacgagacct tcatcaaaaa catcatcatc caggactgta tgtggagccg ctttcggcc	420
gcccggcaagg tcgtctcaga gaagctggcc tcctaccagg ctgcgcgcggc agacaggcc	480
agcccgaaacc cggccggcgcc ccacagcgctc tgctccacctt ccagcttgc cctgcaggat	540
ctgagcgccg cggccctcaga gtgcacatcgac ccctcggtgg tcttccctta ccctctcaac	600
gacagcagct cgcccaagtc ctgcgcctcg caagactcca ggcgccttc tcggctctcg	660
gattctctgc tctccctcgc acggacttcc cccgcaggcc gcccggccg cctgggtgtc	720
catgaggaga caccggccac caccagcgc gactctgagg aggaacaaga agatgaggaa	780
gaaatcgatgt tggttctgtt ggaaaaggagg caggctctg gaaaaagggtc agagtctggaa	840
tcaccttctg ctggaggcca cagcaaacctt cctcacagcc cactggcttcaagagggtc	900
cacgtctcca cacatcagca caactacgc ggcgccttc ccactcgaa ggactatcc	960
gctgccaaga gggtaagtt ggacagtgtc agagtctgaa gacagatcag caacaaccga	1020
aaatgcacca gccccaggc ctcggacacc gaggagaatgtc	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: Synthetic polynucleotide

<400> SEQUENCE: 10

atgcccctca acgttagctt caccaacagg aactatgacc tcgactacga ctgggtgcag	60
ccgtatttct actgcgcacga ggaggagaac ttctaccaggc agcagcagca gagcgagctg	120
cagccccggc cgcccgacga ggatatctgg aagaaattcg agctgctgcc caccggccc	180
ctgtccctcta ggcgcgcgctc cgggctctgc tcgccttc acgttgccgtt cacacccttc	240
tcccttcggg gagacaacga cggcggtggc gggagcttcc caacggccga ccagctggag	300
atggtgaccg agctgctggg aggagacatg gtgaaccaga gtttcatctg cgaccggac	360

---

-continued

---

gacgagacct tcataaaaa catcatcatc caggactgta tgtggagccg cttctcgcc	420
gccgccaagg tcgttcaga gaagctggcc tcctaccagg ctgcgcgcaa agacagccg	480
agccccgaacc ccgccccggg ccacagcgcc tgctccacct ccagcttgta cctgcaggat	540
ctgagcgccg ccgcctcaga gtgcattcgac ccctcggtt tttccctta ccctctcaac	600
gacagcagct cgcccaagtc ctgcgcctcg caagactcca ggcgcctctc tccgtcctcg	660
gattctctgc tctcttcgac ggagtccctcc cgcgcaggcc gccccgagcc cctgggtgtc	720
catgaggaga caccgccccac caccaggcgc gactctgagg aggaacaaga agatgaggaa	780
gaaaatcgat ttgtttctgt ggaaaagagg caggctcctg gaaaaaggcc agagtctgga	840
tcaccttctg ctggaggcca cagcaaacct cctcacagcc cactggtcct caagagggtgc	900
cacgtctcca cacatcagca caactacgca ggcgcctccct ccactcgaa ggactatcct	960
gtgtgcaaga gggtaagtt ggacagtgtc agagtccctgaa gacagatcag caacaaccga	1020
aaatgcacca gccccaggcc ctcggacacc gaggagaatg tcaagaggcc aacacacaac	1080
gtcttggagc gccagaggag gaacgagactt aaacggagct ttttgcct gcgtgaccag	1140
atcccgaggat tggaaaacaa tgaaaaggcc cccaaaggtag ttatccttaaaaagccaca	1200
gcatacatcc tgcgttccca agcagaggag caaaagctca ttctgaaga ggacttgg	1260
cgaaaacgac gagaacagtt gaaaacacaaa cttgaacagc tacggactc ttgtgcg	1317

<210> SEQ\_ID NO 11  
<211> LENGTH: 1317  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 11	
atgcgcctca acgttagctt caccaacagg aactatgacc tcgactacga ctgcggcag	60
ccgtatttct actgcgcacga ggaggagaac ttctaccagg agcgcgcga gagcgagctg	120
cagccccccgg cgccccaggca ggttatctgg aagaaattcg agctgtgcc cgccccggcc	180
ctgtccctcta gccgcgcgtc cgggcctctgc tcgcgcctct acgttgcgtt cacaccctc	240
tcccttcggg gagacaacga cggcggtggc gggagcttcc ccacggccga ccagctggag	300
atggtgaccg agctgctggg aggagacatg gtgaaccaga gtttcatctg cgaccggac	360
gacgagacct tcataaaaa catcatcatc caggactgta tgtggagccg cttctcgcc	420
gccgccaagg tcgttcaga gaagctggcc tcctaccagg ctgcgcgcaa agacagccg	480
agccccgaacc ccgccccggg ccacagcgcc tgctccacct ccagcttgta cctgcaggat	540
ctgagcgccg ccgcctcaga gtgcattcgac ccctcggtt tttccctta ccctctcaac	600
gacagcagct cgcccaagtc ctgcgcctcg caagactcca ggcgcctctc tccgtcctcg	660
gattctctgc tctcttcgac ggagtccctcc cgcgcaggcc gccccgagcc cctgggtgtc	720
catgaggaga caccgccccac caccaggcgc gactctgagg aggaacaaga agatgaggaa	780
gaaaatcgat ttgtttctgt ggaaaagagg caggctcctg gaaaaaggcc agagtctgga	840
tcaccttctg ctggaggcca cagcaaacct cctcacagcc cactggtcct caagagggtgc	900
cacgtctcca cacatcagca caactacgca ggcgcctccct ccactcgaa ggactatcct	960

---

-continued

---

gctgccaaga gggtaagtt ggacagtgtc agagtccgtga gacagatcag caacaaccga	1020
aaatgcacca gccccagggtc ctggacacc gaggagaatg tcaagaggcg aacacacaac	1080
gtcttggagc gccagaggag gaacgagcta aaacggagct ttttgccct gcgtgaccag	1140
atcccgagt tgaaaacaa tgaaaaggcc cccaaggtag ttatcctaa aaaagccaca	1200
gcatacatcc tgtccgtcca agcagaggag caaaagctca tttctgaaga ggacttgg	1260
cggaaacgac gagaacagtt gaaacacaaa cttgaacagc tacggaactc ttgtgcg	1317

```

<210> SEQ ID NO 12
<211> LENGTH: 1317
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

```

```

<400> SEQUENCE: 12
atgccccctca acgttagctt caccaacagg aactatgacc tcgactacga ctcgggtcag 60
ccgtatttct actgcgacga ggaggagaac ttctaccgc agcagcagca gagcgagctg 120
cagccccccgg cgcccagcga ggatatctgg aagaaattcg agctgctgcc caccggccc 180
ctggcccccta gcccggcgtc cgggctctgc tcgcccctct acgttgccgt cacaccctc 240
tcccttcggg gagacaacga cggcggtggc gggagcttcc ccaacggccga ccagctggag 300
atggtgaccg agctgctggg aggagacatg gtgaaccaga gtttcatctg cgaccggac 360
gacgagacct tcatcaaaaa catcatcatc caggactgtatgtggagccggttcc 420
gcccggcaagg tcgtctcaga gaagctggcc tcctaccagg ctgcgcgcggc agacagccgc 480
agcccgaaacc cccggccggcc ccacagcggtc tgctccacctt ccagcttgcgccttgc 540
ctgagcgccg cccgctcaga gtgcattcgc ccctcggtgg tcttccctca ccctctcaac 600
gacagcagct cggccaaatgc ctgcgcctcg caagactcca ggccttc tccgtctcg 660
gattctctgc tctcctcgac ggagtcctcc cgcaggccgc gccccggcc cctgggtgtc 720
catgaggaga caccggccac caccaggcgc gactctgagg aggaacaaga agatgaggaa 780
gaaatcgatg ttgtttctgt ggaaaaggagg caggctctgc gaaaagggtc agagtctgg 840
tcaccttctg ctggaggccca cagcaaacctt cctcacagcc cactggcttcaagagggtc 900
cacgtctcca cacatcagca caactacgc ggcctccctt ccactcgaa ggactatcc 960
gtgccaaga gggtaagtt ggacagtgtc agagtccgtga gacagatcag caacaaccga 1020
aaatgcacca gccccagggtc ctggacacc gaggagaatg tcaagaggcg aacacacaac 1080
gtcttggagc gccagaggag gaacgagcta aaacggagct ttttgccct gcgtgaccag 1140
atcccgagt tgaaaacaa tgaaaaggcc cccaaggtag ttatcctaa aaaagccaca 1200
gcatacatcc tgtccgtcca agcagaggag caaaagctca tttctgaaga ggacttgg 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

```

---

-continued

---

&lt;400&gt; SEQUENCE: 13

atggcgactg	cgaggacagc	acttccatca	atctcaaacac	tcaactgcact	ggggccattt	60
ccagataccc	aggacgattt	ccttaagtgg	tggcggtccg	aagaggctca	agacatggga	120
cctggtccgc	cggatcccac	cgaacctct	ctgcatgtca	aaagtgaaga	tcagectggc	180
gaggaagagg	atgacgaaag	gggtgccac	gccacttggg	acttggatct	tctcattacc	240
aattttctg	gtccggAACC	tggggggca	ccacagacgt	gcgcctcgc	tccctcagaa	300
gccccgggg	ctcagtaccc	accccctccc	gaaaactctgg	gagcctatgc	tgggggtcct	360
ggactggtgg	ctgggttgc	tggtagtgag	gaccattctg	gtgggtacg	ccccgcttg	420
aggggcccg	ctccggacgc	ctttgtggg	ceggcgctcg	ctccgtcacc	ggctccggaa	480
ccaaaagccc	tcgcgtgca	gcccggtgtac	cccgaccccg	gagccggatc	ctcaggggaa	540
tacttccac	ggacccgact	cagcggtcca	ceggcttccg	gggcgcctata	cggttgcgg	600
agcggttacc	oggctatgt	tcccgctccc	cagtaccaag	gacacttcca	attgttccgg	660
ggtcttcaag	ggcctgcgcc	cggcctgtgt	accagttccca	gtttcctcag	ttgtctgggaa	720
cggggactg	ttggcactgg	acttggcggg	actgcagagg	accaggcgt	tatagcagag	780
acagcgccaa	gtaaaagggg	ccgacgaac	tggggccagg	aacgcctaa	tgccgtcacact	840
tgtccccatc	cagggtgcgg	taaatccatc	acgaagagca	gtcatcttaa	agcacatctt	900
cgcacacaca	cggggagagaa	gccctacgac	tgtacttggg	aaagggtgcgg	ctggagattc	960
gttagatctg	acgagctcac	ccggcattat	cgaaaacaca	ctggccagcg	accgttccgg	1020
tgccaaactct	gccccagggc	gtttagtgc	tcagatcatc	tggctttgca	tatgaaggcga	1080
caccc						1086

&lt;210&gt; SEQ ID NO 14

&lt;211&gt; LENGTH: 1065

&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 polynucleotide

&lt;400&gt; SEQUENCE: 14

atggccctta	gtgaacccat	tcttcccagc	ttttccacgt	tgcgtctcc	ttggcgagag	60
agaggccttc	aggaaaagggt	gcccggggct	gaacccgagt	ctggaggtac	ggatgtatgt	120
cttaacagtg	tgctcgattt	catactctca	atgggactgg	acgggctggg	agcggaggca	180
gtccctgaac	caccaccacc	ccctccgc	ccagcgcccc	actacccgg	gccaggtgcg	240
ccggccgcac	attcagcccc	ggcggtggc	ttgggttccg	agcttccctcg	gcctgaattt	300
gtggccgcgc	tcggcccgcc	gctgcattgt	agatttctgc	tgcgtctcc	gggtcgactc	360
gttaaggctg	aacctcctga	ggctgtatgt	ggagggtggct	acggatgtgc	ccccgggctt	420
acccgaggac	cgagaggctt	taagcggaa	ggggcacctc	gcccggctgc	aagctgtatg	480
cggggggcccg	gtggggaggcc	tccccgc	cctgatacac	ccccctttag	tccagatgg	540
ccagctcgac	ttcccgaccc	tggcccccaga	gcgagtttcc	ccctccatt	tggaggaccg	600
gggtttggcg	ccccagggtcc	tggacttcac	tacggccctc	ctggccccc	agcttttgt	660
cttttcgacg	atgctgtgc	tgccgcagca	gccttggcc	ttgcgcgc	cgcagccagg	720
ggactgtca	cgcacccggc	aagccccctg	gagctccttg	aagccaagcc	gaagcgagga	780

---

-continued

---

cgcagatcat ggccgogcaa gcccacagct acgcataacct gtcataatgc gggctgcgga	840
aaaacacctaca caaagagttc acacctaaa ggcacccctc gcacacacac aggcgagaaa	900
ccatattcatt gtaactggga cggatgttga tggaaatttg ctccgtctga tgagcttacg	960
agacattatc gaaagcatac cggacategg cccttcaat gccatcttg tgacagagct	1020
tttccccgt ctgaccacct cgctctgcac atgaagaggc acatg	1065

```

<210> SEQ ID NO 15
<211> LENGTH: 1035
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

<400> SEQUENCE: 15	
atgctcatgt ttgaccagt tcctgtcaag caagaggcca tggaccctgt ctcatgtca	60
tacccatcta attacatgga atccatgaag cctaacaagt atggggtcat ctactccaca	120
ccattgctg agaagttctt tcagacccca gaaggctgtcg cgcacggaaat acagatggag	180
ccagtgacc tcacggtaa caagcggagt tcacccctt cggctggaa ttcgcctcc	240
tctctgaagt tcccttcctc acacggaga gcctcgcctg ggttgagcat gccttctcc	300
agcccacccga taaaaaaaata ctcacccctt tctccaggcg tgcageccctt cggcgtgccg	360
ctgtccatgc caccagtgtat ggcagctgcc ctctcgeggc atggaatacg gagcccgggg	420
atcctgcggc tcatccagcc ggtgggtgtc cagcccttc cctttatgtt cacaagtac	480
ctccagcgc ctctcatggt ctccctatcg gaggagatgg aaaattccag tagtagcatg	540
caagtacctg taatttaatc atatgagaag cctatatac agaaaaaaaaat taaaatagaa	600
cctggatcg aaccacagag gacagattat tattctgtcaag aaatgtcacc ccccttaatg	660
aactcgtgtt cccccccgcgca agcattgttg caagagaatc acccttcgggt catcgtcag	720
cctggaaaga gaccttacc tttggaaatcc cggataactc aaaggaagcg gaggatacac	780
agatgtgattt atgatggatg caacaatgt tacactaaaa gtccttgcattt gaaagcacac	840
agaagaacac acacaggaga aaaacccctac aaatgtacat gggaaagggtg cacatggaa	900
tttgcgttgtt ctgtatgtt aacaagacat ttccgaaaatc atacttggaaat caaaccttc	960
cagtccccgg actgtgaccg cagttctcc cgttctgacc atcttgcctt ccataggaaa	1020
cggccacatgc tagtc	1035

```

<210> SEQ ID NO 16
<211> LENGTH: 1371
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

<400> SEQUENCE: 16	
atggctacaa ggggtgtgag catgagcgcc cgcctggac ccgtgcggccca gcccggcg	60
ccgcaggacg agccgggttt cgcgcagtc aagccgggtc tggggccgcgaatccggcc	120
cgcgcacggcgg cgctttccc cggcgaggag ctgaagcagc cgcaccacccg cccgcaggcg	180
cagcccgccg cccgcaggc cccgcaggccg gcccagccg cccgcaccgg cccgcggctg	240

---

-continued

---

cctccagagg acctggtcca gacaagatgt gaaatggaga agtatctgac acctcagctt	300
cctccagttc ctataattcc agagcataaa aagtatacgac gagacagtgc ctcagtcgta	360
gaccagttct tcactgacac tgaagggtta cttacatgtca tcaacatgaa cgtttcttc	420
cctgacatca ctcacccgtac aactggcctc tacaaatccc agagaccgtg cgtaacacac	480
atcaagacag aacctgttgc catttcagc caccagatgt aaacgactgc ccctctccg	540
gccccgaccc aggccctccc tgagttcacc agtatattca gtcacacca gaccgcagct	600
ccagaggtga acaatatttt catcaaacaa gaacttccta caccagatct tcattttct	660
gtccctaccc agcaggccca cctgttccac ctactgaata caccggatct agatatgcc	720
agttctacaa atcagacagc agcaatggac actcttaatg tttctatgtc agctgccatg	780
gcaggcctta acacacacac ctctgctgtt ccgcagactg cagtgaaaca attccaggc	840
atggcccttgc acatacacac aatgcacatc cagtttcttc cacaacagc cacttactt	900
cccccgctac caccaagctc agagcctgga agtccagata gacaaggcaga gatgtccag	960
aatttaaccc cacccatccatc ctatgctgtc acaatttgctt ctaaactggc aattcacaat	1020
ccaaattttac ccaccaccct gccagttaac tcacaaaaca tccaacctgt cagatacaat	1080
agaaggagta accccgattt ggagaaacga cgcacatccact actgcgatcc cccctgggtgc	1140
acaaaatgtt ataccaagtc ttctcatatc aaagctcacc tgaggactca cactggtgaa	1200
aagccatatac agtgtacactg ggaaggctgc gactggaggt tcgcgcgatc ggtgagctg	1260
acccggccact accggaagca cacaggcgcc aagcccttcc agtgcggggt gtgcaaccgc	1320
agtttctcgc gctctgacca cctggccctg catatgaaga ggcaccagaa c	1371

<210> SEQ\_ID NO 17  
<211> LENGTH: 849  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 17

atggacgtgc tccccatgtc cagcatcttc caggagctcc agatcggtca cgagaccggc	60
tacttctcg cgctgcccgc tctggaggag tactggcaac agacctgcct agagctggaa	120
cgttacccctcc agagcgagcc ctgttatgtt tcagcctcag aaatcaaatt tgacagccag	180
gaagatctgt ggacaaaaat cattctggct cgggagaaaa aggaggaatc cgaactgaag	240
atatcttcca gtcctccaga ggacactctc atcagcccgaa gttttgtta caacttagag	300
accaacagcc tgaactcaga tgcacgcgcga gaatcctctc acagctccgc ggaactttct	360
cccacggccca agtttacctc cgacccattt ggcgaagttt tggcagctc gggaaaattt	420
agtcctctg tcacccatct gcctccatct tctccggaaac tgagcaggaa accttctcaa	480
ctgtgggggtt gggtggccgg ggagctgccc tcgcgcggaa aggtgcgcag cgggacttcg	540
gggaagccag gtgacaaggaa aaatggcgat gcctccccc acggcaggag gaggggtgcac	600
cgggtgcact ttaacggctg caggaaatgt tacacaaaaa gttccactt gaaaggcacac	660
cagcggacgc acacaggaga aaagccctac agatgctcat gggaaagggtg tgagtggcgt	720
tttgcaagaa gtgatgagtt aaccaggcac ttccgaaagc acaccggggc caagccttt	780

---

-continued

---

aaatgctccc actgtacag gtgtttcc aggtctgacc acctggccct gcacatgaag	840
aggcacctc	849

```

<210> SEQ_ID NO 18
<211> LENGTH: 906
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

<400> SEQUENCE: 18	
atggacgtgt tggtagtta tagtatattc caggagctac aacttgtcca cgacaccggc	60
tacttctcag cttaaccatc cctggaggag acctggcgc agacatgcct tgaattggaa	120
cgcctacccatc agacggagcc ccggaggatc tcagagaccc ttggtgagga cttggactgt	180
ttccctccacg ctccccctcc cccgtgcatt gaggaaagct tccgtcgctt agacccctg	240
ctgctccccg tggaagcggc catctgtgag aagagctcg cagtggacat cttgtctct	300
cgggacaagt tgctatctga gacctgcctc agcctccgc cggccagctc ttctctagac	360
agctacacag ccgtcaacca ggcccagctc aacgcagtga cctcattaac gccccatcg	420
tccccctgagc tcagccgcca tctggtcaaa acctcacaaa ctctctctgc cgtggatggc	480
acgggtacgt tgaaactggt ggcaagaag gctgctcta gctcgtaaa ggtggggaggg	540
gtcgcaacag ctgcagcagc cgtgacggct gcggggggccg ttaagagtgg acagagcgcac	600
agtgaccaag gagggctagg ggctgaagca tgtccgaaa acaagaagag ggttcacgc	660
tgtcagtttta acgggtgccg gaaagtttat acaaaaagct cccacttaaa ggcccacccag	720
aggactcaca caggtgagaa gccttataag tgctcatggg agggatgtga gtggcgttt	780
gcacgaagcg atgagctac gaggcaactac agggaaacaca caggtgcataa gcccctcaaa	840
tgcaaccact gcgacaggtg ttttccagg tctgaccatc ttgcctcca catgaagaga	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: Synthetic
polynucleotide

```

<400> SEQUENCE: 19	
atggtcata tggataact cataaacaac ttggaggctc aacttaatc agaagggtggc	60
tcaatgcagg tattcaagca ggtcaactgct tctgttcggc acagagatcc ccctgagata	120
gaatacagaa gtaatatgac ttctccaaca ctcctggatg ccaacccat ggagaaccca	180
gcactgttta atgacatcaa gattgagccc ccagaagaac tttggctag tgattcagc	240
ctgccccaaag tggaaccagt tgacccctcc tttcacaagc ccaaggctcc tctccagct	300
gctagcatgc tacaagctcc aatacgtccc cccaaaggccac agtcttctcc ccagaccctt	360
gtgggtcaca cgtcaacatc tgacatgagc acttcagcaa acatccctac tggctgacc	420
ccaggctctg tcctgaccatc ctctcagac actggtagcc agcagatctt acatgtcatt	480
cacactatcc cctcagtcag tctgccaat aagatgggtg gcctgaagac catcccaat	540

---

-continued

---

gtagtgcagt ctctgccccat ggtgtatact actttgcctg cagatggggg ccctgcagcc	600
attacagtcc cactcattgg aggagatggt aaaaatgctg gatcagtgaa agttgacccc	660
acctccatgt ctccactgga aattccaagt gacagtgagg agagtaacaat tgagagtgg	720
tcctcagcct tgcagagtct gcagggacta cagcaagaac cagcagcaat ggcccaaatg	780
cagggagaag agtcgcttga cttgaagaga agacggattc accaatgtga ctttgaggg	840
tgcagcaaaag tgtacaccaa aagtcctcac ctgaaagctc accgcagaat ccatacagga	900
gagaagcctt ataaatgcac ctgggatggc tgctcctgga aatttgcctg ctcagatgag	960
ctcactcgcc atttccgcaa gcacacaggg atcaagcctt ttccgtgcac agactgcaac	1020
cgcagctttt otcgttctga ccacctgtcc ctgcategcc gtcgcctatga caccatg	1077

<210> SEQ ID NO 20  
<211> LENGTH: 732  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 20	
atgtccgcgg ccgcctacat ggacttcgtg gctgcccagt gtctggtttc catttcgaaac	60
cgcgcgtgcgg tgccggagca tggggtcgct ccggacgccc agcggctgctg actacctgag	120
cgcgagggtga ccaaggagca cggtgaccgg ggggacacct ggaaggattt ctgcacactg	180
gtcaccatcg ccaagagctt gttggacctg aacaagtacc gacccatcca gacccctcc	240
gtgtgcagcg acagtctggaa aagtccagat gaggatggat gatccgacag cgacgtgacc	300
accgaatctg ggtcgagtcc ttcccacagc ccggaggaga gacaggatcc tggcagcgcg	360
cccagccgcg tctccctccct ccatcctgga gtggctgcga agggaaaca cgcctccgaa	420
aagaggcaca agtgcctcta cagtggctgt gggaaagtct atggaaaatc ctcccatctc	480
aaagccatt acagagtgca tacaggtgaa cggcccttc cctgcacgtg gccagactgc	540
ctaaaaaaagt tctcccgctc agacgagctg acccgccact accggaccca cactggggaa	600
aaggcagtcc gctgtccgct gtgtgagaag cgcttcatgaa ggagtgacca cctcacaaag	660
cacgcggcggc ggcacaccga gttccacccc agcatgatca agcgatcgaa aaaggcgctg	720
gcacacgcgtt tg	732

<210> SEQ ID NO 21  
<211> LENGTH: 1440  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 21	
atgctcaact tcgggtccctc tctccagcag actgcggagg aaagaatgga aatgatttct	60
gaaaggccaa aagagagttat gtattcctgg aacaaaactg cagagaaaag tgattttgaa	120
gctgtagaag cacttatgtc aatgagctgc agttggaaatg ctgatattaa gaaatacgat	180
gaaaacagac ctgttacacc agtatctgat ttgtcagagg aagagaatct gcttccggaa	240
acacctgatt ttcatacaat cccagcattt tgtttgactc caccttacag tccttctgac	300

-continued

---

tttgaacct	ctcaagtgtc	aaatctgatg	gcaccagcgc	catctactgt	acacttcaag	360
tcactctcag	atactgccaa	acctcacatt	gcccaccc	tcaaagagga	agaaaagagc	420
ccagtatctg	ccccaaact	ccccaaagct	caggcaacaa	gtgtgattcg	tcatacagct	480
gatgccagc	tatgtAACCA	ccagacctgc	ccaatgaaag	cagccagcat	cctcaactat	540
cagaacaatt	cttttagaag	aagaacccac	ctaaatgttg	aggctgcaag	aaagaacata	600
ccatgtgccc	ctgtgtcacc	aaacagatcc	aatgtgaga	gaaacacagt	ggcagatgtt	660
gatgagaaag	caagtgtgc	actttatgac	ttttctgtgc	cttcctcaga	gacggtcac	720
tgcaggctc	agccagcccc	tgtgtccca	caacagaagt	cagtgttgtt	ctctccacct	780
geagtatctg	cggggggagt	gccacctatg	ceggcatct	gccagatgtt	tcccctttct	840
gecaacaacc	ctgttgtgac	aacagtcgtt	cccagcactc	ctcccaagcc	gccaccagcc	900
gtttggcccc	ctgttgtgtt	catggcaca	caagttccca	aaggcgctgt	catgtttgt	960
gtaccccaagc	ccgttgtca	gagttcaaag	cctccgggtgg	tgagcccgaa	tggcaccaga	1020
ctctctccca	ttgccccctgc	tcctgggttt	tcccttcag	cagcaaaagt	cactctcag	1080
attgattcat	caaggataag	gagtcacatc	tgtagccacc	caggatgtgg	caagacatac	1140
tttaaaagg	cccatctgaa	ggcccacacg	aggacgcaca	caggagaaaa	gcctttcagc	1200
tgtagctgga	aagggtgtga	aaggagggtt	gcccgttctg	atgaactgtc	cagacacagg	1260
cgaacccaca	cgggtgagaa	gaaatttgcg	tgcggcatgt	gtgacggcg	gttcatgagg	1320
agtgaccatt	tgaccaagca	tgccggcgc	catctatcag	ccaagaagct	accaaactgg	1380
cagatggaaag	tgagcaagct	aaatgacatt	gctctacctc	caacccctgc	tcccacacag	1440

<210> SEQ ID NO 22  
<211> LENGTH: 1536  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 22

atgcataactc	ctgatttcgc	tggacctgac	gacgccccag	ccgtggacat	tatggacatt	60
tgtgaatcta	tactcgaaag	aaagagacat	gattcagagc	gaagtacatg	ctctatcctc	120
gagcaaaacag	acatggaggc	ggtagaaagct	ctgggtgtca	tgtccagttt	gggtcagaga	180
tcccagaagg	gggacttgct	tagaatccga	ccgcttactc	cagttccga	tagcggcgac	240
gtaacaacta	ctgttcatat	ggacgcagcc	acgcctgagc	tgcggaaa	ctttcacagc	300
ctctcaactc	tttgcacatc	tccaccacag	tcccccgtat	ttgtcgaacc	atcaacccgg	360
acccctgtta	gcccccaagt	tacagattca	aaggcgtgt	cgcgcaccga	tgttctgcag	420
agttcagcg	ttgttagcgc	ggcattgac	ggaggggtct	aacgaggctt	gttgggtctt	480
gaacccgtac	cgagttctcc	ttgttagagcc	aagggtacta	gtgttattcg	gcataccggc	540
gagagtccgg	cagcttgttt	ccccaccata	caaaccccaag	actgtcgct	tagtattcc	600
cgggaagggg	aggaacagct	gttggccac	ttcgagacac	ttcaagatac	acacttgaca	660
gatagtttgc	tgtccaccaa	cctgggtgtca	tgtcaacatt	gtttgcacaa	gtccgggggt	720
ctccttctga	ctgacaaagg	tcaacaagcg	ggatggctg	gctgtgtcca	aacatgcagt	780
cctaaaaact	acgaaaatga	tttgccctagg	aaaaccacgc	cgcttatcag	tgtgagttt	840

---

-continued

---

cccgctccac	ctgttctgtg	ccagatgatc	cctgttaaccg	ggcaatcatc	tatgttgct	900
gcgttcttga	agccccccccc	acaactgtcc	gttggtaactg	ttcgcggat	ccttgcgaa	960
gcagcgcccg	ccccgcaacc	cgtgttctgt	ggggcccgctg	tcccgaggg	tgcagtcatg	1020
ttgggttcttcc	cccagggggc	cctcccgcca	ccagctccgt	gtgcagcgaa	tgtcatggct	1080
gccggaaaca	cgaaaattgtt	gccccttgtca	cccgctccag	ttttcataaac	gagctcacag	1140
aatttgtgtc	cacaagtgtcga	cttcttcacga	agacggaact	atgtgtgtc	tttcccagggt	1200
tgcagaaaaaa	catatttcaa	atccctctcat	ctgaaagcac	atcttcggac	ccatacagga	1260
gagaaggcctt	ttaattgttag	ctgggatggc	tgtgataaaaa	aattcgcaag	aagtgtatgag	1320
ctcagtcgac	atcgccaggac	gcataccggg	aaaaaaaaat	tcgtttgtcc	agtttgcac	1380
agaagattta	tgagggtccga	ccatctcacc	aagcacgcgc	gacgcccacat	gactacaaag	1440
aaaatttcctgt	gtctggcaagc	cgaggtggga	aaactcaacc	gaatcgcttc	cgctgaatcc	1500
cccgccagcc	cgctggtaag	tatgcctgtcc	agtgtcc			1536

<210> SEQ\_ID NO 23  
<211> LENGTH: 1206  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 23

atgaacattc	acatgttgcg	caagacgata	aagaacatca	atacattcga	gaaccgaatg	60
ttgtatgttgg	atggcatgcc	cgctgtacgg	gtaaaaaccg	agctcttgg	gtctgaacaa	120
ggatccccaa	acgtccacaa	ctacccggat	atggaggcag	tgcgcgttt	gctcaacaat	180
gtgaaggggag	agccgcctga	ggactcttc	tccgttagatc	atttccagac	acagactgag	240
cccgtagatc	tttcaattaa	caaagccaga	acatctccta	ctgcggtaag	ttcttctccc	300
gttaagtatga	cagcaagtgc	atcttagtca	agttctacga	gcactagcag	ttcttcatct	360
atgtacttttgc	ctagttcacc	aacgggtgtc	acaagtgttt	ctagcgccag	cagcagctca	420
acggtaactgt	ctcccggtcc	actcggtggc	agcgctagtg	gcgtgggtgg	ccaacaattt	480
ctccatatta	ttcaccccggt	gcctccgtct	agtccgtatc	atctccagag	caacaagctt	540
atgtcacgtac	ataggatccc	cgtcgctgtc	cagtcagttc	ccgtcgctca	cacagctgt	600
cgatccccctg	ggaatgtcaa	taatactata	gttggccctt	tgcttgagga	tggttagggc	660
catggaaag	cacagatgga	cccccgccgc	ttgtcaccga	gacagtctaa	atccgatagt	720
gacgacgatg	atttgcctaa	cgtaaacactg	gactctgtg	acgagaccgg	gagtaccgct	780
ctgtcaatcg	ctagggccgt	acaggaggc	cacccaagcc	ctgtgtcacg	agtccgaggt	840
aacaggatga	ataatcagaa	atttccctgt	agcatcagcc	cattttctat	agagtccact	900
cgggagacagc	gacgaagtga	atcaccgcac	tccagaaaaaa	ggaggataca	tcgctgtgac	960
tttgagggt	gtaacaagg	ctacacaaaa	agttcacacc	tcaaggcgc	tcgacggacg	1020
catactgggg	aaaaaccgt	caaatgcacc	tgggagggt	gcacgtggaa	atttgcacgc	1080
tctgacgagt	tgacacgcca	ctatcgaaag	catacggcg	taaagccgtt	taaatgcgc	1140
gattgcgaca	ggagtttag	ccgctctgtat	cacccgtcc	ttcaccggag	gacacatg	1200

---

-continued

---

cttggtt	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

atggctgcgg ctgcataatgt ggatcatttt gcggctgagt gcctgggtgtc aatgtctagt	60
agagcggtgg tacacgggcc cagagaaggc ccagaatcac gcccagaggg cgccgcccgtc	120
gtgtcaaacac cgacgctgcc tcgggtcgag gagcgcccg acgggaagga cagtgcgtca	180
cttttcgttag tagcgagaat attggcagat ctgaatcaac aggctccagc acctgcgccc	240
gtgtgaacgccc gggaggggcgc cgctgcccaga aaggccagaa caccatgcgg cttggcccca	300
cctgcgcccag aaccacaacg tccaggtgcc gaaggtgcgg cggctgcccc tccttcacccg	360
gcctggtctg aaccagaacc agaggcaggt ctgtggaccc agcgcgaacc cggccctgca	420
ggctctgggg aacctggccct gaggcagcgg gtgaggcgcg gccggagcag ggccgacttg	480
gaatcaccgc aaaggaaaca taaatgccat tatgctgggtt gcgaaaaggt ttatggaaag	540
tcatccacc tgaaagcaca cctccgcact cacacgggtg agcgcaccc ttgcgtgtcc	600
tggcaagact gcaataaaaaa gtttgctaga tctgtatgaa ttgcacggca ttatcgaact	660
cataccggtg aaaagaagtt ctcatgcct atatgtgaga aacggttcat gcgctctgac	720
cacttgacga aacatgcaag acgacatgtc aattttcatc cggggatgtt gcagagacgg	780
ggagggggaa gttaggactgg aagtctctcc gactattccc gatccgacgc ttccctacca	840
acgattagcc ccgcaagcag tccc	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 cagtcgcattt ctttgattac ttgcggcccg agtgtttgtt ttccatgtca	60
gcgggggctg tcgttcacag aagaccacca gacccggagg gagcgggagg ggcagctgga	120
tctgaagtgc gcgcggctcc acctgaatca ggcgttcccg gcccgggtcc tccagggtccc	180
gctagegtgc cccaaactccc acaagtgcct gctccgagtc ctggagcggg cggagcagcc	240
ccgcataatcc ttgcagcatc agtgtggcc gatcttcggc gaagctccgg ggagggctcc	300
tggggaaaca gcccggagggc cccgcggact tcaagcggtt tttccgatcc aatcccttgc	360
agtgttcaaa ccccatgtc cggactcgcc cccgcgtcccg gagctgcggc agtgtgcgc	420
cctgaaagct catccgatgc gcccggccgtt ccatctgcgc cagctgcgtcc cgggtgcaccc	480
gcagcatctg gcggctttag ttggggggatc cttggggccg gtcccgcccc tgccggcgat	540
caagctccctc gcaggcgcag tgtaacgcggc gcagcaaaac ggcataatgc cccctttcc	600
ggttgtacaa aagcatacta taagtcatcc catctcaaga gtcaccagag gacgcataca	660

---

-continued

---

ggtgagagac	cttttagctg	tgactggctc	gattgcgaca	agaaaattac	gcggagcgac	720
gaaccttgcgc	ggcaactaccg	cactcacact	ggagaaaaga	ggttctcttg	tccctgtgt	780
cccaaggagt	tctcacgcag	tgatcaacttg	acaaaacatg	ctaggagaca	tccaaacatac	840
catcccaca	tgatagagta	tcgaggtagg	cgacgcacac	ctagaattga	tcctccgctg	900
actagtgaag	tcgagtcaag	tgccagtgga	agcggaccgg	gtcccgcc	ctcatttaca	960
acctgttctt						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: Synthetic
      polynucleotide

```

<400> SEQUENCE: 26						
atgggtggacc	acttacttcc	agtggacgag	aacttctcg	cgccaaaatg	cccagttgg	60
tatctgggtg	ataggctgg	tggccggcgg	gcatatcaca	tgctgccctc	acccgtctct	120
gaagatgaca	gcgatgcctc	cagccccctgc	tcctgttcca	gtcccga	ctc tcaagccctc	180
tgctcctgct	atggtgagg	cctggggcacc	gagagccagg	acagcatctt	ggacttccta	240
ttgtccca	aggccctggg	cagtggcggg	ggcagcggca	gtagcattgg	ggccagcagt	300
ggccccgtgg	cctggggg	ctggcgaagg	gcagcggccc	ctgtgaaggg	ggagcatttc	360
tgcttgc	agtttcc	tttggatcct	gatgacgtcc	cacggccctt	ccagcctacc	420
ctggaggaga	ttgaagagtt	tctggaggag	aacatggagc	ctggagtc	aa ggaggtccct	480
gaggggcaaca	gcaaggactt	ggatgcctc	agccagctct	cagctggg	cc acacaagagc	540
cacccatc	ctgggtcc	cgggagagag	cgctgttccc	ctccaccagg	tttgtccagt	600
gcaggagg	tgcc	cccagg	gggtggggc	cccacgc	ct atggcccat	660
ctgcagatcc	agcc	ccgt	tgcc	gtc	aatcggc	720
caagcccc	ca	aaatgt	gttgc	ctc	cttgc	780
gcactcgtgc	ccc	agg	tggt	acc	cttgc	840
attgc	ccc	att	gc	act	gttgc	900
gcgg	cc	tcc	at	gg	atgc	960
cacaaatgt	at	tttcc	ttt	gg	tttgc	1020
cacccatc	cc	ctg	tc	gg	cc	1080
aggttctc	gc	tct	gtc	cg	cc	1140
taccagtgt	tc	gt	gt	gc	cc	1200
aagg	tg	ca	cc	gg	cc	1248

```

<210> SEQ ID NO 27
<211> LENGTH: 756
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

```

<400> SEQUENCE: 27

-continued

---

atgtcagccg cggtcgcgtg cgtggattat tttgcagcac	60
tccgggtgcag tagttcatcg cggaagacca ggtcctgagg gtgcggggcc tgccggccgg	120
ttggatgttc ggcgcgcgcg cagggaaagcc gtttctcccg gaacacactgg ccctctctct	180
cctccgcggg cggcatcagg cccgggtctt ggtgcagctg cggctctca cctgttgca	240
gcctccatac tggctgaccc tggagggggg ccaggcgcgtg cacctgggtgg cgcgagtcca	300
gaaagttcca gtcggcgcc gtcctccccg agtagtgggc gagctccggg cgccggcacct	360
tctgctgcgcg ctaaatcaca ccgcgcgcct ttcccgact gcgcgaaggc gtattataag	420
tcacagtcatt tgaaatcaca cttgaggaca cataccggcg agagacctt tgcgtgcac	480
tggcagggtt gtgataagaa atttgcgaga agcgacgaa tggccgcaca tcaccgcacc	540
cacacagggg aaaaaagatt ctcatgcaca ctctgttcta agcgcttcac gcgaagcgac	600
catcttgcaa agcgcgttag gagacaccct gggttccacc cgcacccctt gcgcacgcac	660
ggcgcccggt ctactagccc gtctgactca ttggccgtgct ctctcgagg gtccctgct	720
ccgagccccc caccgtcccc agtcctgcc 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: Synthetic polynucleotide

<400> SEQUENCE: 28

atgtacggcc gaccgcagcc tgagatggaa caggaggctg gggagctgag ccgggtggcag	60
gcggcgcacc aggctgcaca ggataacgag aactcagcgc ccatcttcaa catgtcttca	120
tcttctggaa gctctggagt gcacacctct tggaaaccaag gcctaccaag cattcagcac	180
tttcctcaca ggcgcagagat gctggggtcc cttttgggtt ctgttgaggc gcccgggcag	240
aatgtgaatg aaggggggcc acatgtcaatg atgcactgc ctgagctgg tatgagctac	300
tgcccccaag cgactctcac tccttccccg atgatttact gtcagagaat gtctccccc	360
cagcaagaga tgacgatttt cagtggggcc caactaatgc ccgttaggaga gcccataatt	420
ccaaaggtag ccaggccctt cgggtggaaat ctaaggatgc ccccaatgg gctgcagtc	480
tccgcttcca ctggaaatccc aataatgtcc cacactggga accctccagt gccttaccct	540
ggcctctcga cagttaccc tcacgaaaca ttgttggcc cgtactgtgcc ttccactgag	600
geccaggcgc tgctccccc catggctcaatg atgttgcacc cgcacatgc ccatgaccc	660
gggatgcccc cagctgagtc ccagtcattt ctggtttttag gatctcaggaa ctctcttgc	720
atgcagccag actctcaaga aggcacattt ctaccagagc agcccgacc tgctccacag	780
acagtagaga agaactccag gcctcaggaa gggactggta gaaggggcctc ctcagaggca	840
aggccttact gctgcacta cgagaactgc ggaaaagctt ataccaaacgc ctcccaccc	900
gtgagccacc agcgcaagca cacaggtgag aggcacattt ctgcactg ggaaagtgt	960
tcatggtctt tcttcgttc tggatgagctt agacgacata tgcgggtaca caccagat	1020
cgaccatata aatgtatca tgcagccgg gagttcatga ggtctgacca tctcaagcaa	1080
caccagaaga ctcatcgcc gggaccctca gacccacagg ccaacaacaa caatggagag	1140
caggacagtc ctccctgctgc tggctct	1167

---

-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 5

<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> SEQ 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
acacttttc octacacgac gcttccga tctagaacta ttccctggct gttacgcg 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
gactggaggt cagacgtgtg ctcttccat cttgtcttcg ttgggagtgta attacg 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
atgggtgagca agggcgagga ggataaacatg gccatcatca aggagttcat gcgcttcaag 60
gtgcacatgg agggctccgt gaacggccac gagttcgaga tcgaggcgaa gggcgaggc 120
cgccccctacg agggcaccca gaccgccaag ctgaagggtga ccaagggtgg cccccctgcc 180
ttcgccctggg acatcctgtc ccctcagttc atgtacggct ccaaggctca cgtgaagcac 240
cccgccgaca tccccgacta cttgaagctg tccttccccg agggcttcaa gtgggagcgc 300
gtgatgaact tcgaggacgg cggcgtggtg accgtgaccc aggactccctc cctgcaggac 360
```

---

-continued

---

ggcgagttca tctacaaggtaaaggt gaagctgcgc ggcaccaact tcccctccga cggccccgta	420
atgcagaaga agaccatggg ctgggaggcc tcctccgagc ggatgtaccc cgaggacggc	480
gccctgaagg gcgagatcaa gcagaggctg aagctgaagg acggcggcca ctacgacgct	540
gaggtaaga ccacctacaa ggccaagaag cccgtgcagc tgccggcgc ctacaacgtc	600
aacatcaagt tggacatcac ctcccacaac gaggactaca ccatacggttga acagtagaa	660
cgcgccgagg gccgcactc caccggcggc atggacgagc tgtacaag	708

<210> SEQ ID NO 34  
<211> LENGTH: 708  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 34

atggagtctt ctgtctaaat ggagtccgga ggcgccggac aacaaccaca accgcaacca	60
caacaaccct tcctgcccgc ggcgcgtatgt ttttcgcga cgcgtgtgc tgctgcagcg	120
ggggccggctg ctgcggccgc gcaatccgc caacagcaac aacaacaaca gcagcagcag	180
caacaagcgc ctcaacttcg acccgctgca gacggcggcgc ctcagggggg agggcacaag	240
agcgctccga agcaggtaa aaggcagagg agcagtagtc cccgactgtat gcgatgtaa	300
aggcgectca attttagcggt ttttggttac tctttggccc agcagcagcc ggctgcgtt	360
gctcgccgaa atgagccggaa aaggaaccgc gttaaacttg tgaatctcggttgcgaca	420
cttcgagagc acgtacaaa tggggcagct aacaagaaaa ttagttaagt tgagacactg	480
cggctctgcag tggagttat tagacttca caacaattgc ttgacgagca cgatgcgtt	540
ttagcccat tcacggccgg ggtgtgtcc ccaacaatat ctccgaacta cagcaatgtat	600
cttaatagca tggccggaaag tcccggttcc tcctactctt ctgtatgggg cagctacgac	660
cctctcagtc ccgaggagca agagttttt gacttcacta actgggttca	708

<210> SEQ ID NO 35  
<211> LENGTH: 573  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 35

atgatggaca acagaggca ctctagtcta cctgacaaac ttcttatctt ccctgattct	60
gccccgttgc cacttaccag gtccttctat ctggagccca tggctactt ccacgtgcac	120
ccagaggccc cgggtgtcatc tccttactctt gaggagctgc cacggctgcc tttccagc	180
gactcttta tcctggaaa ttacagtcaa ccctggccct tctttttccc 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
ggaaaaagttt cctccatgtat agcaaccacc agccaccatg ctgaccctat gttcagaatt	540

---

-continued

---

gtttgcccaa ctttcttgta caaagttgtc ccc	573
--------------------------------------	-----

```

<210> SEQ ID NO 36
<211> LENGTH: 516
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

```

<400> SEQUENCE: 36

atggagacgc gtaaacccggc ggaacggctg gccttgcatt actcgctgctg caccggcccc	60
ctggggcggttc cggggccctt gcccggactc cccggggaggc accccccttag ggtcggccctg	120
cgtctggacg ccgcgtgtctg ggagtggggcg cgcagcggct ggcacgggg atggcgttac	180
ttagcccggtc cgctggacag cgccttcgag cccgccttcc tccgcaagcg caacgagcgc	240
gagcggcagc gggtgccgtg cgtgaacgag ggctatgcgc ggctccgaga ccacccgtccc	300
cgggagctgg cagacaagcg cctcagcaaa gtggagacgc tccgcgttgc catcgactac	360
atcaagcacc tgcaggagct gctggagcgc caggcctggg ggctcgaggg cgccggccggc	420
gcccgtcccc agcgcggggc ggaatgcac acgcacgggg agtccaaggc ctttcgccg	480
ccttcgcacca gcagegagcc cgaggagggg ggcagc	516

```

<210> SEQ ID NO 37
<211> LENGTH: 833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

```

<400> SEQUENCE: 37

atgccgatgg gggcagcaga aagagggtgt gggcccaat catctgcagc accatggct	60
ggttcagaaa aggccggcaaa gagagggcca tcaaaaagct ggtacccaag agctgtgtca	120
tctgtatgtca cgtccccac tgggtgtat ggagctgacc caaaacctgg accttttgg	180
ggtgtgttag cttagggcc tgcgcccaga ggaacaatga ataataattt ctgcaggggcc	240
cttggtaaca gaaggcctt aggacccct tcatgtatgc aattaggtgt aatgccacccg	300
ccaaagacaag cgccttccc gccggctgaa ccccttggaa atgtacctt cctccataac	360
cctggcccaag ctgaaccacc atattatgtat gcatatgtgt gtgtttccc atatgtgcct	420
tccctgggtg cttttgggtat atatgaatac cttttggc cggctttat ccaaaagagg	480
aatgaaagag agagacagag agtgaagtgt gtgaatgaag gatacggccag attgagaggc	540
catttgcctg gtgccttggc agaaaagaga ttatcaaag ttgaaacccct gagggcggca	600
atcagatata taaaataacct ccaagaactc ctttcatcag cacctgtatgg atcgacacca	660
ccggcttcaa gaggtttacc tggaaactgga ccatggccctg caccggcttc tacaccaagg	720
ccagacagac ctggagatgg agaagcaaga gcacccctt cccttgcctt tgaatcttct	780
gaatcatcat gttttcgcc ttccccctttt ttagaaagtg aagaatccctg gca	833

```

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

```

---

-continued

---

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

<400> SEQUENCE: 38

atgggagacg acagaccgtt tggcgtcaat gccccggct gtggacagag atttacaaac	60
gaggaccacc tggcgttca taaacacaag catgagatga cattgaaatt tggcccgagcc	120
cgaactgact cagtcatcat tgcagatcaa acgcctactc caacttagatt cctgaagaac	180
tgtgaggagg tgggactctt caatgaacta gctagctctt ttgaacatga attcaagaaa	240
gtgcagatg aggatgagaa aaaggcaaga agcaggactg ttgccaaaaa actggtgct	300
gtgcgtggc cccttgacat gtctctgcct tccacaccag acataaaaat caaagaagaa	360
gagccagtgg aggttagactc atccccaccc gatagccctg cctctagtcc ctgttcccc	420
ccactgaagg agaaggaggt taccccaaag cctgttctga tctctacccc cacacccacc	480
attgtacgtc ctggctccct gcctctccac ttgggctatg atccacttca tccaaccctt	540
ccctcccaa cctctgtcat cacacaggctt ccaccatcca acaggcaaat ggggtctccc	600
actggctccc tccctttgt catgcattt gctaattggac agaccatgcc tgggtggca	660
gggcctccag tacatgtcc gtcgttata tgcgtggccca gacccatgtc catgggtccc	720
aacatccctg gtatccctgg cccaccagtt aacagtagtg gtcatttc tccctctggc	780
caccctatac catcagaagc caagatgaga ctgaaagccca ccctaactca ccaagtctcc	840
tcaatcaatg gtgggttgtt aatgggttgtt ggtactgcca gcaccatggt gacagccgc	900
ccagagcaga gccagattct catccagcac cctgatgccat cccctctgc ccagccacag	960
gtctcaccag ctcagccac ccctagactt gggggggcgc acggccgcac agtagatgaa	1020
gatccagatg agcgacggca gcgtttctg gagcgcaacc gggctgcagc ctcccgctgc	1080
cgccaaaggc gaaagctgtt ggtgtccctcc ctagagaaga aggccgaaga actcaactct	1140
cagaacattc agctgagtaa tgaagtcaca ttactacgca atgagggtgc ccagttgaaa	1200
cagctactgt tagctataa agactgccc gtcactgcac tacagaaaaa gactcaaggc	1260
tattnagaaa gcccccaagga aagctcgag ccaacgggtt ctccagcccc tggattcag	1320
cacagctcg tagcaatggc ctcagtttgc gtcactgcac tgaagctgtt	1380
gcccaccccg tcctactca gatggccagc caaaggacag aactgagcat gcccataaa	1440
tcgcattaa tcatgacccc acagtcccgat 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> SEQUENCE: 39

atgtacgtca gtcacccctt ggacaaggac gtgagcatgt accctagctc cgtgcgcac	60
tctggggcc tcaacctggc gcccagaac ttcgtcagcc ccccgagta cccggactac	120
ggcggttacc acgtggccgc cgccagctca gccggcagcga acttggacag cgcgcagtc	180
cggggccat cctggccggc agcgatggc gccccactcc gggaggactg gaatggctac	240
gccccccggag gcgccggcgc cgccgccaac gcccgtggctc acggcctcaa cggtggctcc	300

---

-continued

---

ccggccgcag ccatgggcta cagcagcccc gcagactacc atccgcacca ccacccgcat	360
caccacccgc accacccggc cgccgcgcct tcctgcgcctt ctgggtgtct gcaaacgctc	420
aaccccgccc ctccctggcc cgccgcacc gctgcgcgc agcagctgtc tcccgccgc	480
cagcggcggg acctgtgcga gtggatgegg aagccggcgc agcagtccct cggcagccaa	540
gtgaaaacca ggacgaaaga caaatatcga gtggtgtaca cggaccacca gggctggag	600
ctggagaagg agtttacta cagtcgctac atcaccatcc ggagggaaagc cgagctagcc	660
gccacgctgg ggctctctga gaggcagggtt aaaatctggt ttcagaaccg cagagcaaag	720
gagagaaaaa tcaacaagaa gaagttgcag cagcaacagc agcagcagcc accacagccg	780
cctccgcgc caccacagcc tccccagect cagccagggtc ctctgagaag tgtcccagag	840
cccttgagtc cggtgtcttc cctgcaagcc tcagtgtctg gctctgtccc tggggttctg	900
gggccaactg ggggggtgct aaacccccc accgtcacc	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 gggggccac tattctgtca acgccttggc cctaagtggc	60
cccaagtgtgg atctgtatgca ccaggctgtg ccctacccaa gggcccccag gaagcagcgg	120
cgggagcgcacccac ccggagccaa ctggaggagc tggaggact gtttgcac	180
acccagtacc cagacgtcta tgcccgttag gaggtggctc tgaagatcaa tctgccttag	240
tccagggttc aggtttgggtt caagaacccgg agggttaat gcaggcagca ggcacacag	300
cagaaacagc agcagcagcc cccagggggc caggccaaagg cccgctgc caagggaaag	360
gccccacgt ccccaagacc ctccacagat gtgtgtccag accctctggg catctcagat	420
tccatcagtc cccctctgca cgcccccctca ggcccccaaa ccacggcagt ggccactgtg	480
tccatctgca gcccagcctc agagtccct ttgcctgagg cgcaggggc tgggtgtgt	540
gcctcaggcc cgtctctgac ctccggggcc tatgcatgaa cctacgcccc ggccctcgct	600
ttctgtcttt cccctccgc ctatgggtct ccagactctt atttcagcgg cctagacccc	660
tacctttctc ccatggtgcc ccagctaggg ggccggctc tttagccccct ctctggcccc	720
tccgtgggac ctccctggc ccagttttttt acctccat caggccagag ctatggcgcc	780
tacagcccccg tggatagctt ggaattcaag gaccccacgg gcacctggaa attcacctac	840
aatccccatgg accctctgga ctacaaggat cagagtgcct ggaagttca gatcttgc	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	
atggccagca ctattaagga agccttatca gttgtgagtg aggaccagtc gttgttttag	60

---

-continued

---

tgtgcctacg	gaacgccaca	cctggctaag	acagagatga	ccgcgtcctc	ctccagcgac	120
tatggacaga	cttccaagat	gagccccacgc	gtccctcagc	aggattggct	gtctcaaccc	180
ccagccaggg	tcaccatcaa	aatggaatgt	aacccttagcc	aggtgaatgg	ctcaaggAAC	240
tctcctgatg	aatgcagtgt	ggccaaaggc	gggaagatgg	tgggcagccc	agacaccgtt	300
gggatgaact	acggcagcta	catggaggag	aagcacatgc	caccccaaaa	catgaccacg	360
aacgagcgca	gagttatcgt	gccagcagat	cctacgctat	ggagtgacaga	ccatgtgcgg	420
cagtggctgg	agtgggcgggt	gaaaagatat	ggccttccag	acgtcaacat	cttgttattc	480
cagaacatcg	atgggaagga	actgtgcaag	atgaccaagg	acgacttcca	gaggctcacc	540
cccaagctaca	atgcccacat	ccttctctca	catctccact	acctcagaga	gactccttct	600
ccacatttga	cttcagatga	tgttgataaa	gccttacaaa	actctccacg	gttaatgcat	660
gttagaaaca	cagggggtgc	agcttttatt	ttcccaaata	tttcagata	tcctgaagct	720
acgcaaagaa	ttacaactag	gccagattta	ccatatgagc	cccccaggag	atcagcctgg	780
acccggtcacg	gccacccac	gccccagtcg	aaagctgctc	aaccatctcc	ttccacagtg	840
cccaaaaactg	aagaccagcg	tcctcagtt	gatccttatac	agattttgg	accaacaagt	900
agccgccttg	caaatccagg	cagtggccag	atccagcttt	ggcagttcct	cctggagctc	960
ctgtcgac	gtctcaactc	cagctgcata	acctggaaag	gcaccaacgg	ggagttcaag	1020
atgacggatc	ccgacgaggt	ggccggcgc	tggggagagc	ggaagagcaa	acccaacatg	1080
aactacgata	agctcagccg	cgcctccgt	tactactatg	acaagaacat	catgaccaag	1140
gtccatggga	agcgctacgc	ctacaagttc	gacttccacg	ggatcgccca	ggccctccag	1200
ccccacccccc	cggagtcata	tctgtacaag	tacccctcag	acctcccgta	catggctcc	1260
tatcacgccc	acccacagaa	gatgaacttt	gtggcgcccc	accctccacg	cctcccccgt	1320
acatcttcca	gttttttgc	tgcccaaacc	ccatactgg	attcaccaac	tgggggtata	1380
taccccaaca	ctaggctccc	caccagccat	atgccttctc	atctggcac	ttactac	1437

<210> SEQ ID NO 42  
<211> LENGTH: 1326  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 42

atgtcaaaca	aagatgcaca	cattgattcc	agctgttcgt	cttcatcaa	gacggAACCT	60
tccagcccg	cctccctgac	ggacagcgtc	aaccaccaca	gccctgggtgg	ctcttcagac	120
gccagtggg	gtcacagttc	aaccatgaat	ggccatcaga	acggacttga	ctcgccacct	180
ctctaccctt	ctgctcctat	cctggggaggt	agtgggcctg	tcaaggaaact	gtatgtatgc	240
tgctccagca	ccattgtga	agatccccag	accaagtgt	aatacatgt	caactcgatg	300
cccaagagac	tgtgtttagt	gtgtgggtac	atcgcttctg	ggttaccacta	tggggtagca	360
tcatgtgaag	cctgcaaggc	attttcaag	aggacaattc	aaggcaatata	agaatacagc	420
tgccctgcca	caaatgaatg	tgaatcaca	aagcgcagac	gtaaatcctg	ccaggcttgc	480
cgcttcatga	agtgtttaaa	agtgggcac	ctgaaagaag	gggtgegtct	tgacagagta	540

-continued

---

cgtggaggtc ggcagaagta caaagcgcagg atagatcggtt agaacacccc atacctgaac	600
cctcagctgg ttcagccagc caaaaagcca ttgctctggt ctgatccctgc agataacaag	660
attgtctcac atttgttggt ggctgaaccg gagaagatct atgccccatgcc tgaccctact	720
gtccccgaca gtgacatcaa agccctcaact acactgtgtg acttggccga ccgagagttg	780
gtgggttatca ttggatgggc gaagcatatt ccaggcttct ccacgctgtc cctggcggac	840
cagatgagcc ttctgcagag tgcttggatg gaaattttga tccttggtgtt cgataacgg	900
tctcttcgtt ttgaggatga acttgtctat gcagacgatt atataatggc cgaagaccag	960
tccaaatttag caggecttct tgcataat aatgcatacc tgcaagctggt aaagaataac	1020
aagagcatga agctggaaaa agaagaattt gtcaccctca aagctatagc tcttgctaat	1080
tcaagactcca tgcacataga agatgttcaa gccgttcaga agcttcagga tgtcttacat	1140
gaagcgctgc aggattatga agctggccag cacatggaa accctcgctcg agctggcaag	1200
atgctgatga cactgcccact cctgaggcag acctctacca aggccgtgca gcatttctac	1260
aacatcaaac tagaaggcaa agtcccaatg cacaacttt 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: Synthetic
polynucleotide

```

<400> SEQUENCE: 43

atggatcttt ggaactggga tgaagcttcc cctcaagaag ttccccccgg aaataaactc	60
gcggggcttg gaagactccc tcgccttccg caacgcgtct ggggccccatg ccctgggtgg	120
gcctcagccgg acccaaaccctt tttgtctcca gcccggggggg caaagttggg tttctgttcc	180
ccggatcttg ctttgcagg cgataactcca acggcgacgg cagagacccgtt tggaaaggc	240
accagtagct ccctggccag cttccgcag ctcgattggg ggtcagccct tctccatccc	300
gaagttccct gggggccgga acccgactcc caagcccttc cctggagtgg tgattggaca	360
gatatggcat gcacagccctg ggacagttgg tccggggcggt cacagacatt gggaccagcc	420
ccacttggac cggggccatat ccccgccagca ggaagcgaag gagctgtctgg tcagaactgt	480
gtgcccgtgg ctgggtgaggc taccagttgg tccaggccccc aggcagccagg cagtaacacc	540
agctgggatt gctcgtggg gcctgacggg gataacttatt ggggctctgg tcttgggtgg	600
gaaccggagaa cggactgtac gataagttgg ggcgggtccag ctgggcctga ttgtactacg	660
tcatggaaatc ctgggttgca cgcggccggc acgacaagcc ttaagagata tcaaaggatca	720
gcccttacag tttgtctaga accttccccg caaaatgtacc gagcgtcaact ggcggatgt	780
cctaataacta atcatcgagg gcccgtccag ttgtggcagt ttttgcgttga actccttcac	840
gatggcgccga ggagcagttt catcagatgg accggtaaca gcagggagtt ccaattgtgt	900
gaccccaagg aagtggctcg actgtgggggt gagcgcacaaac ggaagcctgg tatgaattac	960
gaaaaggatgtt gtaggggttt gcgatattac tataggcgccg acatcggttgg aaagtccgg	1020
ggtcgaaagt acacatacag attcggccgtt cgcgtaaccat ctcttgatccat ccctgtatgc	1080
gcaggcgcccc gtaggggtgc ggaaacacaaa	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 acgaccaggc cctctttgac	60
tca g cgtacg gagcggcagc ccatctcccc aaggccgaca tgactgcctc ggggagtcct	120
gactacgggc agccccacaa gatcaacccc ctccccaccac agcaggaggc gatcaatccag	180
ccagtgggg tcaacgtcaa gcgggagttt gaccacatga atggatccag ggagtctccg	240
gtggactgca gcgttagcaa atgcagcaag ctgggtggcc gaggcggagtc caacccatg	300
aactacaaca gctataatgga cgagaagaat ggccccccctc ctcccaacat gaccaccaac	360
gagaggagag tcatacgcccc cgcagacccc acactgtggaa cacaggagca tgtgaggcaa	420
tggctggagt gggccataaaa ggagtacagc ttgtatggaga tggacacatc cttttccag	480
aacatggatg gcaaggaaact gtgtaaaatg aacaaggagg acttcctccg cgccaccacc	540
ctctacaaca cggaaagtgtt gttgtcacac ctcagttacc tcaggaaag ttcaactgt	600
gcctataata caacccccc caccgaccaa tcctcacat tgagtgtcaa agaagaccc	660
tcttatgact cagtcagaag aggagcttgg ggcaataaca tgaattctgg cctcaacaaa	720
agtcctcccc ttggggggc acaaacgatc agtaagaata cagagcaacg gccccagcca	780
gatccgtatc agatcctggg cccgaccgc agtcgcctag ccaaccctgg aagcgggcag	840
atccagctgt ggcaattccct cctggagctg ctctccgaca ggcaccaacgc cagctgtatc	900
acctggggagg ggaccaacgg ggagttcaaa atgacggacc ccgatggagt ggccaggcgc	960
tggggggcggc ggaaaagcaa gccaacatg aattacgaca agctgagccg ggcctccgt	1020
tattactatg ataaaaacat tatgacccaa gtgcacggca aaagatatgc ttacaaattt	1080
gacttccacg gcattgcccc ggctctgcag ccacatecgca ccgagtcgtc catgtacaag	1140
tacccttctg acatctcccta catgcctcc taccatgcc accagcagaa ggtgaacttt	1200
gtccctcccc atccatccctc catgcctgtc acttcctccca gttctttgg agccgcata	1260
caataactgga cttcccccac ggggggaatc tacccttacc ccaacgtccc ccgcattct	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
gatacccaag aagcgtattt ttcagttccc gtaagcaata tgaactccgg attggggagc	120
atgaatagta tgaacacgtt aatgacaatg aatacgatga ccaccagcgg caacatgaca	180
cggccctctt ttaatatgtc atatgcgaa cctggcttttgc ggcgtggcct ctcaccagg	240

---

-continued

---

gcggtcgctg	gaatgcccg	ggggagcgcc	ggagcgatga	actccatgac	cgctgcggc	300
gtgacggcca	tgggtacggc	cctgtcaccc	agtggatgg	gagctatggg	ggcccagcaa	360
gccgcctcaa	tgaatggatt	ggggccctat	gcccggcga	tgaatccctg	catgtccct	420
atggcttatg	cccccagcaa	tttgggtcgc	agttagagccg	cggttggtgg	cgatgccaaa	480
accttcaagc	gaagttatcc	tcatgcgaag	cctccttatt	catatatatac	cttgattacg	540
atggcgatac	agcaggcccc	gtctaagatg	ctgactctga	gtgagatata	ccagtggatc	600
atggaccttt	ttccttacta	ccggcaaaac	caacagagat	ggcaaaactc	aatacggccat	660
agccttcct	tcaatgattt	cggctcgaaa	gtcgctcgaa	gcccgtacaa	gcccggtaaa	720
gggtcctatt	ggacccttca	tccagatagc	ggcaatatgt	tcgagaatgg	ttgttatctt	780
agacggcaga	aacgattcaa	atgtgagaaa	cagccaggtg	ccggcggtgg	tggcgccagc	840
gtttcaggcg	gaagtgggtgc	caagggtggg	cctgagtcta	aaaaagaccc	cagcgagca	900
agcaatccaa	gcgcggactc	tccctgcac	cgcgggtttc	atggtaagac	aggtaagctt	960
gagggggcgc	ctgctccagg	cccggtcgcg	tcaccgc当地	cactggacca	tagtggagct	1020
acagcgaccg	gaggtgtttc	agaactcaag	acgcctgcgt	cctccactgc	gcctccgatc	1080
tccagtggtc	ccgggtgcact	tgcctctgtt	cctgcatactc	atccagcaca	cggactcgcg	1140
ccgcacgagt	cccagtc当地	tttggaaaggg	gaccacact	acagctttaa	ccacccattc	1200
tctattaaca	atttgatgtc	atcctcagaa	cagcagcata	aactcgactt	caaaggctat	1260
gaacaggccc	tgcagtattc	tccatatggc	tctacacttc	ctgcttctct	tccattgggg	1320
tctgcaagtg	tgacaacgcg	ctccccaatc	gagccaagtg	ccctcgagcc	tgcttattat	1380
caaggagtat	attcccgacc	agtttgaat	acaagt			1416

<210> SEQ ID NO 46  
<211> LENGTH: 1374  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 46						
atgctggag	cggtgaagat	ggaagggcac	gagccgtccg	actggagcag	ctactatgca	60
gagcccgagg	gtactcctc	cgtgagcaac	atgaacgccc	gcctgggat	gaacggcatg	120
aacacgtaca	tgagcatgtc	ggcgcccgcc	atgggcagcg	gctcgccaa	catgagcg	180
ggctccatga	acatgtcg	gtacgtggc	gctggcatga	gcccgtccct	ggcgggatg	240
tcccccggcg	cgggccat	ggcgccatg	ggcggtcg	ccggggcgcc	tggcggtgg	300
ggcatggggc	cgactttag	tcccagcctg	agcccgctcg	gggggcaggc	ggccggggcc	360
atggggggcc	tggcccccta	cgccaacatg	aactccatg	gccccatgta	cgggcaggcg	420
ggcctgagcc	gcgcggcgca	ccccaaagacc	tacaggcgca	gotacacgca	cgaaaggccg	480
ccctactcg	acatctcg	catcaccatg	gccatccagc	agagccccaa	caagatgt	540
acgctgagcg	agatctacca	gtggatcatg	gaccttcc	ccttctaccc	gcagaaccag	600
cagcgctggc	agaactccat	ccgcccactcg	ctctccttca	acgactgtt	cctgaaggt	660
ccccgtcg	ccgacaagcc	cggaaggcc	tccttctgga	ccctgcaccc	tgactcgcc	720

---

-continued

---

aacatgttcg	agaacggctg	ctacctgcgc	cggcagaagc	gcttcaagtg	cgagaaggcag	780
ctggcgctga	aggaggccgc	aggcgccgcc	ggcagcggca	agaaggcggc	cgccggggcc	840
caggcctcac	aggctcaact	cggggaggcc	gccggggccgg	cctccgagac	tccggcgggc	900
accgagtcgc	ctcactcgag	cgcctccccg	tgccaggagc	acaagcggagg	gggcctggga	960
gagctgaagg	ggacgcggc	tgccggcgctg	agccccccag	agccggcgcc	ctctcccggg	1020
cagcagcage	aggccgcggc	ccacctgtgt	ggcccgcccc	accacccggg	cctgcggcct	1080
gaggcccacc	tgaagccgga	acaccactac	gccttcaacc	accctgttctc	catcaacaac	1140
ctcatgtctt	cgagcagca	gcaccaccac	agccaccacc	accaccagcc	ccacaaaatg	1200
gacctaagg	octacgaaca	ggtgtatgcac	taccccggt	acggttcccc	catgectggc	1260
agcttggcca	tgggccccgt	cacgaacaaa	acgggcctgg	acgcctcgcc	cctggccgca	1320
gatacctctt	actaccaggg	ggtgtactcc	cgccccattt	tgaactccctc	tttg	1374

<210> SEQ ID NO 47  
<211> LENGTH: 1050  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 47

atgctgggt	cagtgaagat	ggaggcccat	gacctggccg	agtggagcta	ctacccggag	60
gcggggcgagg	tctactcgcc	ggtgacccca	gtgcccacca	tggcccccct	caactcctac	120
atgaccctga	atcctctaag	ctctccctat	ccccctgggg	ggctccctgc	ctccccactg	180
ccctcaggac	ccctggcacc	cccagcacct	gcagcccccc	tggggcccac	tttcccaggc	240
ctgggtgtca	gcgggtggcag	cagcagctcc	gggtacgggg	ccccgggtcc	tgggtctggt	300
cacgggaagg	agatgccgaa	ggggtatcgg	cggccctgg	cacacgccaa	gccaccgtat	360
tcctatatct	cactcatcac	catggccatc	cagcaggcgc	cgggcaagat	gctgaccttg	420
agtgaaatct	accagtggat	catggacctc	ttcccttaact	accgggagaa	tcagcagcgc	480
tggcagaact	ccattcggca	ctcgctgtct	ttcaacgact	gcttcgtcaa	ggtggcgctg	540
tccccagaca	agcctggcaa	gggctctac	tggccctac	accccagctc	agggAACatg	600
tttgagaatg	gctgtcacct	gcgcgcggcag	aaacgcttca	agctggagga	gaagggtaaaa	660
aaagggggca	gcggggctgc	caccaccacc	aggaacggga	cagggtctgc	tgcctcgacc	720
accacccccc	cggccacagt	caccccccgg	ccccagcccc	cgccctccagc	ccctgagcct	780
gaggcccagg	gcgggaaaga	tgtgggggt	ctggactgtg	gctcacccgc	ttccctccaca	840
ccctatttca	ctggcctgga	gtctcccgagg	gagctgaagc	tggacgcggcc	ctacaacttc	900
aaccaccctt	tctccatcaa	caacctaatg	tcagaacaga	caccagcacc	tcccaaactg	960
gacgtgggtt	ttggggctaa	cggggctgaa	ggtggggagc	ctggagtcata	ctaccaggc	1020
ctctattccc	gctctttgct	taatgcatcc				1050

<210> SEQ ID NO 48  
<211> LENGTH: 342  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

---

-continued

---

## polynucleotide

&lt;400&gt; SEQUENCE: 48

atgatgc	aag aatctggac tgagacaaaa agtaacggtt cagccatcca gaatgggtcg	60
ggcggcagca accacttact agagtgcggc ggtttcggg agggggggc caacggagag	120	
acgcggccgg tggacatcg ggacgtac ctcgcccacg cccagcagca gcagcaacag	180	
tggcatctca taaaccatca gccctctagg agtcccagca gttggcttaa gagactaatt	240	
tcaaggcctt gggagttga agtctgcag gtccccctgt ggggagcagt tgctgagacg	300	
aagatgagtg gacctgtgt tcagcctaac cttccccat tt	342	

&lt;210&gt; SEQ ID NO 49

&lt;211&gt; LENGTH: 1005

&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 polynucleotide

&lt;400&gt; SEQUENCE: 49

atggagttcc ctggcctggg gtccctgggg acctcagagc ccctccccca gtttgtggat	60
cctgctctgg tgcctccac accagaatca ggggtttct tccctctgg gcctgaggc	120
ttggatgcag cagttccctc cactgccccg agcacagcca cgcgtgcagc tgccggactg	180
gcctactaca gggacgctga ggcctacaga cactccccc tcttcaggt gtaccattg	240
ctcaactgta tggagggat cccagggggc tcaccatatg cccgctggc ctacggcaag	300
acggggctct accctgcctc aactgtgtgt cccaccccg aggactctcc tccccaggcc	360
gtggaagatc tggatggaaa aggacgacc agttctgg agactttgaa gacagagcgg	420
ctgagccctg acctcctgac cctggacact gcactgcctt catcaactccc tgtccccat	480
agtgcattatg gggccctga ctttccagt accttctttt ctccccccgg gagccccctc	540
aattcagcag cctattccctc tcccaagtt cgtggaaactc tccctgtgcc tccctgtgag	600
gcacggaggt gtgtgaactg cggagcaaca gccactccac tggccggag ggacaggaca	660
ggccactacc tatgcacacgc ctgcggccctc tatcacaaga tgaatggca gaacaggccc	720
ctcatccggc ccaagaagcg cctgattgtc agtaaacggg caggtactca gtgcaccaac	780
tgccagacga ccaccacgac actgtggcgg agaaatgcca gtggggatcc cgtgtcaat	840
gectgeggcc tctactacaa gctacaccac cagcaactact gtggggctc cgctcagctc	900
atgagggcac agacatggc ctccagagga ggggtgggtt cttctccctc tttagccag	960
aattctggac aacccaagtc tctggccccc aggcacccccc tggct	1005

&lt;210&gt; SEQ ID NO 50

&lt;211&gt; LENGTH: 1440

&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 polynucleotide

&lt;400&gt; SEQUENCE: 50

atggaggtgg cgccggagca gcccgcgtgg atggcgcacc cggccgtgct gaatgcgcag	60
caccccgact cacaccaccc gggcctggcg cacaactaca tggaaacccgc gcagctgctg	120

-continued

---

cctccagacg	agggtggacgt	cttcttcaat	cacctcgact	cgcaggggcaa	cccctactat	180
gccaaccccg	ctcacgcgcg	ggcgcgcgtc	tcctacagcc	ccgcgcacgc	ccgcctgacc	240
ggaggccaga	tgtgecgccc	acacttgttgc	cacagccccg	gtttgeccctg	gctggacggg	300
ggcaaaggcag	ccctctctgc	cgctgcccgc	caccaccaca	accctggac	cgtgagcccc	360
ttctccaaga	cgcactgca	ccctctcgat	gctggaggcc	ctggaggccc	actctctgtg	420
tacccagggg	ctgggggtgg	gagcgggggg	ggcagcgaaa	gctcagtggc	ctccctcacc	480
cctacagcaa	cccactctgg	ctcccacctt	ttcggcttcc	caccacgccc	acccaaagaa	540
gtgtctcttg	accctagcac	cacgggggct	gctgtccag	cctcatcttc	cgcgaaaaat	600
agtgcagccc	gaggagagga	caaggacggc	gtcaagtacc	aggtgtcaat	gacggagagc	660
atgaagatgg	aaagtggcag	tccctgtgc	ccaggcctag	ctactatggg	cacccagcct	720
gtacacacacc	acccatcccc	cacctacccc	tcctatgtgc	cgccggatgc	ccacgactac	780
agcageggac	tctccaccc	cgaggagcttc	ctggggggac	cgccctccag	cttcacccct	840
aagcagcgca	gcaaggctcg	ttcctgttca	gaaggccggg	agtgtgtcaa	ctgtggggcc	900
acagccaccc	ctctctggcg	gccccggcggc	accggccact	acctgtgc	tgcctgtggc	960
ctctaccaca	agatgaatgg	gcagaaccga	ccactcatca	agcccaagcg	aagactgtcg	1020
gcccggcagaa	gagccggcac	ctgttgtca	aattgtcaga	cgacaaccac	cacccatgg	1080
cgccgaaacg	ccaaacgggg	ccctgtctgc	aacgcctgtg	gcctctacta	caagctgcac	1140
aatgttaaca	ggccactgac	catgaagaag	gaaggatcc	agacteggaa	ccggaagatg	1200
tccaaacaaat	ccaagaagag	caaaaaagg	gccccggatct	tccggaggatct	gtcaaagtgc	1260
atgcaggaga	agtcatcccc	cttcagtgc	gctgcctgg	ctggacacat	ggcacctgt	1320
ggccacccctc	cgcccttcag	ccactccgg	cacatctgc	ccactccgac	gcccatccac	1380
ccctccatcca	gcctctcctt	cgggccaccc	cacccgtcca	gcatggtgac	cgccatggc	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: Synthetic
polynucleotide

```

<400> SEQUENCE: 51						
atgtaccaga	gcctggctat	ggctgcta	atggacctc	cccctggagc	ctatgaagcc	60
ggaggacctg	g	cgcttttat	gcatggatct	ggcgccgtt	cttctccgt	120
acaccttagag	tgcccagcag	cgtgtgggc	ctttcttata	tgcaggagg	aggagcagga	180
tctgtttctg	gccccggatct	aggcgatct	tctggaggcg	ctgtttcagg	tgctggacct	240
ggaactcaac	agggatctcc	tggatggta	caggcaggag	ctgatggagc	cgcttataacc	300
cctccctctg	tgagcccccag	gtttagcttt	cctggcacaa	caggctctt	agctgccgt	360
gctgctgcag	ccgcagctag	agaagcagct	gcatattcta	gtggccggagg	agctgtgg	420
gccccgttag	ctggaagaga	gcagtaacgg	agagccggat	ttggccggaa	ctatagcagc	480
ccttacccctg	octatatggc	cgatgttggc	gcatcttggg	cagccgccc	agcagcttct	540
gcaggacctt	ttgactcacc	tgtgtttcac	tctctgcctg	gcagagctaa	tcctggcc	600
agacatccca	acctggacat	gttcgacgac	ttcagcgagg	gcagagaatg	cgtgaactgc	660

---

-continued

---

ggagccatga gcacccccc ttggagaaga gacggcaccc gccactacct ttgcaatgcc	720
tgtggectgt accacaagat gaacggcatc aacagacccc tgcataagcc ccagagaaga	780
ctgagcgcta gcagaagagt gggectgtcc tgcccaatt gccagaccac aaccaccaca	840
ctgtggagga gaaatgccga gggegagect gtgtgttaacg cctgtggact gtatcatgaag	900
ctgcacggcg tgcccagacc tctggccatg agaaaggagg gcatccagac cagaaagaga	960
aagcccaaga acctaacaacaa gagcaagacc cccgctgctc cttcttggaaag cgagagctg	1020
cctccagcct ctggagccag cagcaatagc tctaaccgcca ccacatcttc ttctgaggag	1080
atgaggccca taaaaaccga gccaggctg agcagccact acggccacag ctctagcgtg	1140
agccagactt ttagcgtgtc tgccatgtca ggccacggac ctagcattca ccctgtgtg	1200
agcgccctga agttgagccc acagggttat gtttctctg tgtctcagag ccctcagacc	1260
tccagcaaggc aggactcttg gaattctctg gtgtggccg acagccacgg cgatatcatc	1320
accgccc	1326

<210> SEQ\_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 ccgacggcg atgggtgtc cttaaaagat tcggcgccgc tggcgctgtat	60
gtttctgaca gcagagcctt ccccgctagg gaacccagca caccacctag ccccatcagc	120
agctcaagct ctagctgttag cagaggcgga gagagaggac ctggaggcgc ttctaaactgc	180
ggcacacctc agctggatac agaagccgcg gccggaccac cagccagatc tctttactt	240
agcagctacg ccagecaccgc ttggcgct cctcatggac cctctgtcc ttgtgtggcc	300
ggacctggcg gaaaccttagt ctcttggag gaccttctgc tgtttaccga cctggaccag	360
gtgtgccaccg ctagcaagct tctgtggagc agcagggggc ctaagctgag ccctttgcc	420
cctgagcgc ccgaggagat gtaccagacc ctggctgttt taagctctca gggacctgccc	480
gtttatgacg gagccccctgg ttggattttt cactcagcg cagcagccgc agctgctgca	540
gcccgtgcca gtcacactgt gtatgtgcct accacaagag tgggcagcat gttacctggaa	600
cttccttacc atctgcaggc cagcggaaacg ggcctgtcta accatgcggg aggagctgga	660
gttcaccccg gatggcctca ggcttctgc gattctctc cttatggatc ttggaggagga	720
gcagctggag gggggactgc aggaccagggt ggagccggaa ggcgcggc acatgtgtct	780
gccagatttc cctatagcccc tagccctctt atggccaatg ggcgtgttag agaaccggaa	840
ggatatgtcg cggcaggctc ttggcgccgt ggcggagttt ctggagggtgg atcttcactg	900
gcccgtatgg gggggactgc ggcgtgttag tttctctgc ggcgcggc accactgaac	960
ggcacctatac atcaccacca ccatcaccat catcatcacc ccagccctta ctcccttat	1020
gtggggagccccc cccttacacc cgttggctt gcccggccctt tcgagacacc tttgtgtgcac	1080
gccttcagt ctagagctgg cgcacccctta ccagtgccta gaggcccttc tgccgacttg	1140
ctggaggatc tgagcgagag cagagagtgc gtgaactgtg gacacccctg	1200

-continued

---

tggagaagag acggcacccg ccactacctg tgcaacgcgt gggccctgta cagcaagatg	1260
aatgggctga gcagacccc gatcaagccc cagaagaggg tgcccagcag cagacggctg	1320
ggactgagct gcgccaactg tcataccaca acaaaccacac tggggggagaa aaacgcccag	1380
ggcgagcccg tggtaacgc ctgcggcctt tacatgaagc tgacacggctg gcccagacct	1440
ctggccatga agaaggaggg aatccagacc agaaagagaa agcccaagaa catcaacaag	1500
agcaagaccc gcagcggcaa cagcaacaac agcatcccc tggggggac cagcacatct	1560
agcaacacgcg acgactgttag caagaacaca tcacctacca cccagccac agctagcgga	1620
gecggegccc ccgtgtatgac aggcccggaa gagtcacaaa atcccgagaa tagcgaactg	1680
aagtactctg gacaggacgg actgtatatac ggctgtgagcc tggcttctcc cgccgagggt	1740
accagctctg tcagacctga ctcttggtgt gcccctgccc tggcc	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: Synthetic polynucleotide

<400> SEQUENCE: 53

atgttcaact cgatgacccc accaccaatc agtagctatg gcgagccctg ctgtctcgg	60
ccccctccca gtcagggggc ccccagtgtg gggacagaag gactgtctgg cccgccttc	120
tgccaccaag ctaacctcat gtccggcccc cacagttatg ggccagccag agagaccaac	180
agctgcacccg agggcccaact ctttcttct ccccgagtg cagtcaagtt gaccaagaag	240
cgggcactgt ccatctcacc tctgtcgat gccagcctgg acctgcagac gtttatccgc	300
acctcaccacca gtccttcgt agctttcatac aactcgcgtt gcacatctcc aggaggctcc	360
tacggtcatac tctccattgg caccatgagc ccatctctgg gattcccagc ccagatgaat	420
cacaaaaaag ggccctcgcc ttcccttggg gtccagcctt gtggccccca tgactctgcc	480
cggggtggga tgatcccaca tcctcaagtcc cggggacccct tcctaacttg ccagctgaag	540
tctgagctgg acatgtgtgt tggcaagtgc cgggaggaac ctttggaaagg tgatatgtcc	600
agcccaact ccacaggcat acaggatccc ctgttggggta tgctggatgg gcgggaggac	660
ctcgagagag aggagaagcg tgagcctgaa tctgtgtatg aaactgactg ccgttggat	720
ggctgcagcc aggaatttga ctcccaagag cagctgggc accacatcaa cagcgagcac	780
atccacgggg agcggaaagga gtctgtgtgc cactgggggg gctgctccag ggagctgagg	840
cccttcaaag cccagttacat gctggtggtt cacatgcgcac gacacactgg cgagaagcca	900
cacaagtgc acgttgaagg gtgcggaaag tcataactcac gctcgaaaaa cctgaagacg	960
cacctgcggcgt cacacacggg tgagaagcca tacatgtgtg agcacgaggg ctgcagtaaa	1020
gccttcagca atgcccgtga ccgagccaaag caccagaatc ggacccatcc caatgagaag	1080
ccgttatgtat gtaagctccc tggctgcacc aaacgctatac cagatcctag ctcgctgcga	1140
aaacatgtca agacagtgc tggccctgac gcccattgtga ccaaaccggca ccgtggggat	1200
ggcccccgtgc ctggggcacc atccattttt acagtggagc ccaagggaa gggggaaagga	1260
ggtccccatca gggaggaaag cagactgact gtgcagagg gtgcctatgaa gccacagcca	1320
agccctgggg cccagtcatac ctgcagcagt gaccactccc cggcaggagtg cagccaaat	1380

---

-continued

---

acagacagtg	gtgtggaaat	gactggcaat	gcagggggca	gcactgaaga	cctctccagc	1440
ttggacagg	gaccttgcatt	tgctggact	ggtctgtcca	ctcttcgcgg	ccttgagaac	1500
ctcaggctgg	accagactaca	tcaactccgg	ccaataggga	cccggggtct	caaactgccc	1560
agcttgtccc	acacccggta	cactgtgtcc	cgccgcgtgg	gccccccagt	ctctcttgaa	1620
cgccgcagca	gcagctccag	cagcatcagc	tctgcctata	ctgtcagccg	ccgctccccc	1680
ctggcctctc	cttcccccc	tggctccca	ccagagaatg	gagcatcctc	cctgcctggc	1740
cttatgcctg	cccagacta	cctgcttcgg	gcaagatatg	cttcagccag	aggggggtgg	1800
acttcgccta	ctgcagcatc	cagcctggat	cgataggtg	gtcttccat	gcctccttgg	1860
agaagccgag	ccgagtagatcc	aggatacaac	ccaatgcag	gggtcacccg	gagggccagt	1920
gacccagccc	aggctgctga	ccgtcctgct	ccagctagag	tccagaggtt	caagagctg	1980
ggctgtgtcc	atacccccacc	caactgtggca	gggggaggac	agaacttga	tccttaccts	2040
ccaacctctg	tctactcacc	acagcccccc	agcatca	agaatgctgc	catggatgct	2100
agagggctac	aggaagagcc	agaagttggg	acctccatgg	tggcagtgg	tctgaacccc	2160
tatatggact	tcccacctac	tgatactctg	ggatatgggg	gacctgaagg	ggcagcagct	2220
gagcctttag	gagcgagggg	tccaggctct	ctgccttgg	ggcctggtcc	acccaccaac	2280
tatggccca	acccctgtcc	ccagcaggcc	tcatacctc	accccacca	agaaacatgg	2340
ggtgagttcc	tttccactc	tggctgtac	ccaggccca	aggctctagg	tggaaacctac	2400
agccagtg	ctcgacttga	acattatgg	caagtgc	aagccaga	acaggggtgc	2460
ccagtgggg	ctgactccac	aggactggca	ccctgcctca	atgcccaccc	cagtgagggg	2520
cccccacatc	cacagctct	ctttccat	taccccagc	cctctccccc	ccaatatctc	2580
cagtca	ccataccca	gccacccct	gattatctc	ttcagaacc	caggccttgc	2640
ctggactttg	attcccccac	ccattccaca	ggcagctca	aggctcagct	tgtgtgtaat	2700
tatgttcaat	ctcaacagga	gctactgtgg	gggggtgggg	gcagggaa	tgcccccgcc	2760
caggaacctt	cttaccagag	tcccaagttt	ctggggggtt	cccaggtag	cccaagccgt	2820
gtaaaagctc	cagtgaacac	atatggac	ggcttggac	ccaaactg	caatcacaag	2880
tcagg	ttccaccc	ttcaccatgc	catgaaaatt	ttgttagtgg	ggcaaata	2940
gtttcacata	gggcagcagc	accaccc	cttctgcccc	cattgc	ccac	3000
cctctcaa	ggggggccac	aaaccc	tgtggtcatc	ctgagg	ttgg	3060
gggggtct	cttgcaccc	tcctccgaa	ggacaggat	gtaaaa	ccct	3120
gatctt	gaca	acactcag	ggactttgt	gttattctgg	atgagccca	3180
cctc	cctt	ccatgatca	ggggggccac	tctggacata	ccccac	3240
cccaacatgg	ctgtggcaa	catgagtgc	ttactgagat	ccctac	tgg	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

```

---

-continued

---

<400> SEQUENCE: 54

```

atgagtctgg taggtggttt tccccaccac ccgggtggtc accacgaggg ctacccgttt      60
gccgcggccg cgcgcgcag ccgctgcaga catgaggaga acccctactt ccatggctgg      120
ctcatcgccc accccgagat gtcgcggccc gactacagca tggccctgtc ctacagcccc      180
gagtatgcca gcggcaccgc caaccgcaag gagcggcgcg ggactcagag catcaacagc      240
gccttcgccc aactgcgcga gtgcattccc aacgtacccg ccgacaccaa actctccaaa      300
atcaagaccc tgcgcgtggc caccagctac atcgcctacc tcatggaccc gctggccaaag      360
gacgaccaga atggegaggc ggaggccttc aaggcagaga tcaagaagac cgacgtgaaa      420
gaggagaaga ggaagaagga gctgaacgaa atcttgaaaa gcacagttag cagcaacgac      480
aagaaaaacca aaggcoggac gggctggccg cagcacgtct gggccctgga gctcaagcag      540

```

<210> SEQ ID NO 55

<211> LENGTH: 1896

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 55

```

atggtttcta aactgagcca gctgcagacg gagctcctgg cggccctgtct ggagtcaaggg      60
ctgagcaaag aggcactgct ccaggcactg ggtgagccgg ggcctacactt cctggctgg      120
gaaggcccccc tggacaaggg ggagtcctgc ggcggcggtc gaggggagct ggctgagctg      180
cccaatgggc tgggggagac tcggggctcc gaggacgaga cggacgacga tggggaaagac      240
ttcacgcac ccatactcaa agagctggag aacctcagcc ctgaggaggc ggcccacccag      300
aaagccgtgg tggagaccct tctgcaggag gacccgtggc gtgtggcgaa gatgtcaag      360
tcctacactgc agcagcacaa catccacacg cgggaggtgg tgcataccac tggcctcaac      420
cagtccaccat tgcataaca cctcaacaag ggcactccca tgaagacgca gaagcgggccc      480
gcctgtaca cctggtaactt ccgcaagcag cgagagggtgg cgcagcgtt caccatgca      540
gggcaggagg ggctgtattga agagccaca ggtgatgagc taccaaccaa gaagggccgg      600
aggaaccgtt tcaagtgggg cccagcatcc cagcagatcc tggccaggc ctatgagagg      660
cagaagaacc ctagcaagga ggagcggagag acgctagtgg aggagtgc当地 tagggcggaa      720
tgcataccaga gagggggtgtc cccatcacag gcacaggggc tggctccaa cctcgtaacg      780
gaggtgcgtg tctacaactg gtttgcac cggcgc当地 aagaagccctt ccggcacaag      840
ctggccatgg acacgtacag cggggccccc ccaggccag gcccgggacc tgcgtgccc      900
gtcacatggcctt cccatggccctt ggcctctccccc ccacgttgc当地 960
cgctatggac agcctgc当地 cagttagact gcagaagttac cctcaagcag cggcgtccc      1020
tttagtgc当地 tgcataccacc cctccacca gtttgc当地 cggccctggc gcccagccac      1080
agcctgtgtgta gtacagaagc caagctggc当地 tcagcagctg gggggccctt cccctgtc当地 1140
agcaccctga cagcactgca cagcttggag cagacatccc caggccctaa ccagcagccc      1200
cagaacccatcc tcatggccctt acttccctggg gtc当地 gacccatcc tggccctgg tgc当地 1260
tccctggc当地 ctacgttccac caacacaggc gcctccaccc tggccatcc gtc当地 1320
acgcaggc当地 agagtgtgcc ggtcatcaac agcatggc当地 gcagcctgac caccctgca      1380

```

---

-continued

---

cccgteca	tctcccagcc	gctgcacccc	tcctaccaggc	agccgc	tcat	gccac	cgttg	1440
cagagccat	tgacccagag	ccccttcatg	gccaccatgg	ctcag	etgca	gag	cccccac	1500
gccctctaca	gccacaagcc	cgaggtggcc	cagtacacc	acacagg	cttgcag	1560		
actatgtca	tcacccgacac	caccaac	ctgg	ccagc	tcac	gcccac	aa	1620
cagg	tctca	cctcagacac	tgagg	cc	agt	gatcc	gg	1680
cagg	ccacca	ccctccac	gt	cccc	ccag	gatcc	cc	1740
gcccac	ccgc	tcag	ccag	cccc	acag	tggt	gt	1800
tcagactcca	gcaatggca	gagccac	ctg	ccat	ccca	accac	agcgt	1860
ttcatctcca	ccc	agatggc	ctt	cc	cag	ttg		1896

<210> SEQ\_ID NO 56  
<211> LENGTH: 1671  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 56								
atggtagca	aactgacatc	cctccagcag	gaacttctt	ctgc	ccctc	ctcc	agg	60
gt	taacca	aaag	agg	tact	tttgc	ccagg	tttgc	120
gt	aaagt	ttgg	agact	ctccc	cct	ctcc	cttgg	180
gt	attt	cata	cg	tttacaaa	cgg	acac	gc	240
gagg	acgg	cg	atg	attat	ga	cccc	gccc	300
gagg	aaag	cg	gg	tttca	aa	actg	cagg	360
gct	gc	aaaa	tgat	ttat	atgc	ca	acta	420
gat	gtt	acc	gat	tttccc	cc	tc	tc	480
aaa	acac	aga	cg	tttac	aca	tttac	gt	540
cg	ac	atc	atc	atc	atc	atc	atc	600
cag	ctt	ttt	ttt	ttt	ttt	ttt	ttt	660
gat	gtt	gtt	gtt	gtt	gtt	gtt	gtt	720
gc	ct	ct	ct	ct	ct	ct	ct	780
cg	cg	gg	gg	gg	gg	gg	gg	840
ag	ca	aa	gg	tt	tt	tt	tt	900
cca	ac	ac	gg	tt	tt	tt	tt	960
aac	cc	ac	ac	ac	ac	ac	ac	1020
tc	ct	cc	cc	cc	cc	cc	cc	1080
ac	at	tt	tt	tt	tt	tt	tt	1140
tt	tg	ca	ac	cc	cc	cc	cc	1200
aa	at	gt	at	tt	tt	tt	tt	1260
at	gt	tc	tc	tc	tc	tc	tc	1320
gt	at	gg	ct	tt	tt	tt	tt	1380

---

-continued

---

acgcgttagcgg gtcattggc ggcgtccaa ccagtgcagt tctccagca gctccattca	1440
ccccatcaac agcctctgat gcagcagago cctggtagtc acatggctca acagccgtc	1500
atggcagctg tcactcagct ccagaactcc catatgtatg cccacaagca agaaccacca	1560
caatacagtc acacatcaag attccccagt gctatggttg ttactgacac atcctctatc	1620
tcaactctga cgaacatgtc cagtagtaaa caatgtcctc tgcaaggatg g	1671

```

<210> SEQ ID NO 57
<211> LENGTH: 1251
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

```

<400> SEQUENCE: 57

atgcgactct caaaaaccct cgtcgacatg gacatggccg actacagtgc tgcactggac      60
ccagcctaca ccaccttgg aatttggaaat gtgcagggtg tgacgatggg caatgacacg      120
tccccatcg aaggcaccaa cctcaacgcg cccaacagcc tgggtgtcag cgccctgtgt      180
gcccattcg gggaccgggc cacgggcaaa cactacggtg cctcgagctg tgacggctgc      240
aaggggcttct tccggaggag cgtgcggaag aaccacatgt actcctgcag atttagccgg      300
cagtgcgtgg tggacaaaaga caagaggaac cagtgcgcgt actgcaggct caagaaatgc      360
ttccgggctg gcatgaagaa ggaagccgtc cagaatgagc gggaccggat cagcactcga      420
aggtcaagct atgaggacag cagcctgccc tccatcaatg cgctcctgca ggccggaggc      480
ctgtcccgac agatcacctc ccccgctctcc gggatcaacg ggcacattcg ggcgaagaag      540
attgccagca tcgcagatgt gtgtgagtcc atgaaggagc agctgctggt tctcggttag      600
tgggccaagt acatcccagc tttctgcgag ctcccccgtt acgaccagggt ggccctgtctc      660
agagcccatg ctggcgagca cctgctgtc ggagccacca agagatccat ggtgttcaag      720
gacgtgtgc tcctaggcaa tgactacatt gtcctcggtc actgcccggg gctggcgagg      780
atgagccggg tgtccatacg catccttgac gagctggtg tgccttcca ggagctgcag      840
atcgatgaca atgagtatgc ctacctcaaa gccatcatct tctttgaccc agatgccaag      900
gggctgagcg atccaggaa gatcaagcgg ctgcgttccc aggtgcagggt gagcttggag      960
gactacatca acgaccgcca gtatgactcg cgtggccgt ttggagagct gctgtgtcg      1020
ctgcccacct tgcagagcat cacctggcag atgatcgagc agatccaggat catcaagctc      1080
ttcgccatgg ccaagattga caacctgttg caggagatgc tgctggggagg tccgtgcca      1140
geccaggagg ggcgggggtt gagtggggac tccccaggag acaggccctca cacagtgagc      1200
tcaccctca gtccttggc ttccccactg tgcgcgtttt ggcaagttgc t      1251

```

```

<210> SEQ ID NO 58
<211> LENGTH: 1005
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

```

<400> SEQUENCE: 58

atggacaacg cgccggatgaa ttcccttcctc gagtacccaa ttttgtctag tggagacagt      60

```

---

-continued

---

ggcaacttgca	gtgcccggc	ctatccatca	gaccacagaa	ttacaacatt	ccaaagctgt	120
gccccgttcag	ccaacagttt	cggcgaggac	gaccgcttcc	tggtcggaaag	aggggttcaa	180
atggatcac	ctcaccatca	ccatcaccc	caccatacc	acccccaacc	ggcgacttac	240
caaaccagcg	gcaatttggg	cgtgagctat	agccattect	catgtggacc	ttcctatggg	300
tctcagaatt	tctccggccc	ttatagccca	tacgcccgtga	accaagaggc	cgatgtatca	360
ggaggctatac	cccaagtgcgc	gccagcggtt	tactcaggta	atctttctag	cccgatggc	420
cagcaccacc	atcaccatca	aggttatgcc	ggcggtgcag	tccggatcccc	acaatacata	480
caccatagtt	acggccaaga	gcaccaatcc	ctggccctcg	ctacatataa	caactcactg	540
tctccgcgttc	atgcttccca	ccaagaagct	tgtcgaggatc	ccgcctcaga	aacttcctct	600
ccagctcaga	cttttgattt	gatgaaggtc	aagcggaatc	cgccctaaaac	gggcaaagta	660
gggtgaatatg	gttatttggg	acagcctaat	gctgtccgc	ccaaatttac	aacaaaacag	720
cttactgaac	tcgagaagga	atttcatttt	aataagtatt	tgactcgagc	gagacgagtc	780
gaaatcgccg	ctagtcgtca	acttaacgag	acccaggta	agatatggtt	ccagaacacaga	840
agaatgaaac	aaaaaaagcg	ggagaaggaa	ggactcctcc	ctatatacc	agccacaccc	900
ccaggttaacg	acgagaaggc	ggaggaatct	tcaagagaaga	gttccagctc	cccttgcgtt	960
ccttctctcg	gtagtcac	cagegatacc	ctcacgcgca	gtcac		1005

<210> SEQ ID NO 59						
<211> LENGTH: 282						
<212> TYPE: DNA						
<213> ORGANISM: Artificial Sequence						
<220> FEATURE:						
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide						
<400> SEQUENCE: 59						
atgtgtcaag	gcaattccaa	aggtgaaaac	gcagccact	ggctcacggc	aaagagtgg	60
cggaagaagc	gctgcccccta	cacgaagcac	cagacactgg	agctggagaa	ggagttctg	120
ttcaatatgt	accttactcg	agagcggcgc	ctagagatta	gccgcagcgt	ccacccatcg	180
gacagacaag	tgaaaatctg	gtttcagaac	cgcaggatga	aactgaagaa	aatgaatcga	240
gaaaaccgga	tccggagct	cacagccac	tttaattttt	cc		282

<210> SEQ ID NO 60						
<211> LENGTH: 942						
<212> TYPE: DNA						
<213> ORGANISM: Artificial Sequence						
<220> FEATURE:						
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide						
<400> SEQUENCE: 60						
atggattttg	atgagcgtgg	tccctgctcc	tctaacatgt	atttgccaag	ttgtacttac	60
tacgtctcg	gtccagattt	ctccagecctc	ccttcttttc	tgcggccagac	cccgcttctcg	120
cgcctaatga	catactccctaa	ctccctccaa	ctggcccgag	tccaaaccgt	gcgcgaagt	180
accttcagag	agtacgccc	actaaatggc	accccccggg	caatctggcc		240
cactgtact	ccgcggagga	gctcgtgcac	agagactgcc	tgcaggcgcc	cagcgcggcc	300
ggcgtgcctg	gcgacgtgt	ggccaagagc	tccggccaacg	tctaccacca	ccccacccccc	360

-continued

---

gcaatttcta tagcaccgtg ggcaggaacg	420
gctccctgcc acaggcttc	
gaccagttt tcgagacagc ctacggcacc	480
ccggaaaacc tgcctctc cgactacccc	
ggggacaaga gcgccagaa gggggccccc	540
gcccccccg gcccacgg cgacctccgc ggccggggcg	
gcccccccg gggctgcaa cggggcgcc ggcaacttca	600
agttcggaca gggccggcg cggccggctgc	
cgggatgg cggccggcgc agaggagaaa gagccggcg	660
ggcccccga gggccggcgc gggcccccga gggccggcgc	
agccccaggt gtcttcgg ccadacttag gacaaggccg	720
ggggctccag tggccaaacgc	
cccccgcata aacccgcata agcgcgtccc	780
ctataccaaat taccatcccg tggccaaacgc	
taccatcccg tggccaaacgc	840
actgtatgtc aagtcaaaaat ctgggttcag aacaggagaa	900
tgaaggaaaa aaaaattaac	
agagacggtt tacagtacta ctcagcaaat ccactcctct	942
tg	

<210> SEQ ID NO 61  
<211> LENGTH: 672  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 61	
atgagttcct atttcgtgaa ctccaccttc cccgtcactc	60
tggccagcc gcaggagtcc	
ttcctgggcc agctaccgct ctattcgtcg	120
ggctatgcgg acccgctgag acattacccc	
gcgcctacg ggccaggggcc gggccaggac	180
aagggctttg ccacttcctc ctattacccg	
ccggcgcccg gtggctacgg ccgagccgcg	240
ccctgcgact acggggccgc gccggccccc	
taccgcgaga aagagtccgc ctgcgcactc	300
tccggcgccg acgagcagcc cccgttccac	
cccgagccgc ggaagtccga ctgcgcgcag	360
gacaagagcc tgttccgcga gacagaagag	
cagaagtgcg ccactccggc	420
ctaccgtgg atgcagccga tgaattcgtg caacagttcc	
tcctttggcc ccagccgcg gcgaggccgc	480
cagacataca cacgttacca gacgtggag	
ctggagaagg agtttacta caatcgctac	540
ctgacgcggc ggcggccat cgagatcgcg	
cacgcctgt gcctgacgga gaggcagatc	600
aagatatgtt tccagaaccg acgcgttacca	
tggaaaaagg agagcaaact gctcagccgc	660
tctcagctca gtgccgagga ggaggaagaa	
aaacaggccg ag	672

<210> SEQ ID NO 62  
<211> LENGTH: 1410  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 62	
atggctgtca gcaatcgccgt gctcccatct ttctccacgt	60
tgcgtatgg cccgggggaa	
agggagaaga cactcggtca	120
agcagggtgcc cccataacc	
gctggccggaa ggagctctcc	
cacatgaagc gacttcccc	180
agtgtttccc ggcggccccc	
atgacatggc ggccggcacc	
gtggccacag acctggagag	240
cgccggagcc ggtggccgtt	
gcccggtagt caacctggcg	
cccttacctc ggagagagac	300
cgaggagttt aacgatctcc	
tggacacttgc	
ctttatttctc	

-continued

---

tccaaattcgc tgaccatcc tccggagtc a gttggccgcca c cctgttctc gtcagcgta	360
gctctctttt cgtcgctgcc gtcgagcgc ggcctgcca g cgcgcctc cacctgcagc	420
ttcacccata cgatccggc cgggaacgc c cgggcgtgg cgcggggcgg c acggggcgg	480
ggcctctctt atggcaggga gtccgcctcc cctccgacgg ctcccttcaa cctggcgac	540
atcaacacg tgagccccctc gggggcgttc gtggccgagc tcctggggcc agaattggac	600
cgggtgtaca ttccggccca gcagccgcag c cggccaggtg ggggggtat gggcaagtcc	660
gtgctgaagg cgtcgctgag cgcgcctggc agcgactacg cgcgcgcgc ggtcatcagc	720
gtcagcaaag gcageccctga cggcagccac c cgggtgggtgg tggcgcctca caacggcgg	780
cgcgcgcgc a cgtgcgcgc gatcaagcag gaggcggctc cttcgtgcac ccacttggc	840
gtgggacccc ctctcagcaa tggccacccgg c cggctgcac acgacttccc cctggggcgg	900
cagctcccca gcaggactac cccgcacctg ggtcttgagg aagtgtgag cagcaggagc	960
tgtcacccctg ccctgcgcgt tcctccggc ttccatcccc accccggggcc caattaccca	1020
tccttcotgc cgcgcgcgc gatcaagcag gaggcggctc cttcgtgcac ccacttggc	1080
ccacccgggtt octgcgcgc agaggagccc aagccaaaga ggggaagacg atcgtggcc	1140
cggaaaagga cgcgcacca cacttgtat tacgcgggtc gggcaaaac ctacacaaag	1200
agttcccatc tcaaggcaca cctgcgaacc cacacaggta agaaaccta ccactgtgac	1260
tgggaeggct gtggatggaa attcgcgcgc tcagatgaa tgaccaggca ctacggtaaa	1320
cacacggggc accgcgcgtt ccagtgcacaa aaatgcgacc gagcatttc caggcggac	1380
cacctcgcc tacatgaa gaggcattt	1410

<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	
atggaggcgc gcggggagct gggccggcc cgggagtcgg cgggaggcga cctgtgtcta	60
gcactgtgg cgcggagggc ggacactgcgc cgagagatcc cgctgtgcgc tggctgtac	120
cagcacatcc tggaccgcctt cacttcaga gtcctggacc gccactggca cagcaagtgt	180
ctcaagtgc a cgcactgc a cgcgcactg gccgagcgcgt gttcagccg aggaggagc	240
gtttactgc aggacactt ttcaagcgc ttcgggacca a gtgcgcgcgt gtcgcagctg	300
ggcatccgc c acgcagggt ggtgcgcgc gcccaggact tcgtgtacca cctgcactgc	360
tttgcctgc tcgtgtgc a gggcaggctg gccacggggc acgaggctta cctcatggag	420
gacagccgc tcgtgtgc a ggcggactac gaaaccgc a cagcgcaga ggcgcaggcc	480
acggccaaggc ggcgcgcac gaccatcacc gccaagcgc tggagacgc t aagagcgc	540
tacaacaccc cgcgcaccc ggcgcgcac gtcgcgcgc agctctcgac c gacacgggc	600
ctggacatgc cgcgtggc a gtttgggtc cagaaccgc gggccaagg a gaaggagctg	660
aagaaggacg cggcgcggca g cgcgtgggg c agtattcc gcaacatgaa g cgcgcgc	720
ggcggctcca agtcggacaa ggacagcgtt caggaggggc aggacagcga cgcgtgggtc	780
tccttccttc atgagccctc cttggcggaa atggccccc ccaatggcct ctacggagc	840

---

-continued

---

ttgggggaac ccacccaggc cttggggcgg ccctcgggag ccctggcaa cttccctcg	900
gagcatggag gcctggcagg cccagagcag taccgagagc tgcgtccccg cagcccctac	960
ggtgtcccc catccccgc cgccccgcag agectccctg gcccccaagcc cctccctcc	1020
agcctggtgt acccagacac cagttgggc cttgtgecc cgggagcccc cggcggggccc	1080
ccacccatga gggtgctggc agggAACGGA CCCAGTTCTG acctatccac ggggagcagc	1140
gggggttacc ccgacttccc tgccagcccc gcctccctggc tggatgaggt agaccacgct	1200
cagttctcag gcctcatggg cccagcttc ttgtac	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: Synthetic
polynucleotide

```

<400> SEQUENCE: 64	
atggaaggaa tcatgaaccc ctacacggct ctgccccaccc cacagcagct cctggccatc	60
gagcagagtg tctacagctc agatcccttc cgacagggtc tcacccccc accatgcct	120
ggagaccaca tgcaccctta tggtgccgag cccctttcc atgacctgga tagcgacgac	180
acctccctca gtaacctggg tgactgttcc ctagcaaccc cagaagctgg gcctctgcag	240
tccagagtgaa gaaacccat 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: Synthetic
polynucleotide

```

<400> SEQUENCE: 65	
atggggagaa aaaagattca gattacgagg attatggatg aacgtaacag acaggtgaca	60
tttacaaaga ggaaatttgg gttgtatgaaag aaggcttatg agctgagcgt gctgtgtgac	120
tgtgagatttgc cgctgtatcat cttcaacagc accaacaagg tggccagta tgccagcacc	180
gacatggaca aagtgtttct caagtacacg ggttacaacg agccgtatga gagccggaca	240
aactcagaca tcgtggagac gttgagaaag aaggccctta atggctgtga cagcccagac	300
cccgatgcgg acgattccgt aggtcacacg cctgagttctg aggacaagta caggaaaatt	360
aacgaagata ttgatctaattt gatcagcagg caaagattgt gtgtgttcc acctcccaac	420
ttcgagatgc cagtctccat cccagttgtcc agccacaaca gtttgggtgta cagcaaccct	480
gtcagtcac tggaaaccc caacccatttgc ccaactggctc acccttctgcagaggaat	540
agtatgtctc ctgggtgtac acatcgaccc tcaagtgac gtaacacagg tggctgtatg	600
ggtggagacc tcacgtctgg tgcaggcacc agtgcaggaa acgggtatgg caatccccga	660
aactcaccacg gtctgttgtt ctcacccgtt aacttgaaca agaatatgca agcaaaatct	720
cctcccccaa tgaatttgg aatgaataac cgtaaaccac atctccgagt tcttattcca	780
ccaggcagca agaatacgtt gccatcagtgc tctgaggatg tcgacccgtt tttgaatcaa	840
aggataaata actcccaacttgc ggctcagtca ttggctaccc cagtggttcc cgtacact	900

---

-continued

---

cctactttac caggacaagg aatgggagga tatccatcg ccattcaac aacatatggt	960
accgagtaactctgtagt tgccagacctg tcatctctgt ctgggtttaa caccgccagc	1020
gtcttcacc ttggttcagt aactggctgg caacagcaac acctacataa catgccacca	1080
tctgcctca gtcagttggg agcttgact agcactcatt tatctcagag ttcaaatctc	1140
tccctgcctt ctactcaaag cctcaacatc aagtcagaac ctgtttctcc tcctagagac	1200
cgttaccacca ccccttcgag atacccacaa cacacgcgc acgaggcggg gagatctcct	1260
gttgcacagct tgagcagctg tagcagttcg tacgacggga ggcacccgaga ggatcacccg	1320
aacgaattcc actccccat tggactcacc agaccttcg cggacgaaag ggaaagtccc	1380
tcagtcaagc gcatgcgact ttctgaagga tggcaaca	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: Synthetic polynucleotide

<400> SEQUENCE: 66

atggcccagc ccctgtgccc gcccgtctcc gagtccttggg tgctctctgc ggctggggc	60
ccaaactccgcg gcccgcgccttccgacaag gactgcggc gctcccttgt ctgcgtccccca	120
gactcatggg gcageaccccc agccgacacgc cccgtggcga gccccgcgcg gccaggcacc	180
ctccgggacc cccggggcccc ctccgttaggt aggccgcggcgc cgccgcacgc cccgcctggc	240
agcgggcaga ggcagagcgc cagtgagcgg gagaaaactgc gcatgcgcac gctggccgc	300
gcctgcacg agctgcgcgc ctttctaccg ccgtccgtgg ccgcgcggg ccagagctg	360
accaagatcg agacgcgtgcg ccttgcatacg cgctatatacg gcccacgtgc ggccgtgcta	420
ggcctcagcg aggagagtct ccagcgcgg tgccggcagc ggggtacgc ggggtccct	480
cggggtgtgcc cgctgtgccc cgacgactgc cccgcgcaga tgcagacacgc gacgcaggct	540
gaggggcagg ggcaggggcg cgggctgggc ctggatccg ccgtccgcgc cggggcgcc	600
tggggatccc cgctgcctg cccggagaccc cgagctgcac ccgcgcgcgc cgacccgcct	660
cgctgttgc ccgaggccgc gtggccggaa gggcaggcga tggagccaag cccaccgtcc	720
ccgctcttc cgggcgcacgt gctggctctg ttggagaccc ggtatccct ctgcctctg	780
gagttggctgc 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: Synthetic polynucleotide

<400> SEQUENCE: 67

atgctggaaa tgctagaata taatcactat caggtgcaga cccacctcga aaacccacc	60
aagtaccaca tacagcaagc ccaacggcag caggtaaagc agtacccatc taccacttta	120
gcaaataaac atgccaacca agtccctgagc ttgccatgtc caaaccagcc tggcgatcat	180
gtcatgccac cggtgccggg gaggcgcgc cccaaacagcc ccatggctat gcttacgctt	240

---

-continued

---

aactccaact	gtgaaaaaga	gggattttat	aagtttgaag	agcaaaacag	ggcagagagc	300
gagtgeccag	gcatgaacac	acattcacga	gegtccctgtc	tgcagatgga	tgtatgtatc	360
gatgacatca	ttagccctaga	atcaagttat	aatgaggaaa	tcttgggctt	gatggatcct	420
gctttgaaa	tggcaaatac	gttgcctgtc	tcgggaaact	tgattgtatct	ttatggaaac	480
caaggctctgc	ccccaccagg	cctcaccatc	agcaactcct	gtccagccaa	ccttcccaac	540
ataaaaaggg	agctcacaga	gtctgaagca	agagcactgg	ccaaagagag	gcagaaaaag	600
gacaatcaca	acctgattga	acgaagaaga	agattnaaca	taaatgaccg	cattaaagaa	660
ctaggtactt	tgatcccata	gtcaaatacgat	ccagacatgc	gctggaacaa	gggaaccatc	720
ttaaaaagcat	ccgtggacta	tatccgaaag	ttgcaacgag	aacagcaacg	cgaaaaagaa	780
cttgaaaacc	gacagaagaa	actggagcac	gccaacccggc	atttggatgct	cagaatacag	840
gaacttggaaa	tgcaggctcg	agctcatgg	ctttccctta	ttccatccac	gggtctctgc	900
tctccagatt	tggtgaatcg	gatcatcaag	caagaacccg	ttcttgagaa	ctgcagccaa	960
gaccccttc	agcatcatgc	agacctaacc	tgtacaacaa	ctctcgatct	cacggatggc	1020
accatcacct	tcaacaacaa	cctcggaact	gggactgagg	ccaaccaacg	ctatagtgtc	1080
ccccacaaaa	tgggatccaa	actggaagac	atcctgtatgg	acgacaccct	ttctccctgc	1140
ggtgtcaactg	atccactcct	ttcctcagtg	tcccccggag	cttccaaaac	aagcagccgg	1200
aggagcagta	tgagcatgga	agagacggag	cacacttgt			1239

<210> SEQ\_ID NO 68  
<211> LENGTH: 1317  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 68

atgccccctca	acgttagctt	caccaacagg	aactatgacc	tcgactacga	ctcggtgcag	60
cctgttattct	actgegacga	ggaggagaac	ttctaccagg	agcagcagca	gagcgagctg	120
cagccccccg	cgccccagcga	ggatatctgg	aagaaattcg	agctgtgtcc	cacccccc	180
ctgtcccccta	gccggccgtc	cgggctctgc	tcgccccctc	acgttgcgg	cacacccttc	240
tcccttcggg	gagacaacga	cggcggtggc	gggagcttct	ccacggccga	ccagctggag	300
atgggtacccg	agctgtggg	aggagacatg	gtgaaccaga	gtttcatctg	cgaccggac	360
gacgagacct	tcataaaaaa	catcatcatc	caggactgta	tgtggagccg	cttctcgccc	420
gcccggcaagg	tcgtctcaga	gaagctggcc	tcctaccagg	ctgcgcgcaa	agacagccgc	480
agccccgaaacc	ccgccccggcgg	ccacagcgtc	tgctccacct	ccagcttgta	cctgcaggat	540
ctgagcgccg	ccgcctcaga	gtgcacatcgac	ccctcggtgg	tcttccctta	ccctctcaac	600
gacagcagct	cgcccaagtc	ctgcgcctcg	caagactcca	gogccttctc	tccgtctcg	660
gattctctgc	tctccctcgac	ggagtcctcc	ccgcaggggca	gccccgagcc	cctgggtgtc	720
catgaggaga	caccgccccac	caccagcagc	gactctgagg	aggaacaaga	agatgaggaa	780
gaaatcgatg	ttgtttctgt	ggaaaagagg	caggctctg	gcaaaaagg	tc aggtctgga	840
tcacattctg	ctggaggcca	cagcaaacct	cctcacagcc	cactggtcct	caagagggtgc	900

---

-continued

---

cacgtctcca cacatcagca caactacgca	gcccctccct	ccactcgaa ggactatcct	960
gctgccaaga gggtaagtt ggacagtgtc	agagtcctga	gacagatcag caacaaccga	1020
aaatgcacca gccccaggc ctcggacacc	gaggagaatg	tcaagaggcg aacacacaac	1080
gtcttggagc gccagaggag	gaacgagact	aaacggagct ttttgcct gcgtgaccag	1140
atcccgagt tgaaaacaa tgaaaaggcc	cccaaggtag	ttatcctaa aaaagccaca	1200
geatacatcc tgtccgtcca	agcagaggag	caaaagctca tttctgaaga ggacttgtt	1260
cgaaaaacgac	gagaacagtt	gaaacacaaa cttaaacgc tacggaaactc ttgtgcg	1317

<210> SEQ ID NO 69  
<211> LENGTH: 618  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 69

atggactacg actcgtacca	gcactattc	tacgactatg	actgcgggga ggatttctac	60
cgctccacgg cgccccagcga	ggacatctgg	aagaaattcg	agctggtgcc atcgcccccc	120
acgtcgccgc cctggggctt	gggtcccgcc	gcaggggacc	cgccccccgg gattggtccc	180
ccggagccgt ggccgggagg	gtgcacccgga	gacgaagcgg	aatccgggg ccactcgaaa	240
ggctggggca ggaactacgc	ctccatcata	cgccgtgact	gcatgtggag cggcttctcg	300
gcccgggaaac ggctggagag	agctgtgagc	gaccggctcg	ctcctggcgc gccccgggg	360
aacccgcca aggcgtccgc	cgcccggaac	tgcaactccca	gcctcgaagc cggcaacccg	420
gcgcggccg cccctgtcc	gctgggcgaa	cccaagaccc	aggcctgctc cgggtccgag	480
agcccaagcg actcgggtaa	ggacctcccc	gagccatcca	agagggggcc accccatggg	540
tggccaaagc tctgccccctg	cctgaggtca	ggcattggct	cttctcaagc tcttgggcca	600
tctccgcctc tctttggc				618

<210> SEQ ID NO 70  
<211> LENGTH: 1392  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 70

atgcccggat	tttccacgtc	tacgatgcca	ggaatgat	gcaagaaccc cgacttgag	60		
tttgcgttctt	tgcaaccatg	cttttatccg	gatgaagacg	acttttat	tttccacgtc	120	
gacagcaccc	ctcctggaga	ggacatctgg	aaaaaattcg	aactttgcc tacacccca	180		
ctcagtcctt	ctcgaggatt	tgccgaacac	agcagtgaac	cgccgttgc	ggtgacagag	240	
atgctctcg	agaacgaatt	gtggggaaac	cctgcggagg	aagacgtttt	cgggttcgt	300	
ggactcggag	gtctcacg	ccgcggcc	gaacccagtc	atactgcagg	attgcatgt	gtctcgat	360
tca	gactcggg	agaagcttgg	acggcgat	tctgagaaac	tccaaatcg	ccggggccct	420
ccaacacg	ggatgggg	acatccct	ggtgctggag	ccgcttagtcc	cgccggggaga	480	
ggccatgggg	gttctaccgc	acagtccct	ggtgctggag	ccgcttagtcc	cgccggggaga	540	

---

-continued

---

cacccggcg	ctgaatgtgt	agatcccgcg	gtagtgttgc	cgttccccgt	taataaggcga	600
gaaccggcac	cggtgccagc	cgttcctgcg	tctgcacccg	cggcagggtcc	tgctgtcgcc	660
tcaggagcag	gtattgccgc	tcctgcaggg	gcaccaggag	tagccccctcc	aaggccccggc	720
ggtaggcaaa	cctccggcg	cgaccacaaa	gcactctcaa	cgagcggaga	ggatacactg	780
tccgatagtg	atgacgagga	cgacgaaagag	gaggacgagg	aggaggagat	agatgttgc	840
acggtcgaga	agcgaaggag	tttttcaaat	acaaaagcgg	taacgacatt	cacgataaca	900
gtaagaccta	agaacgcgc	cctcggtcca	gggcgggccc	agtccagtga	gcttatactt	960
aagcgctgcc	tgccgattca	ccagcagcat	aactacgegg	cccctagtcc	ctacggttag	1020
agcgaggatg	ccccccca	aaaaaaaaata	aagtctgaag	cgtccccccg	ccccctgaaa	1080
tccgtaatcc	ccccaaaggc	gaagtcactc	agtcccagga	attcagattc	cgaggactcc	1140
gaacggccgc	ggaatcataa	catactttag	agacaacgcac	gcaatgaccc	gaggctttct	1200
tttttgaccc	tccgagatca	cgtccccag	ctggtaaga	atgagaaagc	tgcgaaaggta	1260
gtcataactga	aaaaggccac	cgagtatgtc	catagttgc	aagctgagga	gcaccagctt	1320
ctccttgaaa	aggagaaaact	tcaggcacga	caacagcaat	tgctgaaaaa	gattgagcat	1380
gcacgcactt	gt					1392

<210> SEQ ID NO 71  
<211> LENGTH: 960  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 71

atggagactac	tgtcgccacc	gctccgcgcac	gtagacctga	cggcccccga	cggctctctc	60
tgtcccttg	ccacaacgga	cgtttcttat	gacgaccgt	gtttcgactc	cccgacactg	120
cgtttctcg	aagaccttga	cccgccctcg	atgcacgtgg	gcccgcgtct	gaaacccgaa	180
gagcactcgc	acttccccgc	ggcgggtgcac	ccggcccccgg	gcccacgtga	ggacgagcat	240
gtgcgcgcgc	ccagggggca	ccaccaggcg	ggccgcgtcc	tactgtggc	ctgcaaggcg	300
tgcaagcgca	agaccaccaa	cgccgaccgc	cgcaaggccg	ccaccatgcg	cgagggcg	360
cgcctgagca	aagtaaatga	ggccttttag	acactcaagc	gtgcacgtc	gagcaatcca	420
aaccagcggt	tgcccaggt	ggagatcctg	cgcaacgcaca	tccgctatat	cgagggcctg	480
caggctctgc	tgcgcgacca	ggacgcccgc	ccccctggc	ccgcagccgc	cttctatgcg	540
ccggggccgc	tgccccccgg	ccggggccgc	gagcactaca	gcccgcgtcc	cgacgcgtcc	600
agcccgccgt	ccaaactgtc	cgacggccat	atggactaca	gcccgcgcgc	gagggccgc	660
ccggccggca	actgtacga	aggccctac	tacaacgagg	cgcccaagcga	acccaggccc	720
gggaagagtg	ccggcggtgc	gagcctagac	tgcctgtcca	gcatcgtgg	gcccgcgtcc	780
acccgagagcc	ctgcggcgcc	cgccctccgt	ctggcggacg	tgccttctga	gtgcctccg	840
cgcaggcaag	aggctgccgc	ccccagcgcag	ggagagagca	gcccgcaccc	cacccagtca	900
ccggacgcgc	ccccgcagtg	ccctgcgggt	gccaacccca	acccgatata	ccagggtgtc	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

atggagctgt atgagacatc cccctacttc taccaggaac cccgcttcta tgatggggaa	60
aactacctgc ctgtccaccc ccaggggcttc gaaccaccaag gctacgagcg gacggagctc	120
accctgagcc ccgaggcccc agggccctt gaggacaagg ggctggggac ccccgagcac	180
tgtccaggcc agtgectgcc gtggcggtgt aaggtgtgtta agaggaagt ggtgtccgtg	240
gaccggggc gggcgcccac actgagggag aagcgcaggc tcaagaaggt gaatgaggcc	300
ttcgaggccc tgaagagaag caccctgctc aaccccaacc agcggctgcc caaggtggag	360
atcctgccc gtgcccattca gtacatcgag cgcctccagg ccctgtctag ctccctcaac	420
caggaggagc gtgacccctcg ctaccggggc gggggcgccc cccagccagg ggtgcccagc	480
gaatgcagct ctcacagcgc ctcctgcagt ccagagtggg gcagtgact ggatccagc	540
gccaacccacag gggatcatct gctcacgggt gaccctacag atgcccacaa octgcactcc	600
ctcacctcca tcgtggacag catcacagtg gaagatgtgt ctgtggcctt cccagatgaa	660
accatgccc ac	672

<210> SEQ 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 gagtggctg atggcgagc ctcagccca aggtcctcca	60
agctggacag acgagtgtct cagttcttag gacgaggaggc acgaggcaga caagaaggag	120
gacgaccccg aagccatgaa cgcagaggag gactcactga ggaacggggg agaggaggag	180
gacgaagatg aggacctgga agaggaggaa gaagaggaag aggaggatga cgatcaaag	240
cccaagagac gcggccccaa aaagaagaag atgactaagg ctcgcctgga gcgtttaaa	300
ttgagacgca tgaaggctaa cggccggag cggaccgcg tgcacggact gaacgcggcg	360
ctagacaacc tgcgcaggt ggtgccttc tattctaaga cgcagaagct gtccaaatc	420
gagactctgc gcttggccaa gaactacatc tggctctgt cggagatcct gcgcctcagc	480
aaaagccacag acctggctc cttcggttag acgccttgca agggcttatac ccaacccacc	540
accaacccgg ttgcgggctg cctgcactc aatccctggg ctttctgcc tgagcagaac	600
caggacatgc ccccccaccc gccgacggcc agcgcttcct tccctgtaca cccctactcc	660
taccagtcgc ctggctgccc cagtcgcct tacggtagca tggacagctc ccatgtctc	720
cacgttaagc ctccgcgc cgcctacagc gcagcgctgg agcccttctt tgaaagccct	780
ctgactgatt gcaccagccc ttccctttagt ggacccctca gcccggcgct cagcatcaat	840
ggcaacttctt ctttcaaaaca cgaaccgtcc gccgagttt agaaaaattt tgcccttacc	900
atgcactatc ctgcagcgc acgtggcaggg gcccggcc acggatcaat cttctcaggc	960
accgctgccc ctcgcgtgcgaa gatccccata gacaatatta tgccttcga tagccatca	1020

---

-continued

---

catcatgagc gagtcatgag tgcccagtc aatgccatat ttcatgtat 1068

<210> SEQ ID NO 74  
<211> LENGTH: 711  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

&lt;400&gt; SEQUENCE: 74

atgccagccc gccttgagac ctgcatactcc gacactcgact ggcgcaggcagc cagcggcagt 60  
gaccttatccg gcttccctcac cgacgaggaa gactgtgcaca gactccaaca ggcagccctcc 120  
gttccggggc cgccccgcgc ggccccgcagg ggccgcgcaca atatctcccg ggcgtctgag 180  
gttccagggg cacaggacga cgacgaggag aggccggccgc ggcgcggcccg gacgcgggtc 240  
cgctccgagg cgctgtgcac ctcgcgtgcgc aggagccggc ggcgtcaaggc caacgatgc 300  
gagcgcacacc gcatgcaccaa cttgaacgcg gcccctggacg cactgcgcag cgtgtgccc 360  
tgcgttcccg acgacacccaa gctcacccaa atcgagacgc tgcgttgcgc ctacaactac 420  
atctgggctc tggccgagac actgcgcctg gccggatcaag ggctgcgcgg aggccggtgcc 480  
cgggagcggcc tcctgcccgc gcagtgcgtc ccctgcgtgc cccgtcccc aagccccggcc 540  
agcggcagccg cttccgaaga cttcacctac cggccggccgc accctgtttt ctccctccca 600  
agcctgcacca aagacttgct ccacacaaacg ccctgtttca ttcccttacca c 660  
711

<210> SEQ ID NO 75  
<211> LENGTH: 642  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

&lt;400&gt; SEQUENCE: 75

atgacaccac aaccatctgg tgctcccaca gtccaggta cgcgagagac tgaaagatca 60  
ttccccacgcg cgtccgagga tgaggtaaca tgtccaaacta ggcacccccc ctcttctacc 120  
cgggccccgcg ggaattgtgc tgaggccaa gaggaggat gcagaggagc accaaggaaa 180  
cttcgagccc gacgggggtgg aagaagccgc cccaaatctg agctgcgcct tagcaagcag 240  
cgccgcagtc ggaggaaaaa ggcaaacgc cgggaaaggat ataggatgca taatcttaat 300  
tctgctctgg acgctctgcg aggccgtactt cctactttcc cggatgacgc gaaattgacc 360  
aagatagaga ctctccgggtt tgcacataat tacatctgg ctcttacaca aacactgaga 420  
attgcgcgtc acagtcttta cgctcttgag ccacccggcc cgcactgtgg cgagctgggt 480  
agccccggccg gctctccctgg agactggggg tctttgtatt ctccctgtcag ccaagcggga 540  
tctttgagtc cggctgcctag tctcgaagaa agacccggac tccttgagc gacttttca 600  
gcatgtctgt cccctggctc attggcttc tcagactttt tg 642

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

---

-continued

---

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

<400> SEQUENCE: 76

atggccctgc	ctcccagccc	gctggccatg	gaatatgtca	atgactttga	cttgatgaag	60
tttgaggtaa	agcgggaacc	ctctgagggc	cgacctggcc	cacctacagc	ctcactggga	120
tccacacctt	acagctcagt	gcctccttca	cccacattca	gtgaaccagg	catggtaggg	180
geaaccgagg	gtacacgacc	aggtttggag	gagctgtact	ggcttgctac	cctgcagcag	240
cagcttgggg	ctggggaggc	attgggactg	agtccctgaag	aggccatgga	gctactgcaa	300
ggtcagggcc	oagtccctgt	tgtatggaccc	catggttact	acccagggag	cccagaggag	360
acaggagccc	agcacgttca	gttggcagag	cggtttcccg	acgcggcgct	tgtctcgatg	420
tctgtgcgag	aactaaaccg	gcagctgcgg	ggatgcggga	gagacgaggc	tctacgactg	480
aagcagaggc	gtcgaacgct	gaagaaccgt	ggctatgcgc	aagcatgtcg	ttccaagagg	540
ctgcaacaga	ggcgagggtct	tgaggccag	cgcggccgct	ttgcagocca	gctagatgcg	600
ctacgagctg	aagtagcacg	tttggcaaga	gagcgagatc	tatacaaggc	tgcgtgtgac	660
cggtctaacct	cgagtggccc	cgggtccggg	gatccctccc	acctttcct	ctgccccact	720
ttcttgtaca	aagttgtccc	c				741

<210> SEQ ID NO 77

<211> LENGTH: 1395

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 77

atgaacgcgc	agctgaccat	ggaagcgata	ggcgagctgc	acggggtgag	ccatgagccg	60
gtggccggcc	ctgcccaccc	gctggggcgc	agccccacg	cgcgccagctc	cgtggcgcac	120
cgcggcagcc	acctgcccccc	cgcgcaccccg	cgtccatgg	gcatggcgctc	cctgtgtggac	180
ggcggcagcg	gcggcgaggaa	ttaccaccac	caccacccggg	cccctgagca	cagcctggcc	240
ggccccctgc	atcccacccat	gaccatggcc	tgcgagactc	ccccaggtat	gagcatgccc	300
accacctaca	ccacccctgac	ccctctgcag	ccgcgtgcctc	ccatctccac	agtctcgac	360
aagttcccccc	accatcacca	ccaccacccat	caccaccacc	acccgcacca	ccaccagcgc	420
ctggcgccgca	acgtgagcgg	tagtttcacg	ctcatgcggg	atgagcgccg	gctggctcc	480
atgaataacc	tctataacccc	ctaccacaag	gacgtggccg	gcatggggcca	gagcctctcg	540
ccccccttcca	gctccgggtct	gggcagcata	cacaactccc	agcaagggtct	cccccaactat	600
ccccaccccg	ggccggccat	gcccacccgac	aagatgtca	cccccaacgg	cttcgaagcc	660
caccaccccg	ccatgtcg	ccgcacccgg	gagcagcacc	tcacgcac	ctcgccggc	720
atggtgccca	tcaacggcct	tcctccgcac	catccccacg	cccacctgaa	cgcggcaggc	780
cacgggcaac	tcctggcac	agcccgaggag	cccaaccctt	cggtgaccgg	cgcgcaggc	840
agcaaatggaa	gtaattcagg	gcagatggaa	gagatcaata	ccaaagaggt	ggcgcagcgt	900
atcaccacccg	agctcaagcg	ctacagcata	ccacaggcca	tcttcgcgc	gagggtgtc	960
tgccgctccc	aggggaccct	ctcggacctg	ctgcgcaccc	ccaaaccctg	gagcaaactc	1020

---

-continued

---

aaatccggcc	gggagacctt	ccggaggatg	tggaagtggc	tgcaggagcc	ggagttccag	1080
cgcatgtccg	cgctccgctt	agcagcatgc	aaaaggaaag	aacaagaaca	tgggaaggat	1140
agaggcaaca	cacccaaaaa	gcccgaggtg	gtcttcacag	atgtccagcg	tcgaactcta	1200
catgcaatat	tcaaggaaaa	taagcgtcca	tccaaagaat	tgcaaatacac	catttccag	1260
cagctggggt	tggagctgag	cactgtcagc	aacttcttca	tgaacgcaag	aaggaggagt	1320
ctggacaagt	ggcaggacga	gggcagctcc	aattcaggca	actcatctc	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

atgatgtctt	atcttaagca	accgccttac	gcagtcataatg	ggctgagttct	gaccacttcg	60
ggtatggact	tgctgcaccc	ctccgtgggc	tacccggggc	cctgggcttc	ttgtccccca	120
gcacacccccc	ggaaacagcg	ccgggagagg	acgacgttca	ctcgccgcga	gctagatgtg	180
ctgaaagcac	tgtttgccaa	gaccgggtac	ccagacatct	tcatgcgaga	ggaggtggca	240
ctgaaaatca	acttgcggca	gtcgagggtg	caggtatggt	ttaagaatcg	aagagctaag	300
tgccgcacac	aacagcaaca	acagcagaat	ggaggtcaaa	acaaagttag	acctgcacaa	360
aagaagacat	ctccagctcg	ggaagttagt	tcaagagatg	gaacaagtgg	ccaattcact	420
ccccccctcta	gcacctcagt	cccgaccatt	gccagcgcga	gtgtctctgt	gtctatctgg	480
agcccgacctt	ccatctcccc	actgtcagat	cccttgcaca	cctccctcttc	ctgcatgcag	540
aggtccatcc	ccatgaccta	tactcaggt	tcaggttata	gtcaaggata	tgctggctca	600
acttcctact	ttggggcat	ggactgtgga	tcatatttga	cccctatgca	tcaccagctt	660
cccgccacac	ggggccacact	cagtccatg	ggtaccaatg	cagtcaccag	ccatctcaat	720
cagtccccag	cttctctttc	cacccaggga	tatggagett	caagcttgg	tttaactca	780
accactgatt	gcttggatta	taaggacaa	actgcctcct	ggaagcttaa	cttcaatgct	840
gactgcttgg	attataaaga	tcagacatcc	tcgtggaaat	tccagggttt	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> SEQUENCE: 79

atggcgcccc	ttcccgccac	ggtaccgaga	atgatgcggc	cggctccggg	gcagaactac	60
ccccgcacgg	gattcccttt	ggaagtgtcc	accccgcttg	gcacaggccg	ggtaatcag	120
ctggggagggg	tcttcataaa	tggcgaccc	ctgccttaacc	acatccgcac	caagatagt	180
gagatggccc	accatggcat	ccggccctgt	gtcatctccc	gacagctgcg	tgtctccac	240
ggctgcgtct	ccaagattct	ttgcccgtac	caggagacccg	ggtccatccg	gcctggggcc	300

---

-continued

---

atcggeggca gcaagccag acaggtggcg actccggatg tagagaaaaa gattgaggag	360
tacaagggg aaaacccagg catgttcagc tggagatcc gggacaggtct gctgaaggat	420
gggcactgtc accgaagcac tggccctca gtgagttcgta tttagccggt gctcagaatc	480
aagtccggaa agaaagagga ggaggatgaa gggacaaga aggaggacga cggcgaaaag	540
aaggccaaac acagcatcga cggcatctg ggcgacaagaa ggaaccggct ggacgaggc	600
tccggatgtgg agtcggaaacc tgaccttcca ctgaagcgca akgcggcggc cagtcggacc	660
acattcacgg ccgagcagct ggaggagctg gagaaggcctt ttgagaggac ccactaccca	720
gacatataaca cccgcgagga gctggcgcag aggaccaagc tgacagaggc gcgtgtgcag	780
gtctggttca gtaaccggcg cgcccgttgg cgtaagcagg caggagccaa ccagctggcg	840
gcgttcaacc accttctgcc aggaggcttcc cccgcacccgg gcatgccccac gctgcccccc	900
taccagctgc cggactccac ctacccacc accaccatct cccaaagatgg gggcggcact	960
gtgcacccgc ctcagccctt gccaccgtcc accatgcacc aggggcggctt ggctgcagcg	1020
gtgcagccgc cccgacaccag ctctgcctac ggagccgc acagttctc cagctactct	1080
gacagcttca tgaatccggc ggcgcctcc aaccacatga aaccgggtcag caacggcctg	1140
tctcctcagg tgatgagcat cttggggcaac cccagtgccgg tgccccggca gccacaggct	1200
gacttctcca tctcccccgtc gcatggcggc ctggactcgg ccacccatct ctcagccagc	1260
tgcagccgc gggccgactc catcaagcca ggagacagcc tgccccaccc tcaggccctac	1320
tgcacccacca octacagcac caccggctac agcgtggacc ccgtggccgg ctatcgtac	1380
ggccagctacg gccagagtga gtgcctgggt ccctggcggtt ccccccgtccc cattcctct	1440
cccaccccca gggcctctg cttgtttatg gagagctaca aggtgggtgtc agggtgggaa	1500
atgtccattt cacagatggaa aaaatttgaag tccagccaga tggaaacagtt cacc	1554

```

<210> SEQ_ID NO 80
<211> LENGTH: 903
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    polynucleotide

```

<400> SEQUENCE: 80	
atgagttgcc aagctttac ttccggctgtat acctttatac ctctgaatc tgacgcctct	60
gcaactctgc ctctgataat gcatcacagt gctgccaggt gtctaccagt ctccaaaccat	120
gccaccaatg tgatgtctac agcaacagga cttcattttt ctgttccttc ctgtcatatat	180
ggaaaccagc catcaaccta tggagtgtatg gcaggtatgt taaccccttg tctttataaa	240
tttcctgacc acacctttagt tcatggattt ctccttatac accaggcctt tctggcagag	300
gaccccacag ctgctgatggt caagcaggaa ctcaggcgaa aaagtaattt ggtggaaagag	360
ccaaatagaca tggattctcc agaaatcaga gaacttgaaa agtttgccaa tgaatttaaa	420
gtgagacgaa ttaaattttagg atacacccag acaaattttt gggaggccctt ggcagctgt	480
catggctctg aattcagtca aacaacaatc tgccgatttggaaaatctgca gctcagctt	540
aaaaatgcata gcaaaatgaa agcaaatatta tccaaatggc tggaggaaagc tgagcaagta	600
ggagctttgt acaatgaaaaa agtggggagca aatgaaagga aaagaaaaacg aagaacaact	660

---

-continued

---

ataagcattg ctgctaaaga tgctctggag agacactttg gagaacagaa taaaccttct	720
tctcaagaga tcatgaggat ggctgaagaa ctgaatctgg agaaaagaat agtaagagtt	780
tggtttgca accggaggca gagagaaaaa cgggtgaaaaa caagtctgaa tcagagttta	840
tttctatcc ctaaggaaca tcttgagtgc agatcaggcc tcatgggcc agtttcttg	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: Synthetic polynucleotide

<400> SEQUENCE: 81

atggcgggac acctggcttc agattttgc ttctcgcccc ctccagggtgg tggagggtat	60
gggcgcagggg ggcgggagcc gggctgggtt gatcctcgaa cctggctaag cttccaaggc	120
cctcctggag ggccaggaat cgggcggggg gttggggcag gctctgaggt gtgggggatt	180
ccccccatgcc ccccgccgta tgagttctgt gggggatgg cgtaactgtgg gccccaggtt	240
ggagtggggc tagtgcccca aggccggctt gagacctctc agcctgaggg cgaaggcagga	300
gtcgggggtgg agagcaactc cgatggggcc tccccggagc cctgcaccgt caccctgg	360
gcccgtgaaggc tggagaagga gaagctggag caaaacccgg aggagtccca ggacatcaa	420
gctctgcaga aagaactcga gcaatttgc aagctcctga agcagaagag gatcaccctg	480
ggatatacac aggccatgtt ggggttcacc ctgggggttc tattttggaa ggtattcagc	540
caaacgacca tctgcgcctt tgaggctctg cagcttagct tcaagaacat gtgtaaagct	600
cggcccttgc tgcagaagtg ggtggaggaa gctgacaaca atgaaaatct tcaggagata	660
tgcaaaaggc aaccctcgat gcaggcccga aagagaaagc gaaccagtat cgagaaccga	720
gtgagaggca acctggagaa tttgttcctg cagtgcggc aaccacact gcagcagatc	780
agccacatcg cccagcagct tgggctcgag aaggatgtgg tccgagtgtg gttctgtaac	840
cggcgcaga agggcaagcg atcaagcgc gactatgcac aacgagagga ttttgaggct	900
gctgggtctc ctttctcagg gggaccagtg tccttcctc tggccccagg gccccattt	960
ggtatccccag gctatggag ccctcacttc actgcactgt actcctcggt cccttcct	1020
gagggggaaag ctttcccccc tgtctctgtc accactctgg gctctccat gcattcaa	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: Synthetic polynucleotide

<400> SEQUENCE: 82

atggcttcag acagcatatt tgagtcattt ctttcgtacc cacagtgcatt catgagagaa	60
tgcataacttg gaatgaatcc ttcttagagac gtccacgtat ccagcagcgg ccggcccttc	120
acggccgcctt ccaccgcgtt gagcccgaggc aagatgagcg aggctgtgg cgtggcgcc	180
ccggacgcggc ggcgtgcctt ggccggcaag ctgaggagcg ggcacgcag catggtgag	240

-continued

---

gtgctggccg accacccggg cgagctggtg cgcacccgaca gccccaaactt cctctgtcc	300
gtgctgccta cgcactggcg ctgcaacaag accctgccc tcgcttcaa ggtggtgcc	360
ctaggggatg ttccagatgg cactctggtc actgtgtatgg ctggcaatga tgaaaactac	420
tccggctgagc tgagaaatgc taccgcagcc atgaagaacc aggttgcag atttaatgac	480
ctcaggttg tcgggtcgaaag tggaagaggg aaaagcttca ctctgaccat cactgtcttc	540
acaaacccac cgcaagtgc cacctaccac agagccatca aaatcacagt ggatgggccc	600
cgagaacctc gaagacatcg gcagaaacta gatgatcaga ccaagccgg gagcttgtcc	660
tttccgagc ggctcagtga actggagcag ctgcggcgca cagccatgag ggtcagccca	720
caccacccag occccacgccc caaccctcggt gctccctgaa accactccac tgccttaac	780
cctcagccctc agagtcagat gcaggataca aggcatccc accgtggtcc	840
tacgatcagt octaccataa cctggatcc attgcctctc cttctgtgca cccagcaacg	900
cccatattcac ctggacgtgc cagcggcatg acaaccctct ctgcagaact ttccagtcga	960
ctctcaacgg cacccgacct gacagcgttc agcgaccggc gctcgttccc cgctgtggcc	1020
tccatctccg accccccgcgc gcaactatcca ggccgcctca octactccc gacgcggcgc	1080
acctcgggca tcggcatcgat catgtcgccg atgggctcgcc ccacgcgtca ccacacccat	1140
ctgccgcgc octaccccgctc ctcgtcgca ggcgcaggag gcccgttcca agccagctcg	1200
ccctccattacc acctgtacta cggcgccctcg gccggctctc accagttctc catggtggc	1260
ggcgagcgct cgccgcgcgc ctcgtcgccg ccctgcacca acgcctccac cggctccgcg	1320
ctgctcaacc ccagccccc gaaaccagac gacgtgggtgg aggccgaggc cagccacacg	1380
aactccccca ccaacatggc gccctccgcg cgcctggagg aggccgtgtg gaggcctac	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: Synthetic polynucleotide

<400> SEQUENCE: 83

atgtcgatgc tgccgtcggtt tggcttacg caggagcaag tggcggtcggt gtgcgagggtt	60
ctgcagcaag gcgaaacccctt ggagcgccctg ggcagggttcc tgggtactt gcccgcctgc	120
gaccacactgc acaagaacga gagegtactc aaggccaaagg cgggtggtcgc cttccacccgc	180
ggcaacttccgtt gtcgttccatca caagatccgtt gagagccacc agttctcgcc tcacaaccac	240
cccaaaactgc agcaactgtg gctgaaggcg cattacgtgg aggccgagaa gctgtgcggc	300
cgacccttgg ggcgcgtggg caaatatcggtt gtgcgcggaa aatttccact ggcgcgcacc	360
atctgggacg gcgaggagac cagctactgc ttcaaggaga agtcgagggg tgcctgtcg	420
gagtggtagc cgccacaatcc ctacccatcg ccgcgtgaga agcggggagct ggccgaggcc	480
accggccctca ccaccacccca ggtcagcaac tggtttaaga accggaggca aagagaccgg	540
gcccggagg ccaaggaaag ggagaacacc gaaaacaata actcctccctc caacaaggcag	600
aaccaactctt ctcctctggaa agggggcaag ccgcgtatgtt ccagctcaga agggaaattc	660
tacacccccc aaagtccaga ccagaactcg gtccttctgc tgcaggccaa tatggccac	720
gcgcaggagct ccaaactattc tctcccgggc ttaacagccct cgcagcccgag tcacggctg	780

-continued

```
cagacccacc agcatcaagt ccaagactct ctgtctggcc cccctcaccc tc cagttctggtg 840  
qacttqqqqt 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: Synthetic  
polynucleotide

<400> SEQUENCE: 84

atgtccatgc tgccccacctt cggettcacg caggagaca g tggcgtgcgt gtgcgaggtg  
ctgcagcagg gcgccaacat cgagcggtcg ggccgcgtcc tttggtcgtct gcgcgcctgc  
gagcacccatc acaagaatga aaggcggtctc aaggccaaagg ccgtgggtgc ctccaccgc  
ggcaacttcc gcgagctcta caagatcccg gagagccacc agttctcgcc gcacaaccac  
gccaagctgc agcagctgtg gctcaaggca cactacatcg aggccggagaa gctgcgcggc  
cgaccctgg gcgccgtggg caaataccgc gtgcgcgcga aattcccgct gccgcgcgtcc  
atctgggacg gcgaggagac cagctactgc ttcaaggaaa agagtcgcag cgtgtgcgc  
gagtggtagc cgcacaaccc ctacccttca ccccgcgaga agcgtgagct gacggaggcc  
acgggcctca ccaccacacca ggtcagcaac tgggtcaaga accggccggca ggcgcacccgg  
gcccccgagg ccaaggaaag ggagaacaac gagaactcca attctaaacag ccacaaccccg  
ctgaatggca gcccggcaagtc ggtgttaggc agtcggagg atgagaagac tccatcgaaa  
acgccagacc actcatcata cagccccgcga ctgcgcctca ccccgccggcc ccctggggctg  
ccgtccctgc acagcctggg ccaccctccg ggccccagcg cagtgccagt gccgggtgc  
ggccggaggtg gagcggaccc actgcacac caccatggcc tgcaggactc catccatcaac  
cccatgtca gccaacccctcgat ggacccctggcc 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: Synthetic  
polynucleotide

<400> SEQUENCE: 85

atggcgcgct ccttccttgtt caagaagcat ttcaacgcgt ccaaaaagcc aaactacagc	60
gaactggaca cacatacagt gattatttcc ccgtatctt atgagagttt ctccatgcct	120
gttcataccac aaccagagat cctcagctca ggagcataca gccccatcac tgggtggact	180
accgctgctc cattccacgc ccagctaccc aatggccctt ctcctttc cgatctcc	240
tcatctttgg ggcgagtgag tccccctcctt ccatctgaca ctcctccaa ggaccacagt	300
ggcttcgaaaa gccccattag tgatggaaag gaaagactac agtccaagct ttcagacccc	360
catgccattt aagctgaaaa gtttcagtgc aatttatgca ataagaccta ttcaactttt	420
tctgggttgtt ccaaaccataa gcagctgcac tggatgcggc agtctagaaa atctttcagc	480
tgttaataact gtgacaaggaa atatgtgagc ctggggcccc tgaagatgca tattcggacc	540
cacacattac ctqtgttttq caaqatctqc qqcaaqcqct tttccaaqacc ctqqtqctt	600

---

-continued

---

caaggacaca ttagaactca cacggggag aagcctttt cttgcctca ctgcaacaga	660
gcatttcag acaggtcaa tctgagggct catctgcaga cccattctga tgtaaagaaa	720
taccagtgc aaaaactgctc caaacacctc tccagaatgt ctctcctgca caaacatgag	780
gaatctggct gctgtgtac acac	804

```

<210> SEQ ID NO 86
<211> LENGTH: 1398
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

```

<400> SEQUENCE: 86	
atggcgagg agcaggacct atcggagggtg gagctgagcc cctgtgggtc ggaggaggccc	60
cgtgttgttccccggggag cgccgcctcg ctagggcccg acggcgccgg cggcgatcg	120
ggcctgcggc ccageccccc gccaggccgg ctgggcaagg tcaagaagga gcaggcaggac	180
ggcgaggccgg acgtgacaa gttccccgtg tgcatcccg aggccgtcag ccagggtgtc	240
agcggtacg actggacgct ggtgccatg cccgtgcgc tcaacggcgc cagcaaagc	300
aagccgcacg tcaagccgc catgaacgccc ttcatgggtg gggctcaggc agcgccgcagg	360
aagctegccg accagtaccc gcacccgtcac aacgctgagc tcagcaagac gctggcaag	420
ctctggaggc tgctgaacga aagtgacaag cgcgccttca tggaggaggc tgagggctc	480
cgtatgcacg acaagaaga ccacccggac tacaagtacc agcccgccgg gcggaaagAAC	540
gggaaggccg cccagggcga ggcggagtgc cccgggtggg aggccgagca aggtgggacc	600
gcgcgcatacc agggccacta caagagcgc cacttggacc accggcaccc aggagaggc	660
tccccatgt cagatggaa ccccgagcac ccctcaggcc agagccatgg cccacccacc	720
cctccaaacca ccccgaaagac agagctgcag tggggcaagg cagaccggaa gcgggacggg	780
cgtccatgg gggagggccgg gaagcctcac atcgacttcg gcaacgtgga cattggtgag	840
atcagecacg aggtaatgtc caacatggag acctttgtat tggctgagtt ggaccagtag	900
ctggccgcaca atgggcaccc aggcacatgt agcagctact cagcagccgg ctatggctg	960
ggcagtgcggc tggccgtggc cagtggacac tccgcctggc tctccaagcc accaggcgtg	1020
gtctgtccca cggctcacc acctgggtgt gatgccaag cccaggtgaa gacagagacc	1080
gcggggccccc agggggccccc acactacacc gaccagccat ccacccatcaca gatcgccat	1140
acctccatca gcctgcggcc caatggctca gccttccctt ccacccatccg ccccccagg	1200
gactactctg accatcagcc ctcaggaccc tattatggcc actcggccca ggcctctggc	1260
ctctactctg ctttcctcta tatggggccca tggcagccggc ccctctacac ggccatctct	1320
gaccccaagcc cctcaggccca ccagtcacac agccccacac actggagca gccagttat	1380
acgacactgt cccggccc	1398

```

<210> SEQ ID NO 87
<211> LENGTH: 951
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

```

---

-continued

---

<400> SEQUENCE: 87

atgtacaaca	tatggagac	ggagctgaag	ccgcggggcc	cgcagcaaac	ttcgaaaaaa	60
ggcggcggca	actccaccgc	ggcggcggcc	ggcggcaacc	agaaaaaacag	cccgacccgc	120
gtcaagcggc	ccatgaatgc	cttcatggtg	tggtccccgg	ggcageggcg	caagatggcc	180
caggagaacc	ccaagatgca	caactcggag	atcagcaagc	gcctggcg	cgagtggaaa	240
cttttgtcgg	agacggagaa	gcccgggttc	atcgacgagg	ctaageggct	gcgagcgctg	300
cacatgaagg	agcacccgga	ttataaatac	cgccccccgg	ggaaaaaccaa	gacgctcatg	360
aagaaggata	agtacacgct	gccccgggg	ctgctggccc	ccggcgccaa	tagcatggcg	420
agcgggggtcg	gggtggggcg	cggcctgggc	gccccgggtcg	accagcgcat	ggacagtta	480
gcccacatga	acggctggag	caacggcagc	tacagcatga	tgcaggacca	gctggctac	540
cggcagcacc	cgggcctcaa	tgccgcacggc	gcagcgcaga	tgcagccat	gcaccgcata	600
gacgtgagcg	ccctgcagta	caactccatg	accagctcg	agacctacat	gaacggctcg	660
ccccacctaca	gcatgtccta	ctcgcagcag	ggcacccctg	gcatggctct	tggctccatg	720
ggttcggtgg	tcaagtccga	ggccagctcc	agccccccctg	tggttacctc	ttccctccac	780
tccagggcgc	cctgccaggc	cggggaccc	cgggacatga	tcaagatgta	tctcccccgg	840
gcccgggtgc	cggaacccgc	cggcccccgg	agacttcaca	tgtcccagca	ctaccagagc	900
ggcccggtgc	cgggcacggc	cattaacggc	acactgcccc	tctcacacat	g	951

<210> SEQ\_ID NO 88

<211> LENGTH: 1338

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 88

atgcgacctg	tgcgagagaa	ctcatcagg	gcgagaagcc	cgccgggttc	tgcgtatgg	60
gcccggcga	ttttgataag	cctacccttc	cegcggact	cgctggccca	caggecccca	120
agctccgctc	cgacggagtc	ccagggcatt	ttcacccgtg	ccgctccagc	cccgggagcg	180
ccttctctctc	ccggcacgct	ggcgcaccc	cttcccgccc	cgcaatgta	cagccttctg	240
gagactgaac	tcaagaaccc	cgttagggaca	cccacacaag	cgccggggcac	cgccggcccc	300
gcagccccgg	gaggcgcagg	caagactgt	gcgaacgcag	ccggcgccgc	gaactcgccc	360
ggccggcagca	cgccgttgtc	gagcggaggt	ggccggggta	cagaccagg	ccgtgtggaa	420
cgcccccata	acgccttcata	ggtatggtcc	cgccggcagc	ggcgcacaaat	ggccctggag	480
aaccccaaga	tgcacaattc	tgagatcagc	aagcgcttgg	gcgcgcactg	gaaactgtcg	540
accgacgcgg	agaagcgacc	attcatcgac	gaggccaagc	gacttcgcgc	cgtgcacatg	600
aaggagtata	oggactacaa	gtaccgaccc	cgccgcaga	ccaaagacgt	gctcaagaaa	660
gataagtact	ccctgcccag	cgccctctg	cctcccggtg	ccgcggccgc	cgccgcgc	720
gcccggcccg	cagccgctgc	cgccagcagt	ccgggtgggc	tggccagcg	cctggacacg	780
tacacgcacg	tgaacggctg	ggccaacggc	gcgtactcg	tggtgccagga	gcagctgggc	840
tacgcgcaggc	ccccgagcat	gagcagcccg	ccgcgcgcgc	ccgcgcgtgc	gccgatgcac	900

-continued

---

cgctacgaca	tggccggcct	gcagtagc	ccaatgtac	cgccggcgc	tcagagctac	960
atgaacgtcg	ctgcccggc	cgcgcgcgc	tcgggctacg	ggggcatggc	gccctcagcc	1020
acagcagccg	cggccggcgc	ctacggcag	cagccgcaca	ccggcggcgc	cgcagctgcg	1080
gccccagccg	ccatgagcct	ggggccatg	ggctcggtag	tgaagtctga	gcccaagctcg	1140
ccggccggcg	ccatcgcatc	gcactctcg	cgcgcgtgc	tcggcgcacct	gcgcgcacatg	1200
atcagcatgt	acctgccacc	cgggggggac	gccccggac	ccgcctctcc	gctggccggc	1260
ggtcgcgtgc	acggcggtgca	ccagcaactac	caggcgcccg	ggactgcagt	caacggaaacg	1320
gtggcgctga	cccacatc					1338

&lt;210&gt; SEQ ID NO 89

&lt;211&gt; LENGTH: 813

&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 polynucleotide

&lt;400&gt; SEQUENCE: 89

atgttacagg	cgtcaaaaat	ggaagggttt	ccctctgtcc	cccctcagcc	atcagaagac	60
ctgggtgcct	atgacacgg	tctataccaa	cgc当地aaacgc	acgagtattt	cccttatctc	120
agcagtgtat	gggagagcca	tagcgaccat	tactggact	tccacccca	ccacgtgcac	180
agcgagttcg	agagcttcgc	cgagaacaac	ttcacggagc	tccagagcgt	gcagcccccg	240
cagctgcagc	agctctaccg	ccacatggag	ctggagcaga	tgcacgtcct	cgataacccc	300
atgggtgcac	cccatccag	tcttgccac	caggtctcct	acctgccccg	gatgtgcctc	360
cagtacccat	ccctgtcccc	agccagccc	agctcagatg	aggaggaggg	cgagccgcag	420
agccccccac	tggaggtgtc	tgacggcag	gccccatggcc	tggagccgg	gcctggctc	480
ctgcctgggg	agacaggcag	caagaagaag	atccgcctgt	accagttcct	gttggacctg	540
ctccgcagcg	gc当地acatgaa	ggacagcata	tgggggtgg	acaaggacaa	gggcaccc	600
cagttctcgt	ccaaggaccaa	ggaggcgctg	gccc当地ccgt	ggggcatcca	gaagggcaac	660
cgcaagaaga	tgacctacca	gaagatggcg	cgc当地gtgc	gcaactacgg	caagacgggc	720
gaggtcaaga	aggtgaagaa	gaagctcacc	taccagttca	gccccggcgt	gctggccgc	780
ggggccctgg	ccgagccgc	ccacccgc	cac			813

&lt;210&gt; SEQ ID NO 90

&lt;211&gt; LENGTH: 789

&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 polynucleotide

&lt;400&gt; SEQUENCE: 90

atgctcgccc	tggaggctgc	acagctcgac	ggccacact	tcagctgtct	gtacccagat	60
ggcgtcttct	atgacctgga	cagctgcag	cattccagct	accctgatc	agaggggct	120
cctgactccc	tgtggactg	gactgtggcc	ccacctgtcc	cagccacccc	ctatgaagcc	180
ttcgacccgg	cagcagccgc	tttagccac	ccccaggctg	ccacgtctg	ctacgaaccc	240
cccacctaca	gccctgcagg	gaacctcgaa	ctggccccc	gcctggaggc	ccggggcc	300

---

-continued

---

ggcctccccg	catacccccac	ggagaacttc	gttagccaga	ccctggttcc	cccgccatat	360
gccccgtacc	ccagccctgt	gctatcagag	gaggaagact	taccgttgaa	cagccctgcc	420
ctggagggtct	cgacacgcga	gtcgatgg	gccctcggtt	ctggcccccga	ggggaaaggga	480
tccgaggcag	ggactcgcaa	gaagctgcgc	ctgtaccagt	tcctgtggg	gctactgacg	540
cgccgggaca	tgcgtgagtg	cgtgtgggtt	gtggagccag	gcccggcgt	cttccagttc	600
tcctccaagc	acaaggaaact	cctggcgcgc	cgctggggcc	acgagaaggg	gaaccgcaag	660
cgcatgacct	accagaagct	ggcgcgcgc	ctccgaaact	acgccaagac	cggcgagatc	720
cgcaagggtca	agcgeaagct	cacctaccag	ttegacagcg	cgctgtgcc	tgcagtccgc	780
cgggccttg						789

<210> SEQ ID NO 91  
<211> LENGTH: 744  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 91						
atgacgtgtg	ttgaacaaga	caagctgggt	caagcatttg	aagatgttt	tgagggtctg	60
aggcaacatt	caactggaga	tcttcagtagc	tcggccagatt	acagaaatta	cctggcttta	120
atcaaccatc	gtcctcatgt	caaaggaaat	tccagctgt	atggagtgtt	gcctacagag	180
gagcctgtct	ataattggag	aacggtaatt	aacagtgt	cggaacttcta	ttttgaagga	240
aatatttcatc	aatctctgca	gaacataact	gaaaaccaggc	tggtacaacc	cacttttctc	300
cagcaaaagg	ggggaaaagg	caggaagaag	ctccgactgt	ttgaataacct	tcacgaatcc	360
ctgtataatc	cggagatggc	atcttgtatt	cagtggtag	ataaaaaccaa	aggcatcttt	420
cagtttgtat	aaaaaacaa	agaaaaactt	gccgagctt	ggggaaaag	aaaaggcaac	480
aggaagacca	tgacttacca	gaaaatggcc	agggcactca	gaaattacgg	aagaagtggg	540
gaaattacca	aaatccggag	gaagctgact	taccagttca	gtgaggccat	tctccaaaga	600
ctctctccat	cttatttcct	ggggaaaagag	atcttctatt	cacagtgtgt	tcaacctgat	660
caagaatatac	tcaagttaaa	taactggaat	gcaaattata	attatacata	tgccaattac	720
catgagactaa	atcaccatga	ttgc				744

<210> SEQ ID NO 92  
<211> LENGTH: 612  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 92						
atgcaatcat	atgcttctgc	tatgttaagc	gtattcaaca	gogatgat	tttgcagct	60
gtgcaagaga	atattccgc	tctccggaga	agctcttcct	tccttgcac	tgaaagctgt	120
aactctaagt	atcagtgtga	aacggggagaa	aacagtaaaag	gcaacgtcca	ggatagagtg	180
aagcgaccca	tgaacgcatt	catcgtgtgg	tctcgccatc	agaggcgcaa	gatggctcta	240
gagaatccca	gaatgcgaaa	ctcagagatc	agcaagcagc	tgggataccca	gtggaaaatg	300

-continued

---

cttactgaag ccgaaaaatg gccattcttc caggaggcac agaaattaca ggccatgcac	360
agagagaaaat acccgaaatta taagtatcga cctcgctgga aggccaagat gctgccgaaag	420
aattgcgtt tgctccgcg agatcccgt tcggtaactct gcagcgaagt gcaactggac	480
aacagggtgt acagggatga ctgtacgaaa gccacacact caagaatgga gcaccagcta	540
ggccacttac cgcccatcaa cgccaggcgc tcaccgcgc aacgggaccg ctacagccac	600
tggacaaagg 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: Synthetic polynucleotide

<400> SEQUENCE: 93	
atggccgacg cagacgaggg ctttggcctg ggcacacgc ctctggagcc tgacgaaaaa	60
gacctggccct gcgattcgaa acccgagago ggcgtcgggg cccccagcaa gtcccgctcg	120
tccccgcagg cgccttcac ccagcaggc atggagggaa tcaaagtgtt tctccatgaa	180
agagaactgt ggctaaaatt ccacgaagtg ggcacggaaa tgatcataac caaggctgga	240
aggcggatgt ttcccagttt caaagtgaag gtgacgggca ttaatccaa aacgaagtac	300
attcttcata tggacattgt acctgcccac gatcacatg acaaattcgc agataataaa	360
tggctgtga cggccaaagc tgagccgcg atgcctggcc gcctgtacgt gcacccagac	420
tcccccgcgc cccggggcgca ttggatgagg cagctcgct cttccagaa actcaagctc	480
accaacaacc acctggaccc atttggcat attattctaa attccatgca caaataccag	540
ccttagattac acatcgtgaa agcggatgaa aataatggat ttggctaaa aaatacagcg	600
ttctgcactc acgtcttcc tgagactcg tttatagcag tgacttcata ccagaaccac	660
aagatcacgc aattaaagat tgagaataat cccttgccca aaggattcg gggcgttgat	720
gacatggagc tgcacagaat gtcaagaatg caaagtaaag aatatccgt ggtccccagg	780
agcacccgtga ggccaaaagt ggcctccaaac cacagtcttc tcaagcgcga gtctcgagct	840
ctctccacccat catccaattt ggggtccaa taccagtgtt agaatggtgt ttccggcccc	900
tcccaaggacc tcctgcctcc acccaaccca tacccactgc cccaggagca tagccaaatt	960
taccattgtt ccaagaggaa agaggaagaa tggccatcca cagaccatcc ctataagaag	1020
ccctcatgg agacatcacc cagtgaagaa gattccttc accgctctag ctatccacag	1080
cagcagggcc tgggtgcctc ctacaggaca gagtcggcac agcggcaagc ttgcgttat	1140
gcacagctcg cggcccccag cgacgctgtg cccagcttag aggacatcg ctgcaacacg	1200
tggccaagca tgccttcata cagcagctgc accgtcacca ccgtgcagcc catggacagg	1260
ctacccttacc agcaacttcc cgctcaacttc acctcgggcc ccctggccc tcggctggct	1320
ggcatggcca accatggctc cccacagctg ggagagggaa tggccatcca ccagaccc	1380
gtggccacc accctgttgtt cagggcgtgt gggcctcaga ctggcctgca gtccctggc	1440
acccttcagc cccctgagtt cctctactct catggcgtgc caaggactct atccctcat	1500
cagtaccact ctgtgcacgg agttggcatg gtgccagat ggagcgcacaa tagctt	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 atggaatcc	cggggtcccc cacctcttcc	ccgcggggca gcacctctac	120
agccccgcgc caccctctc	ccacactgga gtcgccaat	atcageccgc accctacttt	180
caccctccct accagcagct	ggcctactcc cagtcggccg	acccctactc gcatctgggg	240
gaagcgtacg ccgcgcgcatt	caaccccttg caccagccgg	cgccccacagg cagccagcag	300
caggcctggc cggcgccca	gagccaggag ggagcggggg	tgcctcgca ccacggggcgc	360
ccggccggcc tactgccccca	cctctccggg ctggaggcg	gagcggtgag cgcccgccagg	420
gatgcctacc gccgcetccga	cctgctgtcg ccccacgcac	acgcctcgga tgccgcgggc	480
ctggccgaga acctggggct	ccacgacatg cctcaccaga	tggacgaggt gcagaatgtc	540
gacgaccagc acctgttgct	gcacgatcg acagtcatcc	gaaagggtcc catttccatg	600
accaagaacc ctctgaacct	cccctgttag aaggagctgg	tggggggcgt aatgaacccc	660
actgaggctct tctgtcgat	cccttggaaa ttgtcgctcc	tcaatctac gtctaaatac	720
aaagtgtacag tggctgaagt	acagaggcga ctgtcccac	ctgaatgtt aaatgcctcg	780
ttactggggat gtgttctcg	aagagccaaa tcgaaaaatg	gaggccgggtc cttggggag	840
aagttggaca agattgggtt	gaatcttccg gcccggaggc	ggaaagccgc tcatgtgact	900
ctcctgacat ctttagtaga	aggtaatgtt gttcatttt	ctaggactt tgcctatgtc	960
tgtgaagccg aatttccat	taaaccatgt gcagaatatt	taaccagacc tcatcttgg	1020
ggacgaaatg agatggcagc	taggaagaac atgctattgg	cgccccagca actgtgtaaa	1080
gaattcacag aacttctcg	ccaagaccgg acacccatg	ggaccagcag gtcgcacca	1140
gtcttggaga cgaacataca	gaactgttg tctcatttc	gcctgattac ccacgggttt	1200
ggcagccagg ccatctgtgc	cgccgtgtct gcccgtcaga	actacatcaa agaaggctg	1260
attgtcatag acaaattcta	catgaaccct ggagaccaga	gtccagctga ttctaaca	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

&lt;400&gt; SEQUENCE: 96

ggtcagcagc atcggttca acataaac

27

&lt;210&gt; SEQ\_ID NO 97

&lt;211&gt; LENGTH: 25

&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 primer

&lt;400&gt; SEQUENCE: 97

tctccgttgt cctgaaggac acata

25

&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;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

&lt;400&gt; SEQUENCE: 98

agccatgtgg tctctctgg ttgttatg

28

&lt;210&gt; SEQ\_ID NO 99

&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 primer

&lt;400&gt; SEQUENCE: 99

tttgtgggcc tgaagaaaac t

21

&lt;210&gt; SEQ\_ID NO 100

&lt;211&gt; LENGTH: 22

&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 primer

&lt;400&gt; SEQUENCE: 100

cttgaatccc gaatggaaag gg

22

&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 primer

&lt;400&gt; SEQUENCE: 101

tacagcatgt cctactcgca g

21

&lt;210&gt; SEQ\_ID NO 102

&lt;211&gt; LENGTH: 22

&lt;212&gt; TYPE: DNA

&lt;213&gt; 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

cagcggaaAC cccAACAGTT A

21

<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

cgagAGCTAC acGTTcacGG

20

<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

agccaACCTT aactgaggAG T

21

<210> SEQ ID NO 108  
<211> LENGTH: 21

-continued

---

<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

gaagtttcgc agacctgaca t

21

<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

atcgcttcc 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

ctgagaccccg 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

cacgatctca tacctggcct gcttc

25

-continued

---

<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 cagggtcagt 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

aggttgctgc tggtgaggc att

23

<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

26

<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

gtgtatatcc cagggtgatc ctc

23

<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

gaggaagagg taaccacagg g

21

---

-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

atgtccctct tgcgcac ac

22

<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

gagggtcagt agaacatgcg t

21

<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

tgccttttc ttagggcaga g

21

<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 atctggcggtt 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 acccattcct 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

gtatgcacca ttcaactcct cg 22

<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

ctcgtaactgg atgggtgaac t 21

<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

tgaatctcg a cggttcctc c 21

---

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.

\* \* \* \* \*