

GENERATION OF HUMAN iPS CELLS BY A SYNTHETIC SELF-REPLICATIVE RNA

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims priority to U.S. Provisional Application Serial Nos. 61/649,876, filed May 21, 2012 and 61/798,229, filed March 15, 2013, each of which are hereby incorporated by reference in their entirety.

TECHNICAL FIELD

[0002] Provided are methods and compositions useful for producing and propagating stem cells from fibroblasts. The disclosure relates to the production of induced pluripotent stem cells (iPS) and methods of use thereof.

BACKGROUND

[0003] Stems cells are a potential source from which organs may be regenerated, tissues may be repaired, biological factors prepared or delivered and disease or disorders treated.

SUMMARY

[0003a] In one aspect, there is provided a purified synthetic RNA replicon comprising: at least one non-structural replicase domain from an alphavirus and at least one non-alphavirus heterologous sequence encoding one or more factors for inducing the generation of pluripotent stem cells when expressed in a somatic cell.

[0003b] In another aspect, there is provided a composition comprising human cells transfected with a replicon as described herein.

[0003c] In yet another aspect, there is provided an in vitro method comprising: transfecting a human somatic cell with a purified synthetic ectopic self-replicating RNA replicon derived from an alphavirus, comprising polynucleotides encoding at least the following four de-differentiation factors (i) KLF4, (ii) OCT4, (iii) SOX2, (iv) c-MYC or n-MYC or L-MYC, and optionally (v) GLIS1 and (vi) NANOG, culturing the somatic cell to express the de-differentiation factor; selecting cells that display a stem cell morphology and/or stem cell markers; and subculturing the cells to obtain a population of induced stem cells.

[0003d] In yet another aspect, there is provided a vector system for producing human stem cells, comprising: at least one purified synthetic self-replicating RNA replicon derived from an alphavirus, comprising one or more polynucleotides encoding de-differentiation factors selected from the group consisting of a KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, GLIS1, NANOG or any combination thereof.

[0003e] In yet another aspect, there is provided an isolated human somatic cell comprising a purified synthetic ectopic RNA replicon derived from an alphavirus comprising one or more de-differentiation polynucleotide sequences.

[0003f] In yet another aspect, there is provided a cell population comprising the cells as described herein.

[0004] Generation of induced Pluripotent Stem (iPS) cells from patients is important to use stem cells therapeutically. Generation of iPS cells requires expression of several pluripotent transcription factors or Reprogramming Factors (RFs), including Oct4, Sox2, Klf4, cMyc, Glis1 (and potentially Nanog and Lin28). However, due to concerns with integration of DNA vectors (viruses and naked DNA) into the genome during iPS cell generation excludes these approaches from being subsequently used in patients.

[0005] The disclosure describes an approach to generate induced Pluripotent Stem (iPS) cells by ectopically expressing RFs using a synthetic self-replicating RNA from a modified alphavirus (e.g., Venezuelan Equine Encephalitis (VEE) virus). The alphavirus was designed to express, in one example, four RFs that resulted in the following advantages over mRNA transfection approaches: 1) utilized a single RNA species capable of self-replicating for a limited number cell divisions, thereby reducing the number of transfections; 2) is capable of encoding at one, two, three, four, or more RF open reading frames (ORFs); and 3) consistently expressed all the RF genes at high threshold levels over multiple

cellular divisions. By using the self-replicating backbone of an alphavirus (the structural genes being removed) to express the RFs requires only 3 to 4 transfections (and even only 1 or 2) into primary human fibroblasts to generate iPS cells. The generation of the alphavirus RF-RNA transcript utilizes a SP6 (or T7) *in vitro* transcription kit that does not require special conditions and thereby, further simplifies the approach for broad use. By expressing the four RFs at consistent, high levels over time in the same cell combined with replication of the alphavirus-RF RNA for a limited number of multiple cell generations, the alphavirus-RF RNA approach solves both of the major inefficiency problems associated with attempting to generate iPS cells by daily repeated daily transfections for >14 days of four individual RF mRNAs. The alphavirus-RF RNA is an ectopic approach that does not utilize a DNA intermediate and therefore, there is no opportunity for integrative mutation that can occur with DNA vector-based iPS cell approaches. Moreover, the timing of RNA replicon loss by degradation can be regulated by B18R withdrawal from the media. Using this approach, >100 independent iPS cell clones were generated from both OCT4/KLF4/SOX2/c-MYC and OCT4/KLF4/SOX2/GLIS1 alphavirus-RF RNA protocols from two independent parental human fibroblast populations. In addition, the approach can be engineered to express alternative RF combinations and/or insertion of additional RF ORFs into the RF-RNA backbone for refining iPS cell generation from specific cell types or for use in driving transdifferentiation.

[0006] The disclosure provides an alphavirus replicon RNA comprising at least one non-structural replicase domain from an alphavirus and at least one non-alphavirus heterologous sequence encoding factors for inducing the generation of pluripotent stem cells when expressed in a somatic cell. In one example, the replicon comprises sequences obtained from an alphavirus selected from the group consisting of Eastern Equine Encephalitis virus (EEE), Venezuelan Equine Encephalitis virus (VEE), Everglades virus, Mucambo virus, Pixuna virus and Western Equine Encephalitis virus (WEE). In another example, the replicon comprises sequences obtained from an alphavirus selected from the group

consisting of Sindbis virus, Semliki Forest virus, Middelburg virus, Chikungunya virus, O'nyong-nyong virus, Ross River virus, Barmah Forest virus, Getah virus, Sagiya virus, Bebaru virus, Mayaro virus, Una virus, Aura virus, Whataroa virus, Babanki virus, Kyzylagach virus, Highlands J virus, Fort Morgan virus, Ndumu virus and Buggy Creek virus. In yet another example, the at least one non-alphavirus heterologous sequence comprises at least 2, 3, 4 or 5 non-alphavirus heterologous sequences. In yet another example, of any of the foregoing the non-alphavirus heterologous sequence is selected from a polynucleotide encoding a KLF polypeptide, a SOX-2 polypeptide, a OCT-3/4 polypeptide, a c-MYC or n-MYC or L-MYC polypeptide, a GLIS1 polypeptide, a NANOG polypeptide and any combination thereof. In a further example, the polynucleotide encoding the KLF polypeptide encodes a KLF polypeptide having at least 95% identity to a sequence of SEQ ID NO:8. In another example, the polynucleotide encoding the KLF polypeptide encodes a KLF polypeptide having a sequence of SEQ ID NO:8. In yet another example, the polynucleotide encoding the KLF polypeptide comprises a sequence as set forth in SEQ ID NO:7, wherein "T" is "U". In another example, the polynucleotide encoding the SOX-2 polypeptide encodes a SOX-2 polypeptide having at least 95% identity to a sequence of SEQ ID NO:6. In another example, the polynucleotide encoding the SOX-2 polypeptide encodes a SOX-2 polypeptide having a sequence of SEQ ID NO:6. In yet another example, the polynucleotide encoding the Sox-2 polypeptide comprises a sequence as set forth in SEQ ID NO:5, wherein "T" is "U". In another example, the polynucleotide encoding the OCT-4 polypeptide encodes a OCT-4 polypeptide having at least 95% identity to a sequence of SEQ ID NO:4. In a further example, the polynucleotide encoding the OCT-4 polypeptide encodes a OCT-4 polypeptide having a sequence of SEQ ID NO:4. In a further example, the polynucleotide encoding the OCT-4 polypeptide comprises a sequence as set forth in SEQ ID NO:3, wherein "T" is "U". In another example, the polynucleotide encoding the c-MYC polypeptide encodes a c-MYC polypeptide having at least 95% identity to a sequence of SEQ ID NO:10. In a further example, the polynucleotide encoding the c-MYC polypeptide

encodes a c-MYC polypeptide having a sequence of SEQ ID NO:10. In yet a further example, the polynucleotide encoding the c-MYC polypeptide comprises a sequence as set forth in SEQ ID NO:9, wherein "T" is "U". In another example, the polynucleotide encoding the GLIS1 polypeptide encodes a GLIS1 polypeptide having at least 95% identity to a sequence of SEQ ID NO:34. In a further example, the polynucleotide encoding the GLIS1 polypeptide encodes a GLIS1 polypeptide having a sequence of SEQ ID NO:34. In yet a further example, the polynucleotide encoding the GLIS1 polypeptide comprises a sequence as set forth in SEQ ID NO:33, wherein "T" is "U". In another example, the polynucleotide encoding the NANOG polypeptide encodes a NANOG polypeptide having at least 95% identity to a sequence of SEQ ID NO:2. In a further example, the polynucleotide encoding the NANOG polypeptide encodes a NANOG polypeptide having a sequence of SEQ ID NO:2. In yet a further example, the polynucleotide encoding the NANOG polypeptide comprises a sequence as set forth in SEQ ID NO:1, wherein "T" is "U". In one embodiment example of any of the foregoing, the replicon comprises from 5' to 3': (VEE RNA replicases)-(promoter)-(RF₁)-(self cleaving peptide)-(RF₂)-(self cleaving peptide)-(RF₃)-(IRES or core promoter)-(RF₄)-(IRES or optional promoter)-(optional selectable marker)-(VEE 3'UTR and polyA tail)-(optional selectable marker)-promoter; wherein RF₁₋₄ are factors that induce de-differentiation of a somatic cell to a pluripotent cells, wherein RF₂₋₃ are optional, RF₃₋₄ are optional, or RF₄ is optional; wherein RF₁₋₄ are selected from the group consisting of Oct-4, Klf4, Sox-2, c-Myc, Nanog, and Glis1. In another example, the replicon comprise a sequence that is 90%, 95%, 98%, 99% or 100% identical to SEQ ID NO:29, 30, 31, or 32 from about position 1 to about position 7561 wherein "T" of the sequence is substituted with "U", followed by one or more RFs, followed by a 3'UTR and polyA tail, wherein the one or more RFs are selected from the group consisting of Oct-3/4, Sox-2, Klf4, c-Myc, Nanog, and Glis1; wherein when more than one RF is present, the coding sequences may be separated by an internal ribosome entry site (IRES) or a small promoter. In a further example, the replicon comprise a sequence that is at least 95%,

98%, 99% or 100% identical to a sequence selected from the group consisting of SEQ ID NO:29, 30, 31, or 32, wherein "T" is "U".

[0007] The disclosure also provides a composition comprising human cells transformed with a replicon as described in any of the foregoing examples and examples further described herein. In one example, the composition further comprises B18R conditioned media. In another example, the human cells are somatic cells. In a further example, the human cells are fibroblasts.

[0008] The disclosure also provides a method of making stem cells comprising culturing the composition described above and elsewhere herein, for at least 30 days under conditions to express the coding sequences of the replicon and isolating stem cells.

[0009] The disclosure also provides a method of making stem cells comprising transforming somatic cells with a replicon of the disclosure, culturing the somatic cells under conditions to promote expression of the replicon and isolating stem cells. In one example, the culturing comprise culturing the cells in media conditioned with B18R. In another example, the B18R conditioned media is produced by transfection of B18R mRNA into primary human fibroblasts.

[0010] The disclosure also provides isolated stem cells obtained from the methods described herein, wherein the stem cells are retroviral DNA- or RNA-free.

[0011] The disclosure also provides a method comprising contacting a human somatic cell with an ectopic self-replicating RNA replicon comprising polynucleotides encoding at least four de-differentiation factors selected from the group consisting of a (i) KLF4, (ii) OCT4, (iii) SOX2, (iv) c-MYC or n-MYC or L-MYC, (v) GLIS1 and (vi) NANOG; culturing the somatic cell to express the de-differentiation factor; selecting cells that display a stem cell morphology and/or stem cell markers; and subculturing the cells to obtain a population of induced stem cells. In one embodiment, the cells are selected by detecting expression of a Tumor Rejection Antigen 1-60 and/or 1-81.

[0012] The disclosure also provides a vector system for producing human stem cells, comprising at least one self-replicating RNA replicon comprising one or more polynucleotides encoding de-

differentiation factors selected from the group consisting of a KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, GLIS1, and NANOG. In one embodiment, the replicon comprises (a) Oct4, Sox2, Klf4, and c-Myc, or (b) Oct4, Sox2, Klf4, and Glis1. In another example, the at least one self-replicating RNA vector is derived from an alphavirus. In a further example, the alphavirus is VEE.

[0013] The disclosure also provides an isolated human somatic cell comprising an ectopic RNA replicon comprising one or more de-differentiation polynucleotide sequences. In a further example, wherein upon culture conditions to express the de-differentiation polynucleotides in the ectopic RNA replicon, the somatic cell de-differentiates.

[0014] The disclosure also provides a cell population comprising the human somatic cell containing an ectopic RNA replicon comprising one or more de-differentiation polynucleotide sequences.

[0015] The disclosure also provides a cell population obtained by contacting a human somatic cell with an ectopic self-replicating RNA replicon comprising polynucleotides encoding at least four de-differentiation factors selected from the group consisting of a (i) KLF4, (ii) OCT4, (iii) SOX2, (iv) c-MYC or n-MYC or L-MYC, (v) GLIS1 and (vi) NANOG; culturing the somatic cell to express the de-differentiation factor; selecting cells that display a stem cell morphology and/or stem cell markers; and subculturing the cells to obtain a population of induced stem cells. In one example, the cells are selected by detecting expression of a Tumor Rejection Antigen 1-60 and/or 1-81.

[0016] The disclosure also provides a recombinant human fibroblast cells containing an ectopic RNA molecule encoding B18R. In one example, the RNA encoding B18R comprise SEQ ID NO:39, wherein "T" is replaced with "U". In another example, the RNA encodes a polypeptide comprising the sequence set forth in SEQ ID NO:40.

[0017] The disclosure also provides a method of making B18R conditioned media comprising culturing a human fibroblast cell transformed with RNA encoding B18R under conditions that

allow expression of B18R and isolating media from the culture.

DESCRIPTION OF THE FIGURES

[0018] **Figure 1A-E** shows construction and Persistence of Synthetic VEE-RF RNA Replicons in Primary Human Fibroblasts. **(A)** Schematic of VEE-RF RNA replicon. 5' end nsP1-4: non-structural proteins1-4; 3' end C, E2, E1: Structural proteins. Locations of 26S internal promoter, ribosome shifting 2A peptide, IRES sequence, Puromycin (Puro) resistance gene and the regions for PCR detections of replicon as indicated. **(B)** Co-transfection of B18R mRNA with VEE RNA replicon enables to express VEE-GFP on day 1. **(C)** B18R Conditioned Media (B18R-CM) and puromycin selection are required for persistence of VEE-GFP RNA over 7 days. **(D)** B18R-CM and puromycin are required for retention of VEE-GFP RNA. Photographs of GFP expression on day 7 as indicated. Bar, 200 μ m. **(E)** Immunoblot analysis of VEE RNA expressed reprogramming factors expressed in HFFs cells on day 1 versus retrovirus (RV-4Fs) expression.

[0019] **Figure 2A-E** shows generation of iPS cells by VEE-RF RNA. **(A)** Schematic of epigenetic VEE-RF RNA iPS cell generation protocol. Human fibroblasts were plated on day 0 (d0) and co-transfected (Tfx) with VEE-RF RNA replicon plus B18R mRNA (3:1 ratio) on day 1 (confluent, $\sim 4 \times 10^5$ cells) and treated with puromycin until day 7 (or 10) as indicated. Cells were cultured in B18R-CM until iPS cell colonies were isolated on day 25-30. **(B)** iPS cell colonies stained with Alkaline Phosphatase were generated with VEE-OKS-iM RNA, but not VEE-OMKS RNA. **(C)** Alkaline Phosphatase staining of iPS cell colonies generated from BJ or HFFs from d1, 4, 7, 10 transfection protocol as indicated. **(D)** Typical images of iPS cell colonies on day 26 by VEE-OKS-iM RNA and day 22 for VEE-OKS-iG RNA from BJ or HFFs fibroblasts as indicated. Bar, 100 μ m. **(E)** Immunohistochemistry staining of pluripotent ES marker genes in isolated iPS cell clones generated as indicated. Similar results obtained for 26 additional iPS cell clones (30 clones total). Bar, 100 μ m; insert, 10x amplification.

[0020] **Figure 3A-E** shows characterization of VEE-RF RNA iPS Cell Clones. **(A)** Expression of ES maker genes by qRT-PCR analysis of BJ

and HFF VEE-OKS-iM iPS clones as indicated. **(B)** DNA methylation analysis of NANOG and OCT4 promoter regions. Solid circle, methylated; Open circle, demethylated. Top numbers indicate CpG number relative to the transcription start site. **(C)** Genome-wide mRNA sequence profile scatter plot analysis of BJ-OKS-iM #2 and BJ-OKS-iG #5 compared to parental human BJ fibroblasts and human HUES9 embryonic stem cells with pluripotency NANOG, OCT4, SOX2 indicated. **(D)** Unsupervised hierarchical dendrogram of genome-wide RNA sequences analysis showing clustering of four independent iPS cell clones with HUES9 compared to BJ fibroblasts. **(E)** Teratoma formation of BJ-OKS-iM #21 clone in nude mice. AE1/AE3 (cytokeratin), NF-1 (neuronal cells) and GFAP (neuronal cells) used for markers of ectoderm; Desmin (muscle cells) used for marker of mesoderm; and AFP (primitive and definitive endoderm) used for marker of endoderm. Bar, 100 μ m.

[0021] Figure 4 shows RT-PCR analysis for checking up the existence of RNA replicon. Measurement of PCR sensitivity with the plasmid of OKS-iM-RNA replicon. PCRs for nsP2, nsP4 and OCT4-T2A-KLF4 (OK) regions were performed with 100, 10 and 1 fg of plasmid (Top Panel). RT-PCR of HFF-OKS-iM iPSCs clones. +; positive control, total RNA was prepared from one day after transfection of OKS-iM-RNA replicon. -; negative control, total RNA was prepared from mock transfected HFFs. Total RNAs from iPS cell clones were prepared from passage 8 (Bottom Panel).

[0022] Figure 5 shows Karyotype Analysis of iPS Cell Clones. G-Banded Karyotyping of HFF-OKS-iM-1, BJ-OKS-iM-2, BJ-OKS-iM-21 and BJ-OKS-iG-5 clones was performed on twenty G-banded metaphase cells from each clone and judged as normal male human karyotype in all clones (Cell line GENETICS).

[0023] Figure 6A-B shows iPS cell clones were cultured with STO feeder cells. Cells were collected, and then intramuscularly or subcutaneously injected into the hind limb muscles or dorsal flank of nude mice. After 5 to 8 weeks of injection, tumors were dissected and fixed with 4% paraformaldehyde. (A) Teratoma analysis of HFF-OKS-iM #1 clone in nude mice. AE1/AE3 (cytokeratin) and NF-1 (neuronal cells) used for markers of ectoderm; Desmin (muscle cells) used for marker of mesoderm; and AFP (primitive and

definitive endoderm) used for marker of endoderm. Bar, 100 μ m.

(B) H&E staining of teratomas from BJ-OKS-iG clones 3 and 5. Bar, 100 μ m.

[0024] Figure 7A-D shows (A) B18R Conditioned Media is useful for persistent existence of VEE RNA replicon. Top; % of GFP positive cells, Bottom: mean value of GFP fluorescence in GFP positive population. (B) Photographs of cells. Bar, 200 μ m. (C) Protein expression of RFs on day 10 as indicated. (D) B18R-CM is required for generation of iPS cells in feeder culture. HFFs were co-transfected with OKS-iM RNA and B18R mRNA as indicated, and then cells were cultured in the presence of B18R-CM and puromycin. Cells were passaged to STO feeder cells on day 10 (d1, 3, 8 transfections) or day 11 (d1, 4, 7, 10 transfections), and cultured in the presence or absence of B18R-CM plus/minus puromycin.

DETAILED DESCRIPTION

[0025] As used herein and in the appended claims, the singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the agent" includes reference to one or more agents known to those skilled in the art, and so forth.

[0026] Also, the use of "or" means "and/or" unless stated otherwise. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0027] It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of."

[0028] Although methods and materials similar or equivalent to those described herein can be used in the practice of the disclosed methods and compositions, the exemplary methods, devices and materials are described herein.

[0029] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs.

Thus, as used throughout the instant application, the following terms shall have the following meanings.

[0030] While induced pluripotent stem cells (iPS cells) are virtually identical to ES cells at molecular and functional levels, there are critical hurdles to translation of their therapeutic potentials into medical applications. One of the issues is that because the current standard protocols for reprogramming and propagation of iPS cells include animal-derived materials that are unsuitable for potential clinical purposes, a fully defined method to generate and expand hiPS cells needs to be developed.

[0031] Induced pluripotent stem cells (iPS) are described by Shinya Yamanaka's team at Kyoto University, Japan. Yamanaka identified genes that are particularly active in embryonic stem cells, and used retroviruses to transfect mouse fibroblasts with a selection of those genes. Eventually, four key pluripotency genes essential for the production of pluripotent stem cells were isolated; Oct-3/4, SOX2, c-Myc, and Klf4. Cells were isolated by antibiotic selection for Fbx15⁺ cells. The same group published a study along with two other independent research groups from Harvard, MIT, and the University of California, Los Angeles, showing successful reprogramming of mouse fibroblasts into iPS and even producing a viable chimera.

[0032] The generation of human iPS cells by retroviral expression of four reprogramming factors (RFs; also referred to as de-differentiation factors) opened the potential for regenerative medicine therapies based on patient-specific, personalized stem cells. However, the insertional mutagenic potential of retroviruses combined with the potential for latent RF gene activation, especially c-MYC, all but eliminates integrative DNA-based approaches for use in regenerative medicine therapies. Other DNA-based iPS approaches using episomal vectors, adenovirus, integrated and excised *piggyBac* transposon or floxed lentivirus have been developed; however, these approaches either suffer from low efficiency of iPS cell generation or require genomic excision strategies that leaves behind an inserted DNA element tag. RNA-based iPS cell approaches using Sendai virus or mRNA transfection avoid potential integration problems associated with DNA-based

approaches and are inherently safer methods for clinical applications. Although Sendai virus offers a reasonably efficient iPS approach, problems associated with persistent Sendai virus replication in iPS cell clones requires a negative selection step followed by several recloning steps from the single cell level to isolate virus-free iPS cells, such processes result in excessive iPS cellular division and passage. One of the more promising non-DNA based approaches involves daily transfection of four individual RF mRNAs (plus GFP mRNA) over 16 days. Unfortunately, this approach remains problematic. For example, experiments to replace *KLF4* and *c-MYC* retroviruses with corresponding transfected mRNAs were performed and the results validated; however *OCT4* and *SOX2* retroviruses could not be replaced with transfected mRNAs. The problem appears to stem from both the rapid degradation of RF mRNAs combined with the inconsistent cell-to-cell threshold expression level variation over time, which derives from attempting to transfect four independent mRNAs into the same cell on a daily basis for >14 days during reprogramming. Consequently, there remains a significant need for a simple and highly reproducible, non-DNA based approach to generate human iPS cells.

[0033] The disclosure provides methods and compositions for generating iPS cells from somatic cells (e.g., fibroblast cells). The compositions and method comprise the use of replicons derived from alphaviruses. The replicons comprise an RNA sequence encoding non-structural alphavirus proteins necessary for replication and 1, 2, 3, 4 or more coding sequences heterologous to the alphavirus and which induce dedifferentiation of somatic cells to stem cell phenotypes.

[0034] As used herein, the term "alphavirus" has its conventional meaning in the art, and includes the various species such as Venezuelan Equine Encephalitis (VEE) Virus, Eastern Equine Encephalitis (EEE) virus, Everglades Virus (EVE), Mucambo Virus (MUC), Pixuna Virus (PIX), and Western Equine Encephalitis Virus, all of which are members of the VEE/EEE Group of alphaviruses. Other alphaviruses include, e.g., Semliki Forest Virus (SFV), Sindbis, Ross River Virus, Chikungunya Virus, S.A. AR86, Barmah Forest Virus, Middleburg Virus, O'nyong-nyong Virus, Getah Virus,

Sagiyama Virus, Bebaru Virus, Mayaro Virus, Una Virus, Aura Virus, Whataroa Virus, Banbanki Virus, Kyzylagach Virus, Highlands J Virus, Fort Morgan Virus, Ndumu Virus, and Buggy Creek Virus. Alphaviruses particularly useful in the constructs and methods described herein are VEE/EEE group alphaviruses.

[0035] The terms "alphavirus RNA replicon", "alphavirus replicon RNA", "alphavirus RNA vector replicon", and "vector replicon RNA" are used interchangeably to refer to an RNA molecule expressing nonstructural protein genes such that it can direct its own replication (amplification) and comprises, at a minimum, 5' and 3' alphavirus replication recognition sequences, coding sequences for alphavirus nonstructural proteins, and a polyadenylation tract. It may additionally contain one or more elements (e.g., IRES sequences, core or mini-promoters and the like) to direct the expression, meaning transcription and translation, of a heterologous RNA sequence. The alphavirus replicon of the disclosure can comprise, in one embodiment, 5' and 3' alphavirus replication recognition sequences, coding sequences for alphavirus nonstructural proteins, a polyadenylation tract and one or more of a coding sequence selected from the group consisting of SOX-2, c-Myc, OCT-3/4, Klf, Glis1 and Nanog.

[0036] The term "polynucleotide," "nucleic acid" or "recombinant nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate (particularly with reference to a replicon), ribonucleic acid (RNA).

[0037] The term "expression" with respect to a gene or polynucleotide refers to transcription of the gene or polynucleotide and, as appropriate, translation of an mRNA transcript to a protein or polypeptide. Thus, as will be clear from the context, expression of a protein or polypeptide results from transcription and/or translation of the open reading frame.

[0038] Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of codons differing in their nucleotide sequences can be used to encode a given amino acid. A particular polynucleotide or gene sequence encoding a polypeptide described herein are referenced merely to illustrate an embodiment of the disclosure, and the disclosure

includes polynucleotides of any sequence that encode a polypeptide comprising the same amino acid sequence of the polypeptides and proteins of the enzymes utilized in the methods of the disclosure. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The disclosure includes such polypeptides with alternate amino acid sequences, and the amino acid sequences encoded by the RNA or DNA sequences shown herein merely illustrate embodiments of the disclosure.

[0039] The disclosure provides polynucleotides in the form of recombinant DNA expression vectors, RNA replicons or plasmids, as described in more detail elsewhere herein, that encode one or more polypeptides.

[0040] A polynucleotide of the disclosure can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques and those procedures described in the Examples section below. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by sequence analysis. Furthermore, oligonucleotides corresponding to nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

[0041] In one embodiment, a replicon of the disclosure comprise a sequence that is 90%, 95%, 98%, 99% or 100% identical to SEQ ID NO:29, 30, 31, or 32 from about position 1 to about position 7561 (including wherein "T" of the sequence can be substituted with "U"), followed by one or more RFs selected from the group consisting of Oct-3/4, Sox-2, Klf4, c-Myc, Nanog, and Glis1. Where more than one RF is present, the coding sequences may be separated by an internal ribosome entry site (IRES) or a small (e.g., a core) promoter such as SP1. The order of the RFs is not critical to the disclosure; thus the order may be Klf4, Oct-3/4, Sox-2, c-Myc or can be Sox-2, Klf4, Oct-3/4, c-Myc, or Oct4, Klf4, Sox2, c-Myc or any variation of the order of the RFs. The replicon may further comprise a selectable marker (e.g., an antibiotic resistance

marker). In other embodiments, coding sequences of RFs may be separated by self-cleaving peptides such as T2A and/or E2A. In another embodiment, the replicon comprises from 5' to 3': (VEE RNA replicases)-(26S promoter)-(RF₁)-(self cleaving peptide)-(RF₂)-(self cleaving peptide)-(RF₃)-(IRES or core promoter)-(RF₄)-(IRES or optional promoter)-(optional selectable marker)-(VEE 3'UTR and polyA tail); wherein RF₁₋₄ are factors that induce de-differentiation of a somatic cell to a pluripotent cells, wherein RF₂₋₃ are optional, RF₃₋₄ are optional, or RF₄ is optional; wherein RF₁₋₄ are selected from the group consisting of Oct-4, Klf4, Sox-2, c-Myc, Nanog, and Glis1. In one embodiment, the replicon of the foregoing is an RNA molecule. In a further embodiment, the replicon is derived from VEE and includes a mutation to reduce pathogenicity. In one embodiment, the VEE is a TC-83 strain (vaccine strain)-based RNA replicon with one point mutation (nsP2P₇₇₃ to S mutation), which reduced the cytopathic effect of replicon.

[0042] In any of the foregoing embodiments, the RFs include variants and degenerate polynucleotide sequences. For example, an RF can comprise homologs and variants of an OCT-4 polypeptide, KLF4 polypeptide, SOX-2 polypeptide, c-MYC polypeptide, NANOG polypeptide or GLIS1. For example, an RF coding sequence for NANOG useful in any of the replicon embodiments described herein can comprise (i) a polynucleotide encoding a polypeptide of SEQ ID NO:2; (ii) a polynucleotide comprising at least 95% identity to SEQ ID NO:1 and which encodes a polypeptide having NANOG activity; (iii) a polynucleotide having a sequence as set forth in SEQ ID NO:1 or (iv) a polynucleotide encoding a polypeptide of SEQ ID NO:2 containing 1 to 10 conservative amino acid substitutions and wherein the polypeptide has Nanog activity; and wherein any of the foregoing nucleic acid sequences can have "T" replaced with "U". For example, an RF coding sequence for Oct-4 useful in any of the replicon embodiments described herein can comprise (i) a polynucleotide encoding a polypeptide of SEQ ID NO:4; (ii) a polynucleotide comprising at least 95% identity to SEQ ID NO:3 and which encodes a polypeptide having Oct-4 activity; (iii) a polynucleotide having a sequence as set forth in SEQ ID NO:3 or

(iv) a polynucleotide encoding a polypeptide of SEQ ID NO:4 containing 1 to 10 conservative amino acid substitutions and wherein the polypeptide has Oct-4 activity; and wherein any of the foregoing nucleic acid sequences can have "T" replaced with "U". For example, an RF coding sequence for Sox-2 useful in any of the replicon embodiments described herein can comprise (i) a polynucleotide encoding a polypeptide of SEQ ID NO:6; (ii) a polynucleotide comprising at least 95% identity to SEQ ID NO:5 and which encodes a polypeptide having SOX-2 activity; (iii) a polynucleotide having a sequence as set forth in SEQ ID NO:5 or (iv) a polynucleotide encoding a polypeptide of SEQ ID NO:6 containing 1 to 10 conservative amino acid substitutions and wherein the polypeptide has SOX-2 activity; and wherein any of the foregoing nucleic acid sequences can have "T" replaced with "U". For example, an RF coding sequence for KLF4 useful in any of the replicon embodiments described herein can comprise (i) a polynucleotide encoding a polypeptide of SEQ ID NO:8; (ii) a polynucleotide comprising at least 95% identity to SEQ ID NO:7 and which encodes a polypeptide having KLF4 activity; (iii) a polynucleotide having a sequence as set forth in SEQ ID NO:7 or (iv) a polynucleotide encoding a polypeptide of SEQ ID NO:8 containing 1 to 10 conservative amino acid substitutions and wherein the polypeptide has KLF4 activity; and wherein any of the foregoing nucleic acid sequences can have "T" replaced with "U". For example, an RF coding sequence for c-MYC useful in any of the replicon embodiments described herein can comprise (i) a polynucleotide encoding a polypeptide of SEQ ID NO:10; (ii) a polynucleotide comprising at least 95% identity to SEQ ID NO:9 and which encodes a polypeptide having c-MYC activity; (iii) a polynucleotide having a sequence as set forth in SEQ ID NO:9 or (iv) a polynucleotide encoding a polypeptide of SEQ ID NO:10 containing 1 to 10 conservative amino acid substitutions and wherein the polypeptide has c-MYC activity; and wherein any of the foregoing nucleic acid sequences can have "T" replaced with "U". For example, an RF coding sequence for GLIS1 useful in any of the replicon embodiments described herein can comprise (i) a polynucleotide encoding a polypeptide of SEQ ID NO:34; (ii) a

polynucleotide comprising at least 95% identity to SEQ ID NO:33 and which encodes a polypeptide having GLIS1 activity; (iii) a polynucleotide having a sequence as set forth in SEQ ID NO:33 or (iv) a polynucleotide encoding a polypeptide of SEQ ID NO:34 containing 1 to 10 conservative amino acid substitutions and wherein the polypeptide has GLIS1 activity; and wherein any of the foregoing nucleic acid sequences can have "T" replaced with "U".

[0043] Nanog is a gene expressed in embryonic stem cells (ESCs) and plays a role in maintaining pluripotency. Nanog is thought to function with SOX2. A polynucleotide and polypeptide encoding a Nanog is set forth in SEQ ID NO:1 and 2, respectively. Furthermore, SEQ ID NO:1 comprises a DNA sequence it will be recognized that "T" can be replaced with "U". Human NANOG protein (see, e.g., Accession number NP_079141, incorporated herein by reference) is a 305 amino acid protein with a homeodomain motif that is localized to the nuclear component of cells. Similar to murine NANOG, N-terminal region of human NANOG is rich in Ser, Thr and Pro residues and the C-terminus comprises Trp repeats. The homeodomain in human NANOG ranges from about residue 95 to about residue 155. Homologs of human Nanog are known.

[0044] An "Oct polypeptide" refers to any of the naturally-occurring members of Octamer family of transcription factors, or variants thereof that maintain transcription factor activity, similar (within at least 50%, 80%, or 90% activity) compared to the closest related naturally occurring family member, or polypeptides comprising at least the DNA-binding domain of the naturally occurring family member, and can further comprise a transcriptional activation domain. Exemplary Oct polypeptides include, Oct-1, Oct-2, Oct-3/4, Oct-6, Oct-7, Oct-8, Oct-9, and Oct-11. e.g. Oct3/4 (referred to herein as "Oct4") contains the POU domain, a 150 amino acid sequence conserved among Pit-1, Oct-1, Oct-2, and uric-86. See, Ryan, A. K. & Rosenfeld, M. G. *Genes Dev.* 11, 1207-1225 (1997). In some embodiments, variants have at least 85%, 90%, or 95% amino acid sequence identity across their whole sequence compared to a naturally occurring Oct polypeptide family member such as to those listed above or such as listed in Genbank accession number NP002692.2 (human Oct4) or NP038661.1 (mouse

Oct4). Oct polypeptides (e.g., Oct3/4) can be from human, mouse, rat, bovine, porcine, or other animals. Generally, the same species of protein will be used with the species of cells being manipulated. Oct-4 (Octamer-4) is a homeodomain transcription factor of the POU family and regulates the expression of numerous genes (see, e.g., J. Biol. Chem., Vol. 282, Issue 29, 21551-21560, July 20, 2007, incorporated herein by reference). A polynucleotide and polypeptide encoding an Oct4 is set forth in SEQ ID NO:3 and 4, respectively. Furthermore, SEQ ID NO:3 comprises a DNA sequence it will be recognized that "T" can be replaced with "U". Homologs of human Oct-4 are known as set forth in the following accession numbers NP_038661.1 and NM_013633.1 (*Mus musculus*), NP_001009178 and NM_001009178 (*Rattus norvegicus*), and NP_571187 and NM_131112 (*Danio rerio*), which are incorporated herein by reference.

[0045] SRY (sex determining region Y)-box 2, also known as SOX2, is a transcription factor that plays a role in self-renewal of undifferentiated embryonic stem cells and transactivation of Fgf4 as well as modulating DNA bending (see, e.g., Scaffidi et al. J. Biol. Chem., Vol. 276, Issue 50, 47296-47302, December 14, 2001, incorporated herein by reference). A "Sox polypeptide" refers to any of the naturally-occurring members of the SRY-related HMG-box (Sox) transcription factors, characterized by the presence of the high-mobility group (HMG) domain, or variants thereof that maintain transcription factor activity similar (within at least 50%, 80%, or 90% activity) compared to the closest related naturally occurring family member, or polypeptides comprising at least the DNA-binding domain of the naturally occurring family member, and can further comprise a transcriptional activation domain. See, e.g., Dang, D. T., et al., Int. J. Biochem. Cell Biol. 32:1103-1121 (2000). Exemplary Sox polypeptides include, e.g., Sox1, Sox-2, Sox3, Sox4, Sox5, Sox6, Sox7, Sox8, Sox9, Sox10, Sox11, Sox12, Sox13, Sox14, Sox15, Sox17, Sox18, Sox-21, and Sox30. Sox1 has been shown to yield iPS cells with a similar efficiency as Sox2, and genes Sox3, Sox15, and Sox18 have also been shown to generate iPS cells, although with somewhat less efficiency than Sox2. See, Nakagawa, et al., Nature Biotechnology 26:101-106 (2007). In some embodiments, variants have at least 85%, 90%, or 95% amino acid sequence

identity across their whole sequence compared to a naturally occurring Sox polypeptide family member such as to those listed above or such as listed in Genbank accession number CAA83435 (human Sox2). Sox polypeptides (e.g., Sox1, Sox2, Sox3, Sox15, or Sox18) can be from human, mouse, rat, bovine, porcine, or other animals. Generally, the same species of protein will be used with the species of cells being manipulated. A polynucleotide and polypeptide encoding a Sox2 is set forth in SEQ ID NO:5 and 6, respectively. Furthermore, SEQ ID NO:5 comprises a DNA sequence it will be recognized that "T" can be replaced with "U". Homologs of human Sox2 are known.

[0046] Kruppel-like factor 4, also known as KLF4 plays a role in stem cell maintenance and growth. A "Klf polypeptide" refers to any of the naturally-occurring members of the family of Kruppel-like factors (Klfs), zinc-finger proteins that contain amino acid sequences similar to those of the Drosophila embryonic pattern regulator Kruppel, or variants of the naturally-occurring members that maintain transcription factor activity similar (within at least 50%, 80%, or 90% activity) compared to the closest related naturally occurring family member, or polypeptides comprising at least the DNA-binding domain of the naturally occurring family member, and can further comprise a transcriptional activation domain. See, Dang, D. T., Pevsner, J. & Yang, V. W., Cell Biol. 32,1103-1121 (2000). Exemplary Klf family members include, Klf1, Klf2, Klf3, Klf-4, Klf5, Klf6, Klf7, Klf8, Klf9, Klf10, Klf11, Klf12, Klf13, Klf14, Klf15, Klf16, and Klf17. Klf2 and Klf-4 were found to be factors capable of generating iPS cells in mice, and related genes Klf1 and Klf5 did as well, although with reduced efficiency. See, Nakagawa, et al., Nature Biotechnology 26:101-106 (2007). In some embodiments, variants have at least 85%, 90%, or 95% amino acid sequence identity across their whole sequence compared to a naturally occurring Klf polypeptide family member such as to those listed above or such as listed in Genbank accession number CAX16088 (mouse Klf4) or CAX14962 (human Klf4). Klf polypeptides (e.g., Klf1, Klf4, and Klf5) can be from human, mouse, rat, bovine, porcine, or other animals. Generally, the same species of protein will be used with the species of cells being

manipulated. To the extent a Klf polypeptide is described herein, it can be replaced with an estrogen-related receptor beta (Essrb) polypeptide. Thus, it is intended that for each Klf polypeptide embodiment described herein, a corresponding embodiment using Essrb in the place of a Klf4 polypeptide is equally described. A polynucleotide and polypeptide encoding an KLF4 is set forth in SEQ ID NO:7 and 8, respectively. Furthermore, SEQ ID NO:7 comprises a DNA sequence it will be recognized that "T" can be replaced with "U". Homologs of human KLF4 are known and include NP_034767, NM_010637 (Mus musculus), which are incorporated herein by reference.

[0047] The MYC family of cellular genes is comprised of c-myc, N-myc, and L-myc, three genes that function in regulation of cellular proliferation, differentiation, and apoptosis (Henriksson and Luscher 1996; Facchini and Penn 1998). A "Myc polypeptide" refers any of the naturally-occurring members of the Myc family (see, e.g., Adhikary, S. & Eilers, M. Nat. Rev. Mol. Cell Biol. 6:635-645 (2005)), or variants thereof that maintain transcription factor activity similar (within at least 50%, 80%, or 90% activity) compared to the closest related naturally occurring family member, or polypeptides comprising at least the DNA-binding domain of the naturally occurring family member, and can further comprise a transcriptional activation domain. Exemplary Myc polypeptides include, e.g., c-Myc, N-Myc and L-Myc. In some embodiments, variants have at least 85%, 90%, or 95% amino acid sequence identity across their whole sequence compared to a naturally occurring Myc polypeptide family member, such as to those listed above or such as listed in Genbank accession number CAA25015 (human Myc). Myc polypeptides (e.g., c-Myc) can be from human, mouse, rat, bovine, porcine, or other animals. Generally, the same species of protein will be used with the species of cells being manipulated. Although myc family genes have common structural and biological activity. N-Myc is a member of the MYC family and encodes a protein with a basic helix-loop-helix (bHLH) domain. The genomic structures of c-myc and N-myc are similarly organized and are comprised of three exons. Most of the first exon and the 3' portion of the third exon contain untranslated regions that carry

transcriptional or post-transcriptional regulatory sequences. N-myc protein is found in the nucleus and dimerizes with another bHLH protein in order to bind DNA. A polynucleotide and polypeptide encoding an c-Myc is set forth in SEQ ID NO:9 and 10, respectively. Furthermore, SEQ ID NO:9 comprises a DNA sequence it will be recognized that "T" can be replaced with "U". Homologs and variants of the Myc family of proteins are known in the art.

[0048] Glis1 (Glis Family Zinc Finger 1) is gene encoding a Krüppel-like protein of the same name whose locus is found on Chromosome 1p32.3. The gene is enriched in unfertilised eggs and embryos at the one cell stage and it can be used to promote direct reprogramming of somatic cells to induced pluripotent stem cells. Glis1 can be used as one of the four factors used in reprogramming somatic cells to induced pluripotent stem cells. The three other transcription factors used are Oct3/4, Sox2 and Klf4. A human Glis1 (NM_147193) is set forth in SEQ ID NO:33 and 34 (cDNA and polypeptide, respectively).

[0049] cDNA coding for the human oct4 (pour5f1), sox2, klf4, c-myc (n-myc or L-myc), Glis1 and nanog, variants and homologs thereof can be cloned and expressed using techniques known in the art. Using the sequences set forth herein polynucleotides encoding one or more de-differentiation factors can be cloned into a suitable vector for expression in a cell type of interest.

[0050] An RF "activity" (e.g., an RF variant activity) refers the ability to de-differentiate a somatic cell when expressed in combination with other RFs as known in the art. For example, an Oct-4 variant can be measured for Oct-4 activity by co-expressing the Oct-4 variant in a somatic cell with klf4, Sox-2 and c-myc and determining if a somatic cell de-differentiates. If the cell de-differentiates than the Oct-4 variant can be said to have Oct-4 activity.

[0051] In another embodiment, the replicon comprises a sequence as set forth in SEQ ID NO:29, 30, 31, or 32. In yet another embodiment, the replicon comprises a sequence that is about 90%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% identical to SEQ ID NO:29, 30, 31, or 32, and wherein when the replicon is transfected into a somatic cells, the somatic cell is "induced" to become a

stem cell. In addition, any of SEQ ID NO:29, 30, 31, or 32, wherein "T" is replaced by "U".

[0052] In one embodiment, SEQ ID NO:29 provides a replicon of the disclosure. In another embodiment the sequence of SEQ ID NO:29 has "T" replaced with "U". The replicon comprises VEE RNA replicases from nucleotide 1 to about nucleotide 7561, a human Oct-4 sequence from nucleotide 7592 to 8671, a coding sequence for a T2A self-cleaving peptide from nucleotide 8678-8731, a human Klf4 sequence from 8738-10147, a coding sequence for a self-cleaving E2A peptide from nucleotide 10154-10213, a human Sox-2 sequence from 10223-11176, an internal ribosome entry site from 11195-11805, a human c-Myc sequence from 11818-13140, an internal ribosome entry site from 13165-13776, a puromycin resistance gene from 13777-14376, the VEE 3'UTR and polyA tail from 14383-14510, an ampicillin resistance gene from 14679-15539 and a SP6 promoter from 16320-16337.

[0053] In one embodiment, SEQ ID NO:30 provides a replicon of the disclosure. In another embodiment the sequence of SEQ ID NO:30 has "T" replaced with "U". The replicon comprises VEE RNA replicases from nucleotide 1 to about nucleotide 7561, a human Oct-4 sequence from nucleotide 7592 to 8671, a coding sequence for a T2A self-cleaving peptide from nucleotide 8678-8731, a human Klf4 sequence from 8738-10147, a coding sequence for a self-cleaving E2A peptide from nucleotide 10154-10213, a human Sox-2 sequence from 10223-11176, an internal ribosome entry site from 11195-11805, a human c-Myc sequence from 11818-13140, an internal ribosome entry site from 13165-13776, a puromycin resistance gene from 13777-14376, the VEE 3'UTR and polyA tail from 14383-14510, an ampicillin resistance gene from 14679-15539 and a T7 promoter from 16319-16336.

[0054] In one embodiment, SEQ ID NO:31 provides a replicon of the disclosure. In another embodiment the sequence of SEQ ID NO:31 has "T" replaced with "U". The replicon comprises VEE RNA replicases from nucleotide 1 to about nucleotide 7561, a human Oct-4 sequence from nucleotide 7592 to 8671, a coding sequence for a T2A self-cleaving peptide from nucleotide 8678-8731, a human Klf4 sequence from 8738-10147, a coding sequence for a self-cleaving E2A peptide from nucleotide 10154-10213, a human Sox-2 sequence from 10223-11176, an internal ribosome entry site from 11195-11805, a human

Glis1 sequence from 11818-13680, an internal ribosome entry site from 13689-14300, a puromycin resistance gene from 14301-14900, the VEE 3'UTR and polyA tail from 14907-15034, an ampicillin resistance gene from 15203-16063 and a SP6 promoter from 16844-16861.

[0055] In one embodiment, SEQ ID NO:32 provides a replicon of the disclosure. In another embodiment the sequence of SEQ ID NO:32 has "T" replaced with "U". The replicon comprises VEE RNA replicases from nucleotide 1 to about nucleotide 7561, a human Oct-4 sequence from nucleotide 7592 to 8671, a coding sequence for a T2A self-cleaving peptide from nucleotide 8678-8731, a human Klf4 sequence from 8738-10147, a coding sequence for a self-cleaving E2A peptide from nucleotide 10154-10213, a human Sox-2 sequence from 10223-11176, an internal ribosome entry site from 11195-11805, a human Glis1 sequence from 11818-13680, an internal ribosome entry site from 13689-14300, a puromycin resistance gene from 14301-14900, the VEE 3'UTR and polyA tail from 14907-15034, an ampicillin resistance gene from 15203-16063 and a T7 promoter from 16843-16860.

[0056] In another embodiment, more than one alphavirus replicon may be used, each replicon comprising one or more coding sequences for factors that induce a somatic cell to become a stem cell, wherein the combination of the more than one alphavirus replicons include all the coding sequence for all RFs necessary for inducing de-differentiation into a stem cell.

[0057] In more specific embodiments, an alphavirus replicon comprises coding sequences for expression of OCT-3/4, SOX-2, KLF, c-MYC, GLIS1 and/or NANOG. In a specific embodiment, the alphavirus replicon comprises coding sequences for OCT-4, KLF4, SOX-2, GLIS1 and c-MYC.

[0058] The replicon may also be engineered to express alphavirus structural proteins. U.S. Pat. Nos. 7,045,335, 7,078,218, 7,425,337 and 7,442,381 describe numerous constructs for such alphavirus RNA replicons consisting of the 5' and 3' alphavirus replication recognition sequences, coding sequences for alphavirus nonstructural proteins, and a polyadenylation tract, and such constructs are incorporated herein by reference. Specific embodiments of the alphavirus RNA replicons may contain one or more attenuating mutations, an attenuating mutation being a nucleotide

deletion, addition, or substitution of one or more nucleotide(s), or a mutation that comprises rearrangement or chimeric construction which results in a loss of virulence in a live virus containing the mutation as compared to the appropriate wild-type alphavirus.

[0059] The terms "alphavirus structural protein/protein(s)" refers to one or a combination of the structural proteins encoded by alphaviruses. These are produced by the virus as a polyprotein and are represented generally in the literature as C-E3-E2-6k-E1. E3 and 6k serve as membrane translocation/transport signals for the two glycoproteins, E2 and E1. Thus, use of the term E1 herein can refer to E1, E3-E1, 6k-E1, or E3-6k-E1, and use of the term E2 herein can refer to E2, E3-E2, 6k-E2, or E3-6k-E2. Attenuating mutations can be introduced into any one or more of the alphavirus structural proteins.

[0060] In addition, and as mentioned above, homologs of enzymes useful for generating metabolites are encompassed by the microorganisms and methods provided herein. The term "homologs" used with respect to an original enzyme or gene of a first family or species refers to distinct enzymes or genes of a second family or species which are determined by functional, structural or genomic analyses to be an enzyme or gene of the second family or species which corresponds to the original enzyme or gene of the first family or species. Most often, homologs will have functional, structural or genomic similarities. Techniques are known by which homologs of an enzyme or gene can readily be cloned using genetic probes and PCR. Identity of cloned sequences as homolog can be confirmed using functional assays and/or by genomic mapping of the genes.

[0061] A protein has "homology" or is "homologous" to a second protein if the nucleic acid sequence that encodes the protein has a similar sequence to the nucleic acid sequence that encodes the second protein. Alternatively, a protein has homology to a second protein if the two proteins have "similar" amino acid sequences. (Thus, the term "homologous proteins" is defined to mean that the two proteins have similar amino acid sequences).

[0062] As used herein, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences have at

least about 30%, 40%, 50% 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity. To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In one embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, typically at least 40%, more typically at least 50%, even more typically at least 60%, and even more typically at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0063] When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill

in the art (see, e.g., Pearson *et al.*, 1994, hereby incorporated herein by reference).

[0064] A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0065] Sequence homology for polypeptides, which can also be referred to as percent sequence identity, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wis. 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. See, e.g., GCG Version 6.1.

[0066] A typical algorithm used comparing a molecule sequence to a database containing a large number of sequences from different organisms is the computer program BLAST (Altschul, 1990; Gish,

1993; Madden, 1996; Altschul, 1997; Zhang, 1997), especially blastp or tblastn (Altschul, 1997). Typical parameters for BLASTp are: Expectation value: 10 (default); Filter: seg (default); Cost to open a gap: 11 (default); Cost to extend a gap: 1 (default); Max. alignments: 100 (default); Word size: 11 (default); No. of descriptions: 100 (default); Penalty Matrix: BLOWSUM62.

[0067] When searching a database containing sequences from a large number of different organisms, it is typical to compare amino acid sequences. Database searching using amino acid sequences can be measured by algorithms other than blastp known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, 1990, hereby incorporated herein by reference). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, hereby incorporated herein by reference.

[0068] As described herein, the compositions and methods of the disclosure provide the ability to de-differentiate somatic cells to form stem cells (e.g., induce the formation of stem cells). Stem cells are cells capable of differentiation into other cell types, including those having a particular, specialized function (e.g., tissue specific cells, parenchymal cells and progenitors thereof). There are various classes of stem cells, which can be characterized in their ability to differentiate into a desired cell/tissue type. For example, "progenitor cells" can be either multipotent or pluripotent. Progenitor cells are cells that can give rise to different terminally differentiated cell types, and cells that are capable of giving rise to various progenitor cells. The term "pluripotent" or "pluripotency" refers to cells with the ability to give rise to progeny cells that can undergo differentiation, under the appropriate conditions, into cell types that collectively demonstrate characteristics associated with cell lineages from all of the three germinal layers (endoderm, mesoderm, and ectoderm). Pluripotent stem cells can contribute to all embryonic derived

tissues of a prenatal, postnatal or adult animal. A standard art-accepted test, such as the ability to form a teratoma in 8-12 week old SCID mice, can be used to establish the pluripotency of a cell population; however identification of various pluripotent stem cell characteristics can also be used to detect pluripotent cells.

"Pluripotent stem cell characteristics" refer to characteristics of a cell that distinguish pluripotent stem cells from other cells.

The ability to give rise to progeny that can undergo differentiation, under the appropriate conditions, into cell types that collectively demonstrate characteristics associated with cell lineages from all of the three germinal layers (endoderm, mesoderm, and ectoderm) is a pluripotent stem cell characteristic. Expression or non-expression of certain combinations of molecular markers are also pluripotent stem cell characteristics. For example, human pluripotent stem cells express at least some, and in some embodiments, all of the markers from the following non-limiting list: SSEA-3, SSEA-4, TRA-1-60, TRA-1-81, TRA-2-49/6E, ALP, Sox2, E-cadherin, UTF-1, Oct4, Rex1, and Nanog. Cell morphologies associated with pluripotent stem cells are also pluripotent stem cell characteristics. In comparison, a multipotent stem cell is capable of differentiating into a subset of cells compared to a pluripotent stem cell. For example, a multipotent stem cell may be able to undergo differentiation into one or two of the three germinal layers. As used herein, "non-pluripotent cells" refer to mammalian cells that are not pluripotent cells. Examples of such cells include differentiated cells as well as multipotent cells. Examples of differentiated cells include, but are not limited to, cells from a tissue selected from bone marrow, skin, skeletal muscle, fat tissue and peripheral blood. Exemplary cell types include, but are not limited to, fibroblasts, hepatocytes, myoblasts, neurons, osteoblasts, osteoclasts, and T-cells.

[0069] Another class of cells even more primitive (*i.e.*, uncommitted to a particular differentiation fate) than pluripotent stem cells are the so-called "totipotent" stem cells (*e.g.*, fertilized oocytes, cells of embryos at the two and four cell stages of development), which have the ability to differentiate into any type of cell of the particular species. For example, a

single totipotent stem cell could give rise to a complete animal, as well as to any of the myriad of cell types found in the particular species (e.g., humans).

[0070] Pluripotent stem cells are a type of cells that undergo self-renewal while maintaining an ability to give rise to all three germ layer-derived tissues and germ cell lineages. Although pluripotent human embryonic stem (hES) cells derived from human blastocysts are promising sources for cell-based therapies to treat diseases and disorders such as Parkinson's disease, cardiac infarction, spinal cord injury, and diabetes mellitus, their clinical potentials has been hampered by their immunogenicity and ethical concerns.

[0071] The term "precursor cell," "progenitor cell," and "stem cell" are used interchangeably in the art and herein and refer either to a pluripotent, or lineage-uncommitted, progenitor cell, which is potentially capable of an unlimited number of mitotic divisions to either renew its line or to produce progeny cells which will differentiate into fibroblasts or a lineage-committed progenitor cell and its progeny, which is capable of self-renewal and is capable of differentiating into a parenchymal cell type. Unlike pluripotent stem cells, lineage-committed progenitor cells are generally considered to be incapable of giving rise to numerous cell types that phenotypically differ from each other. Instead, they give rise to one or possibly two lineage-committed cell types.

[0072] The disclosure demonstrates that terminally differentiated human cells (e.g., human dermal fibroblasts) can be induced to de-differentiate using an ectopic mRNA expression system (e.g., a replicon system). The disclosure contemplates the use of a variety of de-differentiation (also referred to as Reprogramming Factors (RFs)) coding sequence comprising, for example, a polynucleotide that encodes KLF4, OCT4, SOX2, c-MYC or n-MYC (L-Myc), GLIS1, NANOG or any combination thereof (e.g., KLF4, OCT4, SOX2, c-MYC or n-MYC (L-Myc) and optionally NANOG). De-differentiation may be achieved by contacting a cell, *in vivo* or *in vitro*, with one or more self-replicating RNA vectors that remain ectopic to the host cell genome and encode factors that induce de-differentiation. In various embodiments the ectopic self-replicating RNA vector of the

disclosure can be controlled by culturing a host cell transformed with the self-replicating RNA vector in the presence of B18R. Methods for promoting de-differentiation provide methods of promoting regeneration of mammalian cells and tissues damaged by injury or disease. The disclosure also provides methods for enriching for induced stem cells and populations comprising such enriched stem cells.

[0073] The generation of patient-specific pluripotent stem cells has the potential to dramatically speed the implementation of stem cells into clinical use to treat degenerative diseases. The disclosure provides methods to employ easily donated stromal cells, such as dermal fibroblasts, from a patient and generate Human Induced Pluripotent Stem (hiPS or iPS) cells by ectopic expression of a set of de-differentiation factors comprising RNA encoding (i) KLF4, OCT4, SOX2, c-MYC or n-MYC(L-Myc), NANOG or any combination thereof; (ii) KLF4, OCT4, SOX2, and GLIS1; and (iii) KLF4, OCT4, SOX2, and NANOG. The cell lines generated are physiologically and morphologically indistinguishable from Human Embryonic Stem Cells (HESC) generated from the inner cell mass of a human embryo. hiPS cells share a nearly identical gene expression profile with two established HESC lines.

[0074] The term "de-differentiation" is familiar to the person skilled in the relevant art. In general de-differentiation signifies the regression of lineage committed cell to the status of a stem cell, for example, by "inducing" a de-differentiated phenotype. For example, as described further herein KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, GLIS1 and/or Nanog can induce de-differentiation and induction of mitosis in lineage committed mitotically inhibited cells.

[0075] In one embodiment, the disclosure provides a cell culture comprising human somatic cells that have been transformed with a replicon of the disclosure. In one embodiment the somatic cells are fibroblasts. In another embodiment, the somatic cells are keratinocytes. In another embodiment, the replicon comprises a sequence that is 90%, 95%, 98%, 99% or 100% identical to SEQ ID NO:29, 30, 31, or 32 from about position 1 to about position 7561 (including wherein "T" of the sequence can be substituted with

"U"), followed by one or more RFs selected from the group consisting of Oct-3/4, Sox-2, Klf4, c-Myc, Nanog, and Glis1 followed by a VEE 3'UTR and polyA tail. Where when more than one RF is present, the coding sequences may be separated by an internal ribosome entry site (IRES) or a small (e.g., a core) promoter such as SP1. The order of the RFs is not critical to the disclosure; thus the order may be Klf4, Oct-3/4, Sox-2, c-Myc or can be Sox-2, Klf4, Oct-3/4, c-Myc, or Oct4, Klf4, Sox2, c-Myc or any variation of the order of the RFs. In one embodiment, the replicon comprises a sequence that is at least about 95%, 98%, 99% or 100% identical to a sequence as set forth in SEQ ID NO:29, 30, 31, or 32. In yet another embodiment, the cells are cultured in conditioned media comprising B18R and/or are co-transformed with a polynucleotide encoding B18R.

[0076] The disclosure also provide methods of making a stem cell from a somatic cell comprising transforming the somatic cell with an RNA replicon as described in the disclosure and culturing the somatic cell under conditions to promote expression of coding sequences in the replicon and culturing the cells for a sufficient period of time to de-differentiate the cells to stem cells. In one embodiment, the cells are passaged at least 5, 10, 15, 20 or more times. In another embodiment, the cells are cultured for at least 10, 20, 30 or more days. In yet another embodiment, the cells are cultured in conditioned media comprising B18R or are co-transformed with a polynucleotide encoding B18R.

[0077] The disclosure also provides induced stem cell cultures obtained by the methods described herein. In one embodiment, the stem cells do not contain any heterologous RF factors in the genomic DNA of the cell. In another embodiment, the stem cells do not contain any retroviral DNA or RNA (e.g., stem cells that are retroviral DNA- or RNA-free).

[0078] In one embodiment, the disclosure provides isolated induced stem cells, individually or in populations. The term "isolated" or "purified" when referring to stem cells of the disclosure means cells that are substantially free of cells carrying markers associated with lineage dedication. In particular embodiments, the human induced pluripotent stem cells are at least 30%, 35%, 40%,

45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 99% free of such contaminating cell types. In another embodiment, the isolated stem cells also are substantially free of soluble, naturally occurring molecules. As discussed more fully below, a substantially purified stem cell of the disclosure can be obtained, for example, by extraction (e.g., via density gradient centrifugation and/or flow cytometry) from a culture source. Purity can be measured by any appropriate method. A stem cell of the disclosure can be 99%-100% purified by, for example, flow cytometry (e.g., FACS analysis), as discussed herein. Such purified iPS cells will lack any retroviral DNA or RNA.

[0079] In one embodiment, the disclosure provides an enriched population of induced stem cells. An "enriched population of induced stem cells" is one wherein induced stem cells of the disclosure have been partially separated from other cell types, such that the resulting population of cells has a greater concentration of induced stem cells than the original population of cells. The enriched population of induced stem cells can have greater than about a 10-fold, 100-fold, 500-fold, 1,000-fold, 2,000-fold, 3,000-fold, 4,000-fold, 5,000-fold, 6,000-fold, 7,000-fold, 8,000-fold, 9,000-fold, 10,000-fold or greater concentration of induced stem cells than the original population had prior to separation. Induced stem cells of the disclosure can, for example, make up at least 5%, 10%, 15%, 20%, 35%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or more of the enriched population of stem cells. The enriched population of induced stem cells may be obtained by, for example, selecting against cells displaying markers associated with differentiated cells, or other undesired cell types, and/or selecting for cells displaying markers (e.g., TRA-1-81 and/or TRA-1-60) associated with the human induced pluripotent stem cells of the disclosure, and/or by regenerating isolated stem cells in defined culture systems. Alternatively, or in addition to, the enrichment for the expression of a marker, the loss of expression of a marker may also be used for enrichment. Such enriched iPS cells will lack any retroviral RNA or DNA typically used to transform cells with RFs.

[0080] In another embodiment, the disclosure provides cell lines of induced stem cells. As used herein a "cell line" means a culture of stem cells of the disclosure, or progeny cells thereof, that can be reproduced for an extended period of time, preferably indefinitely, and which term includes, for example, cells that are cultured, cryopreserved and re-cultured following cryopreservation. As used herein a "culture" means a population of induced stem cells grown in a medium and optionally passaged accordingly. A stem cell culture may be a primary culture (e.g., a culture that has not been passaged) or may be a secondary or subsequent culture (e.g., a population of cells which have been subcultured or passaged one or more times).

[0081] In one embodiment, the disclosure provides cells that are de-differentiated to stem cells (i.e., induced stem cells) comprising characteristics including the ability of self-renewal and differentiation into mesoderme, endoderm and epiderm, wherein the de-differentiated cells can be produced by expression of one or more RFs ectopic to the host cell genome using a replicating RNA vector. In one embodiment, the replicon vector is derived from an alphavirus (e.g., Venezuelan Equine Encephalitis virus).

[0082] Therapeutic uses of the human induced pluripotent stem cells of the disclosure include transplanting the human induced pluripotent stem cells, stem cell populations, or progeny thereof into individuals to treat a variety of pathological states including diseases and disorders resulting from cancers, neoplasms, injury, viral infections, diabetes and the like. Stem cells or stem cell populations (including genetically altered stem cells) are introduced into a subject in need of such stem cells or progeny or in need of a KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, GLIS1 or any combination thereof protein or molecule encoded or produced by the genetically altered cell. For example, in one embodiment, the human induced pluripotent stem cells can be administered to cancer patients who have undergone chemotherapy that have killed, reduced, or damaged stem cells or other cells of a subject, wherein the induced stems cells replace the damaged or dead cells. In another embodiment, the human induced pluripotent stem cells can be transfected or transformed (in addition to the

de-differentiation factors) with at least one additional therapeutic factor. For example, once human induced pluripotent stem cells of the disclosure are isolated or obtained by the methods of the disclosure, the stem cells may be transformed with a polynucleotide encoding a therapeutic polypeptide. Such a method and compositions can provide stem cell bioreactors for the production of a desired polypeptide or may be used for gene delivery or gene therapy. In this embodiment, the iPS cells may be isolated, transformed with a polynucleotide encoding a therapeutic polypeptide and may then be implanted or administered to a subject, or may be differentiated to a desired cell type and implanted and delivered to the subject. Under such conditions the polynucleotide is expressed within the subject for delivery of the polypeptide product.

[0083] If the human cells are derived from a heterologous (non-autologous/allogenic) source compared to the recipient subject, concomitant immunosuppression therapy is typically administered, e.g., administration of the immunosuppressive agent cyclosporine or FK506. However, due to the immature state of the human induced pluripotent stem cells of the disclosure such immunosuppressive therapy may not be required. Accordingly, in one embodiment, the human induced pluripotent stem cells of the disclosure can be administered to a recipient in the absence of immunomodulatory (e.g., immunosuppressive) therapy. Alternatively, the cells can be encapsulated in a membrane, which permits exchange of fluids but prevents cell/cell contact. Transplantation of microencapsulated cells is known in the art, e.g., Balladur *et al.*, 1995, Surgery 117:189-94, 1995; and Dixit *et al.*, 1992, Cell Transplantation 1:275-79.

[0084] The cells may be introduced directly into the peripheral blood or deposited within other locations throughout the body, e.g., a desired tissue, or on microcarrier beads in the peritoneum. For example, 10^2 to 10^9 cells can be transplanted in a single procedure, and additional transplants can be performed as required.

[0085] Differentiation of the human induced pluripotent stem cells or de-differentiation of lineage committed (mitotically inhibited) cells can be induced ex vivo, or alternatively may be induced by

contact with tissue *in vivo*, (e.g., by contact with fibroblasts or cell matrix components). Optionally, a differentiating agent or de-differentiation agent (e.g., KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, GLIS1, or any combination thereof or an agonist thereof) may be co-administered or subsequently administered to the subject.

[0086] It has been previously demonstrated that transplantation of beta islet cells provides therapy for patients with diabetes (Shapiro *et al.*, 2000). The human induced pluripotent stem cells of the disclosure provide an alternative source of islet cells to prevent or treat diabetes. For example, induced pluripotent stem cells of the disclosure can be generated, isolated and differentiated to a pancreatic cell type and delivered to a subject. Alternatively, the induced pluripotent stem cells can be delivered to the pancreas of the subject and differentiated to islet cells *in vivo*. Accordingly, the cells are useful for transplantation in order to prevent or treat the occurrence of diabetes.

[0087] The disclosure contemplates that the *in vitro* methods described herein can be used for autologous transplantation of de-differentiated or redifferentiated cells (e.g., the cells are harvested from and returned to the same individual). The disclosure further contemplates that the *in vitro* methods described herein can be used for non-autologous transplantations. In one embodiment, the transplantation occurs between a genetically related donor and recipient. In another embodiment, the transplantation occurs between a genetically un-related donor and recipient. In any of the foregoing embodiments, the disclosure contemplates that de-differentiated cells can be expanded in culture and stored for later retrieval and use. Similarly, the disclosure contemplates that redifferentiated cells can be expanded in culture and stored for later retrieval and use.

[0088] The compositions and methods of the disclosure may be applied to a procedure wherein differentiated (lineage committed) cells are removed from the a subject, de-differentiated in culture, and then either reintroduced into that individual or, while still in culture, manipulated to redifferentiate along specific

differentiation pathways (e.g., pancreatic cells, neuronal cells, liver cells, skin cells, cardiovascular cells, gastrointestinal cells and the like). Such redifferentiated cells can then be introduced to the individual. For example, differentiated fibroblasts can be removed, de-differentiated (e.g., with ectopic expression of of a replicon of the disclosure comprising KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, GLIS1, NANOG or any combination thereof) and mitotically expanded and then re-differentiated (e.g., with a KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, GLIS1 antagonists or any combination thereof) or factors (including physical stimuli) known to cause differentiation of hESCs down a lineage committed path. In one embodiment, the method comprises removing differentiated cells from an injured or diseased subject. Cells de-differentiated from cells harvested from an injured subject can later be returned to the injured or diseased subject to treat an injury or degenerative disease. The de-differentiated cells can be reintroduced at the site or injury, or the cells can be reintroduced at a site distant from the injury. Similarly, cells can be harvested from an injured subject, de-differentiated *in vitro*, redifferentiated *in vitro*, and transplanted back to the subject to treat an injury or degenerative disease.

[0089] The human induced pluripotent stem cells of the disclosure can be isolated from a sample obtained from a mammalian subject. The subject can be any mammal (e.g., bovine, ovine, porcine, canine, feline, equine, primate), including a human. The sample of cells may be obtained from any of a number of different sources including, for example, bone marrow, fetal tissue (e.g., fetal liver tissue), peripheral blood, umbilical cord blood, pancreas and the like.

[0090] In another embodiment, the disclosure provides methods of establishing and/or maintaining populations of stem cells, or the progeny thereof, as well as mixed populations comprising both stem cells and progeny cells, and the populations of cells so produced. As with the human induced pluripotent stem cells of the disclosure, once a culture of cells or a mixed culture of stem cells is established, the population of cells is mitotically expanded *in*

vitro by passage to fresh medium as cell density dictates under conditions conducive to cell proliferation, with or without tissue formation. Such culturing methods can include, for example, passaging the cells in culture medium lacking particular growth factors that induce differentiation (e.g., IGF, EGF, FGF, VEGF, and/or other growth factor), in the presence of an agent that stimulates (e.g., an agonist) of KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, GLIS1 or any combination thereof, in the presence of KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, Glis1 or any combination thereof, or any combination of the foregoing. Cultures comprising fibroblast or fibroblast-like cells and mixed cultures comprising stem cells and fibroblast cells can be transferred to fresh medium when sufficient cell density is reached. Some stem cell types do not demonstrate typical contact inhibition-apoptosis or they become quiescent when density is maximum. Accordingly, appropriate passaging techniques can be used to reduce contact inhibition and quiescence. Thus, in one embodiment, for example, transferring a portion of the cells to a new culture vessel with fresh medium. Such removal or transfer can be done in any culture vessel.

[0091] Once the human induced pluripotent stem cells of the disclosure have been established in culture, as described above, they may be maintained or stored in cell "banks" comprising either continuous *in vitro* cultures of cells requiring regular transfer or cells which have been cryopreserved.

[0092] Cryopreservation of stem cells, or other cell of the disclosure, may be carried out according to known methods, such as those described in Doyle *et al.*, (eds.), 1995, Cell & Tissue Culture: Laboratory Procedures, John Wiley & Sons, Chichester. For example, but not by way of limitation, cells may be suspended in a "freeze medium" such as, for example, culture medium further comprising 15-20% fetal bovine serum (FBS) and 10% dimethylsulfoxide (DMSO), with or without 5-10% glycerol, at a density, for example, of about $4-10 \times 10^6$ cells/ml. The cells are dispensed into glass or plastic vials which are then sealed and transferred to a freezing chamber of a programmable or passive freezer. The optimal rate of freezing may be determined

empirically. For example, a freezing program that gives a change in temperature of $-1^{\circ}\text{C}/\text{min}$ through the heat of fusion may be used. Once vials containing the cells have reached -80°C , they are transferred to a liquid nitrogen storage area. Cryopreserved cells can be stored for a period of years, though they should be checked at least every 5 years for maintenance of viability.

[0093] The cryopreserved cells of the disclosure constitute a bank of cells, portions of which can be withdrawn by thawing and then used to produce a stem cell culture comprising stem cells, as needed. Thawing should generally be carried out rapidly, for example, by transferring a vial from liquid nitrogen to a 37°C water bath. The thawed contents of the vial should be immediately transferred under sterile conditions to a culture vessel containing an appropriate medium. It is advisable that the cells in the culture medium be adjusted to an initial density of about $1\text{--}3 \times 10^5$ cells/ml. Once in culture, the cells may be examined daily, for example, with an inverted microscope to detect cell proliferation, and subcultured as soon as they reach an appropriate density.

[0094] The human induced pluripotent stem cells of the disclosure may be withdrawn from a cell bank as needed, and used for the production of new stem cells, either *in vitro*, for example, as a three dimensional tissue culture, as described below, or *in vivo*, for example, by direct administration of cells to the site where new fibroblasts or tissue is needed. As described herein, the human induced pluripotent stem cells of the disclosure may be used to produce new tissue for use in a subject where the cells were originally isolated from that subject's own blood or other tissue (*i.e.*, autologous cells). Alternatively, the cells of the disclosure may be used as ubiquitous donor cells to produce new tissue for use in any subject (*i.e.*, heterologous cells).

[0095] Once established, a culture of stem cells may be used to produce progeny cells and/or fibroblasts capable of producing new tissue. Differentiation of stem cells to fibroblasts or other cell types, followed by the production of tissue therefrom, can be triggered by specific exogenous growth factors or by changing the culture conditions (*e.g.*, the density) of a stem cell culture. Since the cells are pluripotent, they can be used to reconstitute

an irradiated subject and/or a subject treated with chemotherapy; or as a source of cells for specific lineages, by providing for their maturation, proliferation and differentiation into one or more selected lineages. Examples of factors that can be used to induce differentiation include erythropoietin, colony stimulating factors, *e.g.*, GM-CSF, G-CSF, or M-CSF, interleukins, *e.g.*, IL-1, -2, -3, -4, -5, -6, -7, -8, and the like, Leukemia Inhibitory Factor (LIF), Steel Factor (Stl), or the like, coculture with tissue committed cells, or other lineage committed cells types to induce the stem cells into becoming committed to a particular lineage.

[0096] In another embodiment, the human induced pluripotent stem cells are genetically engineered to express genes for specific types of growth factors for successful and/or improved differentiation to fibroblasts, other stromal cells, or parenchymal cells and/or turnover either pre- or post-implantation.

[0097] The cells of the disclosure may be used to treat subjects requiring the repair or replacement of tissue resulting from disease or trauma. Treatment may entail the use of the cells of the disclosure to produce new tissue, and the use of the tissue thus produced, according to any method presently known in the art or to be developed in the future. For example, the induced cells (*e.g.*, cells comprising an ectopic expression vector expressing KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, Glis1 or any combination thereof) of the disclosure may be implanted, injected or otherwise administered directly to the site of tissue damage so that they will produce new tissue *in vivo*. In one embodiment, administration includes the administration of genetically modified stem cells.

[0098] In one embodiment, a formulation comprising the cells of the disclosure is prepared for injection directly to the site where the production of new tissue is desired. For example, and not by way of limitation, the cells of the disclosure may be suspended in a hydrogel solution for injection. Alternatively, the hydrogel solution containing the cells may be allowed to harden, for instance in a mold to form a matrix having cells dispersed therein prior to implantation. Once the matrix has hardened, the cell

formations may be cultured so that the cells are mitotically expanded prior to implantation. A hydrogel is an organic polymer (natural or synthetic) which is cross-linked via covalent, ionic, or hydrogen bonds to create a three-dimensional open-lattice structure, which entraps water molecules to form a gel. Examples of materials which can be used to form a hydrogel include polysaccharides such as alginate and salts thereof, polyphosphazines, and polyacrylates, which are cross-linked ionically, polyethylene oxide-polypropylene glycol block copolymers which are cross-linked by temperature or pH, respectively. Methods of synthesis of the hydrogel materials, as well as methods for preparing such hydrogels, are known in the art.

[0099] Such cell formulations may further comprise one or more other components, including selected extracellular matrix components, such as one or more types of collagen known in the art, and/or growth factors and drugs. Growth factors which may be usefully incorporated into the cell formulation include one or more tissue growth factors known in the art such as, but not limited to, any member of the TGF- β family, IGF-I and -II, growth hormone, BMPs such as BMP-13, and the like. Alternatively, the cells of the disclosure may be genetically engineered to express and produce growth factors such as BMP-13 or TGF- β . Other components may also be included in the formulation include, for example, buffers to provide appropriate pH and isotonicity, lubricants, viscous materials to retain the cells at or near the site of administration, (e.g., alginates, agars and plant gums) and other cell types that may produce a desired effect at the site of administration (e.g., enhancement or modification of the formation of tissue or its physicochemical characteristics, support for the viability of the cells, or inhibition of inflammation or rejection). The cells can be covered by an appropriate wound covering to prevent cells from leaving the site. Such wound coverings are known to those of skill in the art.

[00100] Alternatively, the human induced pluripotent stem cells of the disclosure may be seeded onto a three-dimensional framework or scaffold and cultured to allow the cells to differentiate, grow and fill the matrix or immediately implanted *in vivo*, where the seeded

cells will proliferate on the surface of the framework and form a replacement tissue *in vivo* in cooperation with the cells of the subject. Such a framework can be implanted in combination with any one or more growth factors, drugs, additional cell types, or other components that stimulate formation or otherwise enhance or improve the practice of the disclosure.

[00101] In yet another embodiment, the human induced pluripotent stem cells of the disclosure can be used in conjunction with a three-dimensional culture system in a "bioreactor" to produce tissue constructs which possess critical biochemical, physical and structural properties of native human tissue by culturing the cells and resulting tissue under environmental conditions which are typically experienced by native tissue. The bioreactor may include a number of designs. Typically the culture conditions will include placing a physiological stress on the construct containing cells similar to what will be encountered *in vivo*.

[00102] The human induced pluripotent stem cells, their progeny, and tissue of the disclosure can be used in a variety of applications. These include, but are not limited to, transplantation or implantation of the cells either in a differentiated form, an undifferentiated form, a de-differentiated form. Such cells and tissues serve to repair, replace or augment tissue that has been damaged due to disease or trauma, or that failed to develop normally.

[00103] The human induced pluripotent stem cells and tissue produced according to the disclosure can be used to repair or replace damaged or destroyed tissue or to augment existing tissue.

[00104] In addition, the cells or tissue of the disclosure can be used, for example, to screen *in vitro* for the efficacy and/or cytotoxicity of compounds, allergens, growth/regulatory factors, pharmaceutical compounds, and the like on stem cells, to elucidate the mechanism of certain diseases by determining changes in the biological activity of the stem cells (e.g., changes in KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, Glis1 or any combination thereof expression or activity, proliferative capacity, adhesion), to study the mechanism by which drugs and/or growth factors operate to modulate stem cell biological activity (e.g., KLF4, OCT4, SOX2,

c-MYC or n-MYC or L-MYC, NANOG, Glis1 or any combination thereof expression or activity), to diagnose and monitor cancer in a patient, for gene therapy, gene delivery or protein delivery; and to produce biologically active products.

[00105] The human induced pluripotent stem cells also can be used in the isolation and evaluation of factors associated with the differentiation and maturation of stem cells. Thus, the human induced pluripotent stem cells may be used in assays to determine the activity of media, such as conditioned media, evaluate fluids for cell growth activity, involvement with dedication of particular lineages, or the like. Various systems are applicable and can be designed to induced differentiation of the human induced pluripotent stem cells based upon various physiological stresses.

[00106] The human induced pluripotent stem cells, progeny thereof, and tissues derived therefrom of the disclosure may be used *in vitro* to screen a wide variety of agents for effectiveness and cytotoxicity of pharmaceutical agents, growth/regulatory factors, anti-inflammatory agents, and the like. To this end, the cells or tissue cultures of the disclosure can be maintained *in vitro* and exposed to the agent to be tested. The activity of a cytotoxic agent can be measured by its ability to damage or kill stem cells or their progeny in culture. This can be assessed readily by staining techniques. The effect of growth/regulatory factors can be assessed by analyzing the number of living cells *in vitro*, e.g., by total cell counts, and differential cell counts. This can be accomplished using standard cytological and/or histological techniques, including the use of immunocytochemical techniques employing antibodies that define type-specific cellular antigens. The effect of various drugs on the cells of the disclosure can be assessed either in a suspension culture or in a three-dimensional system. In one aspect, the effect of a test agent on the human induced pluripotent stem cells of the disclosure can be analyzed.

[00107] Stem cells which express a gene product of interest, or tissue produced *in vitro* therefrom, can be implanted into a subject who is otherwise deficient in that gene product. For example, genes that express products capable of preventing or ameliorating symptoms of various types of vascular diseases or disorders, or

that prevent or promote inflammatory disorders are of particular interest. In one embodiment, the cells of the disclosure are genetically engineered to express an anti-inflammatory gene product that would serve to reduce the risk of failure of implantation or further degenerative change in tissue due to inflammatory reaction. For example, a stem cell of the disclosure can be genetically engineered to express one or more anti-inflammatory gene products including, for example, peptides or polypeptides corresponding to the idiotype of antibodies that neutralize granulocyte-macrophage colony stimulating factor (GM-CSF), TNF, IL-1, IL-2, or other inflammatory cytokines. IL-1 has been shown to decrease the synthesis of proteoglycans and collagens type II, IX, and XI (Tyler *et al.*, 1985, *Biochem. J.* 227:69-878; Tyler *et al.*, 1988, *Coll. Relat. Res.* 82:393-405; Goldring *et al.*, 1988, *J. Clin. Invest.* 82:2026-2037; and Lefebvre *et al.*, 1990, *Biophys. Acta.* 1052:366-72). TNF also inhibits synthesis of proteoglycans and type II collagen, although it is much less potent than IL-1 (Yaron, I., *et al.*, 1989, *Arthritis Rheum.* 32:173-80; Ikebe, T., *et al.*, 1988, *J. Immunol.* 140:827-31; and Saklatvala, J., 1986, *Nature* 322:547-49). Also, for example, the cells of the disclosure may be engineered to express the gene encoding the human complement regulatory protein that prevents rejection of a graft by the host. See, for example, McCurry *et al.*, 1995, *Nature Medicine* 1:423-27. In another embodiment, the human induced pluripotent stem cells may be engineered to include a gene or polynucleotides sequence that expresses or causes to be expressed an angiogenic factor.

[00108] The induced stem cells of the disclosure express one or more markers associated with a human pluripotent stem cell phenotype and/or lack one or more markers associated with a differentiated cell (*e.g.*, a cell having a reduced capacity for self-renewal, regeneration, or differentiation) and/or a cell of neuronal origin. A molecule is a "marker" of a desired cell type if it is found on a sufficiently high percentage of cells of the desired cell type, and found on a sufficiently low percentage of cells of an undesired cell type. One can achieve a desired level of purification of the desired cell type from a population of cells comprising both desired and undesired cell types by selecting for

cells in the population of cells that have the marker. A marker can be displayed on, for example, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or more of the desired cell type, and can be displayed on fewer than 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5%, 1% or fewer of an undesired cell type.

[00109] As discussed above, the induced stem cells of the disclosure or induced stem cells that have been differentiated are characterized by the presence and/or the absence of certain markers that are specifically recognized by a molecule. Accordingly, in one aspect, the disclosure provides methods of labeling induced stem cells of the disclosure. In one embodiment, the human induced pluripotent stem cells are labeled with a molecule (e.g., an antibody) that specifically recognizes a marker that is associated with an induced stem cell of the disclosure. In another embodiment, a population of cells is contacted with a molecule that specifically binds to a marker (e.g., TRA-1-81) under conditions that allow the molecule to bind to the marker, wherein the population of cells comprises at least one stem cell having said marker. In another embodiment, a population of cells is contacted with a molecule that specifically binds to a marker under conditions that allow the molecule to bind to the marker, wherein the population of cells comprises stem cells that do not have the marker and non-stem cells that do have the marker. The molecule used can be, for example, an antibody, an antibody derivative, or a ligand. The molecule optionally can comprise an additional moiety, for example, one that is detectable (e.g., a fluorescent or colorimetric label) or one that aids in the isolation of the labeled cells (e.g., a moiety that is bound by another molecule or a magnetic particle).

[00110] In one embodiment, the population of transformed somatic cells undergoes live staining for a Tumor Rejection Antigen 1-61 and 1-81 (TRA-1-60, TRA-1-81). TRA-1-60 and TRA-1-81 may be obtained commercially, for example from Chemicon International, Inc (Temecula, Calif., USA). The immunological detection of these antigens using monoclonal antibodies has been used to characterize pluripotent stem cells in combination with other markers (Shamblott M. J. *et al.* (1998) PNAS 95: 13726-13731; Schuldiner M. *et al.*

(2000). PNAS 97: 11307-11312; Thomson J. A. *et al.* (1998). Science 282: 1145-1147; Reubinoff B. E. *et al.* (2000). Nature Biotechnology 18: 399-404; Henderson J. K. *et al.* (2002). Stem Cells 20: 329-337; Pera M. *et al.* (2000). J. Cell Science 113: 5-10.). In one embodiment, a population of somatic cells that have been transformed with at least one ectopic RNA vector comprising a KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, and optionally or alternatively NANOG or Glis1 are enriched for cells comprising TRA-1-81 or TRA-1-60 expression. In a further embodiment, the cells may also be enriched for the loss of a detectable marker associated with a retroviral vector.

[00111] In another aspect, the disclosure provides methods of isolating induced stem cells of the disclosure. The human induced pluripotent stem cells of the disclosure can be isolated by, for example, utilizing molecules (*e.g.*, antibodies, antibody derivatives, ligands or Fc-peptide fusion molecules) that bind to a marker (*e.g.*, a TRA-1-81, a TRA-1-60 or a combination of markers) on the human induced pluripotent stem cells and thereby positively selecting cells that bind the molecule (*i.e.*, a positive selection). Other examples of positive selection methods include methods of preferentially promoting the growth of a desired cell type in a mixed population of desired and undesired cell types. Alternatively, by using molecules that bind to markers that are not present on the desired cell type, but that are present on an undesired cell type, the undesired cells containing such markers can be removed from the desired cells (*i.e.*, a negative selection). Other negative selection methods include preferentially killing or inhibiting the growth of an undesired cell type in a mixed population of desired and undesired cell types. Accordingly, by using negative selection, positive selection, or a combination thereof, an enriched population of stem cell can be made.

[00112] Procedures for separation may include magnetic separation, using antibody-coated magnetic beads, affinity chromatography, cytotoxic agents joined to a monoclonal antibody, or such agents used in conjunction with a monoclonal antibody, *e.g.*, complement and cytotoxins, and "panning" with antibody attached to a solid matrix (*e.g.*, plate), or other convenient technique. Techniques

providing accurate separation include fluorescence activated cell sorters, which can have varying degrees of sophistication, e.g., a plurality of color channels, low angle and obtuse light scattering detecting channels, and impedance channels. Conveniently, antibodies may be conjugated with markers, such as magnetic beads, which allow for direct separation, biotin, which can be removed with avidin or streptavidin bound to a support, fluorochromes, which can be used with a fluorescence activated cell sorter, or the like, to allow for ease of separation of the particular cell type. Any technique may be employed which is not unduly detrimental to the viability of the human induced pluripotent stem cells. In one embodiment, the cells are incubated with an antibody against a marker (e.g., a TRA-1-81 antibody) and the cells that stain positive for the marker are manually selected and subcultured.

[00113] Combinations of enrichment methods may be used to improve the time or efficiency of purification or enrichment. For example, after an enrichment step to remove cells having markers that are not indicative of the cell type of interest the cells may be further separated or enriched by a fluorescence activated cell sorter (FACS) or other methodology having high specificity. Multi-color analyses may be employed with a FACS. The cells may be separated on the basis of the level of staining for a particular antigen or lack thereof. Fluorochromes may be used to label antibodies specific for a particular antigen. Such fluorochromes include phycobiliproteins, e.g., phycoerythrin and allophycocyanins, fluorescein, Texas red, and the like.

[00114] Any cell type-specific markers can be used to select for or against a particular cell type. Induced stem cell markers useful for enrichment comprise expressed markers such as TRA-1-81 and loss of markers (e.g., GFP) associated with a retroviral vector or other exogenous vector.

[00115] Once stem cells have been isolated, they optionally can be propagated in appropriate medium in the presence of absence of a feeder layer. In addition, the human induced pluripotent stem cells of the invention may be cultured in a bioreactor system.

[00116] Once the human induced pluripotent stem cells of the disclosure have been established in culture, as described above,

they may be maintained or stored in cell "banks" comprising either continuous *in vitro* cultures of cells requiring regular transfer or cells which have been cryopreserved. In some embodiments, the banked cells are used for autologous treatment of a subject.

[00117] Fibroblasts may be readily isolated by disaggregating an appropriate organ or tissue which is to serve as the source of the fibroblasts. This may be readily accomplished using techniques known to those skilled in the art. For example, the tissue or organ can be disaggregated mechanically and/or treated with digestive enzymes and/or chelating agents that weaken the connections between neighboring cells making it possible to disperse the tissue into a suspension of individual cells without appreciable cell breakage. Enzymatic dissociation can be accomplished by mincing the tissue and treating the minced tissue with any of a number of digestive enzymes either alone or in combination. These include but are not limited to trypsin, chymotrypsin, collagenase, elastase, and/or hyaluronidase, DNase, pronase, dispase *etc.* Mechanical disruption can also be accomplished by a number of methods including, but not limited to, the use of grinders, blenders, sieves, homogenizers, pressure cells, or insonators to name but a few. For a review of tissue disaggregation techniques, see Freshney, *Culture of Animal Cells. A Manual of Basic Technique*, 2d Ed., A.R. Liss, Inc., New York, 1987, Ch. 9, pp. 107-126.

[00118] Once the tissue has been reduced to a suspension of individual cells, the suspension can be fractionated into subpopulations from which the fibroblasts and/or other stromal cells and/or elements can be obtained. This also may be accomplished using standard techniques for cell separation including, but not limited to, cloning and selection of specific cell types, selective destruction of unwanted cells (negative selection), separation based upon differential cell agglutinability in the mixed population, freeze-thaw procedures, differential adherence properties of the cells in the mixed population, filtration, conventional and zonal centrifugation, centrifugal elutriation (counterstreaming centrifugation), unit gravity separation, countercurrent distribution, electrophoresis and fluorescence-activated cell sorting. For a review of clonal

selection and cell separation techniques, see Freshney, Culture of Animal Cells. A Manual of Basic Techniques, 2d Ed., A.R. Liss, Inc., New York, 1987, Ch. 11 and 12, pp. 137-168.

[00119] The isolation of fibroblasts may, for example, be carried out as follows: fresh tissue samples are thoroughly washed and minced in Hanks balanced salt solution (HBSS) in order to remove serum. The minced tissue is incubated from 1-12 hours in a freshly prepared solution of a dissociating enzyme such as trypsin. After such incubation, the dissociated cells are suspended, pelleted by centrifugation and plated onto culture dishes. All fibroblasts will attach before other cells, therefore, appropriate stromal cells can be selectively isolated and grown.

[00120] Where the de-differentiated cells are to be used for transplantation or implantation *in vivo* it is useful to obtain the stromal cells from the patient's own tissues.

[00121] Oligonucleotide probes and primers can be used to identify expression of various factors described herein as well as in cloning and amplification procedures. An oligonucleotide probe or a primer refers to a nucleic acid molecule of between 8 and 2000 nucleotides in length. More particularly, the length of these oligonucleotides can range from about 8, 10, 15, 20, or 30 to 100 nucleotides, but will typically be about 10 to 50 (e.g., 15 to 30 nucleotides). The appropriate length for oligonucleotides in assays of the disclosure under a particular set of conditions may be empirically determined by one of skill in the art.

[00122] Oligonucleotide primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis based upon the known KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG or any combination thereof polynucleotide and polypeptide sequence. Various orthologs from other species are known in the art.

[00123] Oligonucleotide probes and primers can comprise nucleic acid analogs such as, for example, peptide nucleic acids, locked nucleic acid (LNA) analogs, and morpholino analogs. The 3' end of the probe can be functionalized with a capture or detectable label to assist in detection of a KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, Glis1 or any combination thereof nucleic acid.

[00124] Any of the oligonucleotides or nucleic acid of the disclosure can be labeled by incorporating a detectable label measurable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, such labels can comprise radioactive substances (^{32}P , ^{35}S , ^3H , ^{125}I), fluorescent dyes (5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin), biotin, nanoparticles, and the like. Such oligonucleotides are typically labeled at their 3' and 5' ends.

[00125] The oligonucleotide primers and probes can be immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, glass and the like. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips and the like are all suitable examples. Suitable methods for immobilizing oligonucleotides on a solid phase include ionic, hydrophobic, covalent interactions and the like. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. The oligonucleotide probes or primers can be attached to or immobilized on a solid support individually or in groups of about 2-10,000 distinct oligonucleotides of the disclosure to a single solid support. A substrate comprising a plurality of oligonucleotide primers or probes of the disclosure may be used either for detecting or amplifying KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, Glis1 or any combination thereof. For example, the oligonucleotide probes can be used in an oligonucleotide chip such as those marketed by Affymetrix and described in U.S. Pat. No. 5,143,854; PCT publications WO 90/15070 and 92/10092, the disclosures of which are incorporated herein by reference. These arrays can be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis. The disclosure further contemplates antibodies capable

of specifically binding to a KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, or Glis1 polypeptide.

[00126] A reference or control population refers to a group of subjects or individuals who are predicted to be representative of the general population. A test sample is measured for the amount of KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, Glis1 or any combination thereof in the sample, wherein the amount is compared to a control sample.

[00127] In another aspect, the disclosure provides methods of differentiating stem cells along a committed lineage comprising inhibiting the expression or activity of KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, NANOG, Glis1 or any combination thereof. Differentiation agents useful in this regard include, for example, antibodies, antisense oligonucleotides, RNAi constructs, or ribozymes.

[00128] Culture techniques useful in the methods of the disclosure are disclosed in International Patent Publication No. WO 2010/120785, which is incorporated herein by reference.

[00129] The following Examples are provided to illustrate certain aspects of the disclosure and to aid those of skill in the art in practicing the disclosure. These Examples are in no way to be considered to limit the scope of the disclosure in any manner.

EXAMPLES

Example 1

[00130] Cells. BJ foreskin fibroblasts and STO cell line were obtained from ATCC. Primary human foreskin fibroblasts (HFF) and HUES-9 human ES cell line were obtained from existing sources. BJ, HFFs and STO were cultured in DMEM containing 10% FBS, MEM Non-Essential Amino Acids (NEAA), Pyruvate, penicillin, and streptomycin. HUES-9 and iPS cells were cultured with ES culture medium in Knockout D-MEM containing 20% Knockout SR, GlutaMAX, NEAA, 2-Mercaptoethanol (all from Invitrogen), penicillin, streptomycin, and bFGF (10 ng/ml). STO feeder cells were prepared by mitomycin C treatment (10 µg/ml, Sigma). For feeder free culture of iPS cell clones and HUES-9, cells were passaged on Matrigel™ (BD Bioscience) coated wells and cultured in the

conditioned medium prepared from STO feeder cells with ES culture medium.

[00131] Plasmid construction. cDNAs coding for OCT4 (accession no. NM_002701), c-MYC (accession no. NM_002467) and GLIS1 (accession no. BC104911) were obtained from Open biosystems. SOX2 (accession no. NM_003106), KLF4 (accession no. NM_004235), NANOG (accession no. BC099704) are available from ATCC. B18R (accession no. D01019) was obtained from Addgene. The polynucleotide and polypeptide sequences associated with each of the foregoing accession nos. are incorporated herein by reference. The cDNAs were used as templates for PCR amplification to add restriction enzyme sites and/or Kozak sequence, and cloned into pBluescript SK+ vector for checking of cDNA sequences. Then cDNAs were cloned into pTNT vector (Promega) for mRNA synthesis and pCX4bsr1 for the retrovirus production. For the multicistronic expression using viral 2A peptide sequences, F2A oligos, T2A oligos and E2A oligos (Table 1) were annealed and cloned into EcoRI/SpeI, SpeI/XbaI and XbaI/NotI sites of pBluescript SK+ vector, respectively. cDNAs of reprogramming factors were linked with 2A peptide sequences in frame, and then cloned into pVEE-S-IRES-Puro. pVEE-S-IRES-Puro were constructed from p5'VEE/S/GFP/Pac3 to clone reprogramming factors. Briefly, GFP/Pac genes and partial 3'UTR in p5'VEE/S/GFP/Pac were deleted with XbaI/MfeI digestion, and then introduced the multiple cloning sites (MCS; NdeI, AscI, BbvCI, ClaI, MfeI, FseI and NotI) (Table 1), IRES and Puromycin resistance gene from pCX4puro. This vector was renamed as pVEE-IRES-Puro for convenience. To generate RNA with T7 RNA polymerase, the SP6 promoter (ATTTAGGTGACACTATAG (see, e.g., SEQ ID NO:31 from 16844-16861)) was replaced to T7 promoter (TAATACGACTCACTATAG (see, e.g., SEQ ID NO:32 from 16843-16860)) by PCR (**Table 1**) using the SacI/BstZ17I fragment of VEE vector as a template (SP6 promoter is located on next to the SacI site).

[00132] Table 1: PCR Cloning Primers

F2A-Forward	5'-AATTCACCGGTGTGAAACAGACTTTGAATTTGACCTTCTCAAGTTGGCGGGAGACGTGGAGTCCAACCCAGGGCCAGATCTA (SEQ ID NO:11)	F2A-oligo
F2A-Reverse	5'-CTAGTAGATCTGGGCCCTGGGTTGGACTCCACGTCTCCGCGCAACTTGAGAAAGGTCAAAATTCAAAGTCTGTTTCACACCGGTG (SEQ ID NO:12)	F2A-oligo
T2A-F	5'-CTAGTGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCAACAATTGT (SEQ ID NO:13)	T2A-oligo
T2A-R	5'-CTAGACAATTGTGGGCCAGGATTCTCCTCGACGTCACCGCATGTTAGCAGACTTCCTCTGCCCTCA (SEQ ID NO:14)	T2A-oligo
E2A-F	5'-CTAGACAATGTACTAACTACGCTTTGTTGAAACTCGCTGGCGATGTT	E2A-oligo

	GAAAGTAACCCCGGTCCTGGCGCGCCCGC (SEQ ID NO:15)	
E2A-R	5'-GGCCGCGGGCGCGCCAGGACCGGGTTACTTTCAACATCGCCAGC GAGTTTCAACAAAGCGTAGTTAGTACATTGt (SEQ ID NO:16)	E2A-oligo
VEE-MCS-F1	5'-CTAGCATATGGGCGCGCCCTCAGCATCGATGGCCGGCCTCTAGAGC GGCCGC (SEQ ID NO:17)	MCS-oligo
VEE-MCS-R1	5'-GGCCGCGGGCGGCTCTAGAGGCCGGCCATCGATGCTGAGGGCGCGC CCATATG (SEQ ID NO:18)	MCS-oligo
nsP2a-F1	5'-CAGGACGATCTCATTCTCAC (SEQ ID NO:19)	PCR, nsP2
nsP2a-R1	5'-GCTTGCCACTCCTCTATCGTG (SEQ ID NO:20)	PCR, nsP2
nsP4a-F1	5'-CCACAATACGATCGGCAGTG (SEQ ID NO:21)	PCR, nsP4
nsP4a-R1	5'-ATGTCCTGCAACATATTCAAA (SEQ ID NO:22)	PCR, nsP4
hOct4RTa-F1	5'-CGGCGCCAGAAGGGCAAGCG (SEQ ID NO:23)	PCR, OK
hKlf4RTb-R1	5'-CACCTGCTTGACGCAGTGTC (SEQ ID NO:24)	PCR, OK
hKlf4GC2For	5'-GCAGGAGGCGGTCTCTTCGTGCACC (SEQ ID NO:35)	PCR, Klf4
hKlf4GC2Rev	5'-CAGGTGTGCCTTGAGATGGGAAGCTC (SEQ ID NO:36)	PCR, Klf4
Bis-Oct-10F	5'-GGAGTAGAAGGATTGTTTTGGTTTA (SEQ ID NO:25)	bisulfite,
Bis-Oct-9R	5'-AAACCTTAAAACTTAACCAAATCC (SEQ ID NO:26)	bisulfite
Bis-Nanog-4F	5'-AGAGTAGTTGGGATTATAGATATTTA (SEQ ID NO:27)	bisulfite
Bis-Nanog-3R	5'-AACAAACAAACCTAAAAACAAACC (SEQ ID NO:28)	bisulfite
EcoR1-Sac1- T7M1-VEE	5'-CGGAATTCGAGCTCTAATACGACTCACTATAGATGGGCGGCGCATGA GAGAAGCCCAG (SEQ ID NO:37)	T7 VEE PCR
Xba1-BstZ171- VEE	5'-GCTCTAGAGTATACATCCTGGTAAACAGCGACTTGCCC (SEQ ID NO:38)	T7 VEE PCR

[00133] **mRNA and Replicon RNA synthesis.** pTNT-B18R plasmid was used for the synthesis of B18R mRNA. The pTNT vector contains a 5' β -globin leader sequence and a synthetic poly (A) tail (30 bases) to enhance the expression of genes. 30 bases of poly(A) were not enough to stabilize mRNA, so additional poly(A) tail was added by poly(A) tail polymerase. B18R-mRNA synthesis was performed with modified nucleotides using the RiboMAX Large Scale RNA Production System-SP6 (Promega) kit. Modification was performed with replacement of 100% of UTP with pseudouridine (Psi) (TriLink Biotechnologies) or 25% of UTP and CTP with Psi and 5-methylcytidine (5mc) (TriLink Biotechnologies), respectively. After the transcription reaction, DNA template was removed by DNase digestion. The mRNA was purified by extraction with Phenol/Chloroform/Isoamyl alcohol (PCI) and Chloroform/Isoamyl alcohol (CI), and then concentrated by ammonium acetate precipitation (2.5 M), which is selectively precipitates RNA, while leaving most of the protein, DNA and unincorporated NTPs in the supernatant according to the manufacture's protocol (Epicentre). Typically 10 μ g of linearized plasmid for 100 μ l reaction scale was used and received about 400 μ g mRNA. For the 5'-Capping of mRNA, ScriptCap m7G Capping System™ was used and ScriptCap 2'-O-Methyltransferase (Epicentre, currently available from CELLSRIPT)

to produce cap 1-capped RNA, which proceeds to quantitative completion of capping. After 5'-Capping, mRNA was briefly purified by ammonium acetate precipitation, and then additional poly(A) tail was added by Poly(A) Polymerase (Epicentre, currently available from CELLSRIPT). The mRNA bearing 5'-Capping and poly(A) tail was purified by extraction with PCI and CI, followed by ammonium precipitation. For the synthesis of replicon RNA, template plasmid was linearized by digestion with MluI, and then used for RNA synthesis in the same way with mRNA synthesis. The synthesis of RNA replicon was performed without RNA modification. After the DNase treatment, the synthesized RNA was purified by ammonium acetate precipitation without organic purification because most of large RNA was trapped into intermediate phase after organic extraction. The replicon RNA was added 5'-Capping and poly(A) tail as described above, and then purified by ammonium acetate precipitation without organic purification. All RNAs were resuspended in the RNA Storage Solution (Ambion) at 1µg/µl concentration and stored at -80 °C until use.

[00134] Preparation of B18R conditioned medium (B18R-CM). 25% double modified B18R mRNA (1 µg for 1 well of 6-well plate) was transfected into HFFs with Lipofectamine 2000 (Invitrogen). After 3 hr, cells were cultured in Advanced DMEM (Invitrogen) containing 15% FCS (ES cell qualified, Millipore), penicillin, and streptomycin, or ES culture medium. Culture medium was collected on next day, filtrated, and diluted into 5 times with cell culture medium, and then used as B18R-CM (20% B18R-CM). The activity of B18R-CM was briefly measured by the efficiency of repeated transfection of mRNAs.

[00135] iPS generation by replicon transfection. BJ or HFFs were passaged to 6-well plate on day-0 and cultured to ~90-100% confluency (4×10^5 cells/well) on day-1. 1 µg RNA mixture (3:1 ratio VEE RNA Replicon to B18R mRNA) was transfected with Lipofectamine 2000. 25% double modified B18-mRNA or 100% Psi modified mRNA were used for co-transfection. After 3 hr, transfection medium was changed to the Advanced DMEM (Invitrogen) containing 15% FCS (ES cell qualified, Millipore), penicillin, and streptomycin. Cells were cultured in medium containing B18R-CM and

puromycin (0.8 µg/ml) from day-2. Medium was changed every day and transfections were performed every 3 days (day-1, 4, 7, 10 or 14). ES medium was used from day-7. Puromycin was removed at day-7 or day-11. One day after the final transfection, cells were passaged to STO feeder and cultured in ES medium containing B18R-CM. ES medium was changed every day and cultured until iPS cell colonies were generated. Colonies were mechanically picked for isolation of clones or stained with Alkaline Phosphase Detection kit (Millipore) or manually prepared AP-staining solution containing 1 mg/ml of FastRed TR (Sigma) and 0.4 mg/ml of 1-Naphthyl phosphate (Sigma) in AP buffer (100 mM Tris, 100 mM NaCl and 50 mM MgCl₂, pH 9.5)

[00136] RT-PCR for the detection of RNA replicon. Total RNAs were isolated with RNeasy mini kit (Qiagen) or TRIzol (Invitrogen). TRIzol purified RNAs were then purified with ammonium acetate precipitation. Synthesis of cDNAs was performed with QuantiTect Rev. Transcription Kit (Qiagen) or iScript cDNA synthesis kit (Bio-Rad) from 1 µg of total RNA. 1-2 µl of 20 µl RT reaction was used for PCR amplification. PCR was performed with Taq DNA polymerase (NEB) supplemented with PCRx enhancer (Invitrogen): 3 min at 94 °C for initial denature; 36 cycles of 94 °C for 25 sec, 56 °C for 25 sec, 68 °C for 30 sec; followed by 72 °C for 5 min. Primer sequences used RT-PCR were described in Table 1.

[00137] TaqMan RT-PCR. Total RNAs from feeder free culture of iPSCs clones, HUES-9, BJ and HFFs were isolated with RNeasy mini kit. TaqMan RT-PCR reactions were carried out using RNA-to-Ct one-step reaction (Applied Biosystem) according to manufacturer's protocol. 10 ng of total RNA were used per reaction. Primers and probes were obtained from AB TaqMan Gene Expression Assay catalog (GAPDH, Hs99999905_m1; POU5F1 Hs03005111_g1; Sox2 Hs01053049_s1; DNMT3B Hs00171876_m1; TERT Hs00972656_m1; Lin28 Hs00702808_s1; Nanog Hs02387400_g1; TDGF1 Hs02339499_g1). Quantitative PCR reactions were carried out in triplicate, and conditions were as followed: 20 min 55 °C, 10 min 95 °C, 40 cycles of 95 °C for 0.15 min, 65 °C for 1 min. Data were analyzed on the 7300 real-time PCR system (Applied Biosystems) using the delta-delta Ct method.

[00138] Bisulfite genomic sequencing. Conversion of unmethylated cytosines into uracil of genomic DNA was performed with EZ DNA

Methylation-Gold Kit (Zymo Research) according to manufacturer's protocol. Converted genomic DNAs were then used for PCR amplification of promoter region of OCT4 or NANOG with ZymoTaq™ DNA Polymerase (Zymo Research). PCR products were cloned into the T-vector from pBluescript SK+, and then sequenced. Primer sequences used for PCR were described in Table 1.

[00139] Teratoma formation. iPSC clones were cultured with STO feeder cells. Cells were collected by accutase treatment, and then intramuscularly or subcutaneously injected into the hind limb muscles or dorsal flank of nude mice (approximately 10 cm dish cultured cells for 1 shot of injection). After 5 to 8 weeks of injection, tumors were dissected and fixed with 4% paraformaldehyde. Tumors were embedded into paraffin, and sectioning, and then hematoxylin and eosin (H&E) staining or immunostaining of three germ layers markers was performed. AE1/AE3 (cytokeratin), NF-1 (neuronal cells) and GFAP (neuronal cells) were used for markers of ectoderm, Desmin (muscle cells) for marker of mesoderm, and AFP (primitive and definitive endoderm) for marker of endoderm.

[00140] Immunofluorescence staining. Cells were washed twice in PBS and fixed in 4% paraformaldehyde for 10 min. Washed cells were treated with 0.1% Triton X-100 in PBS for 10 min. Cells were blocked with 2% BSA for 1hr at room temperature (RT), and then incubated with primary antibodies in PBS at 4 °C overnight. Cells were washed and incubated with secondary antibodies followed by incubation with DAPI or Hoechst 33342, and then washed and stored in PBS. Primary antibodies such as rabbit anti-Oct4, goat anti-Nanog and anti-Sox2, mouse anti-SSEA4, anti-Tra-1-60 and anti-Tra-1-81 antibodies were used at 1:100 to 1:500 dilutions. Alexa Fluor 488 (BD Biosciences) secondary antibodies were used at 1:800 dilutions.

[00141] Antibodies. Antibodies used in this research are as follows; anti-OCT4 (sc-9081), anti-KLF4 (sc-20691), anti-GLIS1 (sc-67584), anti-c-MYC (sc-42), anti-LIN28 (sc-54030), TRA-1-60 (sc-21705), SSEA1 (sc-21702) and SSEA4 (sc-21704) from Santa Cruz; anti-SOX2 (AF2018) and anti-NANOG (AF1997) from R&D Systems; TRA-1-81 (09-0011) from Stemgent; AE1/AE3 (RB-9010P0), Desmin (MS-376-

S0), AFP (RB-365) and GFAP (RB-087) from Labvision; NF-1 (NB-300-155) from Novus Biological.

[00142] RNA Sequence. Total RNAs were isolated with RNeasy mini kit (Qiagen), and cDNA library of each cells were synthesized and analyzed as known in the art.

[00143] To develop an RNA-based iPS generation strategy, efforts were focused on an approach that: 1) utilized a single RNA species capable of self-replicating for a limited number cell divisions, thereby reducing the number of transfections; 2) was capable of encoding at least four reprogramming factor open reading frames (ORFs); and 3) consistently expressed all four RF genes at high threshold levels over multiple cellular divisions. To ectopically express all four RFs, a modified non-infectious, self-replicating, Venezuelan Equine Encephalitis (VEE) virus RNA replicon was used that is currently being investigated as an expression platform for vaccine development. The VEE replicon is a positive-strand, single RNA species that mimics cellular mRNA with a 5'-Cap and poly(A) tail that does not utilize a DNA intermediate, so there is no potential for genomic integration. VEE encodes four non-structural replication complex proteins (nsP) as a single ORF in the 5' end of the RNA that is separated from the viral structural protein ORF in the 3' end (Fig. 1a). Petrakova *et al.* showed the ability to express exogenous proteins by replacing the 3' structural proteins ORF with GFP. However, exposure of cells to single stranded VEE RNA induces a strong IFN-alpha/beta innate immune response that has severely limited this approach.

[00144] To evaluate the VEE RNA replicon, the 3' ORF was replaced with GFP, followed by an internal ribosomal entry site (IRES) and a Puromycin resistance gene (Puror) (Fig. 1a). VEE-GFP RNA was produced using a standard SP6 polymerase *in vitro* transcription kit followed by 5'-capping, and poly(A) tail addition resulting in a high yield, full length 11,500 nt RNA transcript. To mitigate the innate immune response to VEE-GFP RNA, the B18R protein from Western Vaccinia virus was used, which binds to and neutralizes type I IFNs. A comparison of transfection of primary human foreskin fibroblasts (HFFs) with VEE-GFP RNA alone was performed, in the presence of recombinant B18R protein or with co-transfection

of B18R mRNA. Consistent with induction of a strong innate immune response to cells exposed to single stranded RNA, in the absence of B18R, little to no GFP expression was observed (Fig. 1b). Although addition of recombinant B18R protein increased GFP expression, the GFP fluorescence level was very low. However, co-transfection of VEE-GFP RNA replicon with B18R mRNA resulted in high levels of GFP expression in HFFs (Fig. 1b-d), showing that B18R is required for efficient expression of proteins from the VEE RNA replicon.

[00145] The generation of iPS cells requires consistent, high level expression of reprogramming factors for >7 days; therefore, the persistence of the VEE-GFP replicon in fibroblasts was examined. HFFs were co-transfected with VEE-GFP RNA replicon and B18R mRNA (3:1 ratio) on day 1, then cultured in the presence or absence of B18R conditioned media (CM) plus/minus puromycin on day 2. Although untreated VEE-GFP RNA/B18R mRNA transfected cells showed a high level of GFP expression on day 1, the expression level was rapidly reduced over the next several days to baseline values by day 7 (Fig. 1e). Moreover, in the absence of continuous B18R-CM exposure, VEE-GFP RNA transfected cells stopped growing and/or were killed by the innate immune response (Fig. 1d). In contrast, B18R-CM/puro treated VEE-GFP RNA/B18R mRNA transfected cells maintained persistent high levels of GFP expression in >90% of cells with healthy growth characteristics (Fig. 1d,e). These results showed the ability of B18R exposure to overcome the VEE RNA-induced innate immune response problem and also demonstrated the ability to selectively retain or degrade the VEE RNA replicon from cells by exposure or withdrawal of B18R-CM.

[00146] The VEE RNA replicon 3' ORF was engineered to encode a single combined ORF of three reprogramming factors, OCT4, KLF4, SOX2, separated by internal ribosomal skipping 2A peptides. The ORFs were followed by an IRES then either c-MYC (OKS-iM) or GLIS18 (OKS-iG), which avoids the genomic instability induced by c-MYC, followed by a second IRES and the Puromycin resistance gene (PuroR) (Fig. 1a; Table 1). Similar to the VEE-GFP RNA protocol, VEE-RF RNAs were produced by SP6 *in vitro* transcription, 5'-capping, and poly(A) tail addition resulting in a high yield, full length ~14,500 nt VEE-OKS-iM RNA or ~15,000 nt VEE-OKS-iG RNA. Co-

transfection of VEE-OKS-iM RNA or VEE-OKS-iG RNA replicons plus B18R mRNA (3:1 ratio) into BJ or HFF human fibroblasts resulted in extended high levels of expression of all four RFs that exceeded RF expression levels from retroviruses (Fig 1f). These observations demonstrated the ability to express four reprogramming factors from a single, synthetic VEE-RF RNA replicon in primary human cells, while utilizing B18R to block the innate immune response.

[00147] To develop an RNA-based generation iPS cell protocol, several parameters were evaluated, including number and timing of VEE-RF RNA transfections, selection for VEE-RF RNA replicon retention by puromycin, and the genetic organization of the VEE-RF RNA replicon (Fig. 1a, 2a). Although even a single or double transfection of RF-RNA resulted in iPS cell generation, three or four transfections in the presence of B18R consistently resulted in the highest generation of Alkaline Phosphatase positive (AP+) colonies (Fig. 2b-d). >100 iPS cell colonies were mechanically isolated from the VEE-OKS-iM and VEE-OKS-iG RNA protocols and had a >95% success rate for the ability of isolated iPS-like clones to continuously divide and retain a human embryonic stem cell (hESC) morphology. Of the >100 iPS-like clones isolated, 30 clones were isolated for expression of stem cell markers by immunofluorescence. All 30 VEE RF-RNA iPS clones analyzed (6x HFF-OKS-iM clones, 12x BJ-OKS-iM clones, 6x HFF-OKS-iG clones, 6x BJ-OKS-iG clones) showed strong nuclear staining of endogenous OCT4, SOX2 and NANOG, and strong cell surface staining of SSEA4, TRA-1-60 and TRA-1-81, with negative staining of SSEA1 (Fig. 2e). To eliminate the VEE-RF RNA replicon, all iPS protocols removed B18R-CM and puromycin on day 7 or 10 during reprogramming (Fig. 2a). To confirm the complete loss of VEE RF-RNA replicons, a highly sensitive and specific qRT-PCR protocol was developed capable of detecting <10 femtogram of the VEE RF-RNA replicon (Fig. 4). As expected, qRT-PCR analysis showed that all iPS cell clones had lost the VEE RF-RNA replicon (Table 2). Moreover, karyotype analysis of 4 independent iPS cell clones (BJ-OKS-iM #2 & #21, BJ-OKS-iG #5, HFF-OKS-iM #1) showed normal diploid karyotypes (Fig. 5).

[00148] Table 2: Detection of RF-RNA replicon by qRT-PCR

^a Clones	Passage #								Tfx times						
	P4	P5	P6	P7	P8	P9	P11								
	^b R1R2R3	R1R2R3	R1R2R3	R1R2R3	R1R2R3	R1R2R3	R1R2R3	^c PL	^d FD						
BJ-iM-1		+	+	+		-	-	-		-	-	ND	5	2	
BJ-iM-2			-	+	+/-		-	-	-				5	2	
BJ-iM-3			-	-	+/-		-	-	-				5	2	
BJ-iM-14			-	-	-		-	+/-	-		-	-	ND	1	2
BJ-iM-15			-	-	-		-	-	-				1	2	
BJ-iM-16			-	-	-		-	-	-				1	2	
BJ-iM-20			-	-	-					-	-	-	5	0	
BJ-iM-21			-	+	-		-	+/-	+/-		-	-	-	5	0
BJ-iM-22			-	-	-					-	-	-	5	0	
BJ-iM-23			-	-	-					-	-	-	5	0	
BJ-iM-24			-	+	-		-	-	-		-	-	-	2	0
BJ-iM-25			-	-	-					-	-	-	2	0	
HFF-iM-1	+	+	ND			-	-	-		-	-	-	2	2	
HFF-iM-2	+	+	ND			-	+	+		-	-	-	2	2	
HFF-iM-3	+	+	ND						-	-	-		2	2	
HFF-iM-4	-	+	-			-	-	-		-	-	-	2	2	
HFF-iM-5	-	-	-			-	-	-		-	-	-	2	2	
HFF-iM-6	+	+	+			-	+/-	-		-	-	-	2	2	
HFF-iM-7	+	+	ND						-	-	-		5	2	
HFF-iM-8	+	+	ND	+	+	ND			-	-	-		5	2	
HFF-iM-9	-	+	ND	+/-	+/-	ND			-	-	-		5	2	
HFF-iM-10	+	+	+			-	+/-	+/-		-	-	-	5	2	
HFF-iM-11	-	-	-			-	-	-		-	-	-	5	2	
HFF-iM-12	-	-	-			-	-	-		-	-	-	5	2	
BJ-iG-1	-	-	-			-	-	-		-	-	-	5	0	
BJ-iG-2	-	-	-			-	-	-		-	-	-	5	0	
BJ-iG-3	-	-	-			-	-	-		-	-	-	5	0	
BJ-iG-4	-	-	-			-	-	-		-	-	-	5	0	
BJ-iG-5	-	-	-			-	-	-		-	-	-	5	0	
BJ-iG-6	-	+/-	-			-	-	-		-	-	-	5	0	
HFF-iG-7	-	-	-			-	-	-		-	-	-	4	0	
HFF-iG-8	-	+/-	-			-	-	-		-	-	-	4	0	
HFF-iG-9	-	-	-			-	-	-		-	-	-	4	0	
HFF-iG-10	-	+/-	-			-	+/-	-		-	-	-	4	0	
HFF-iG-11	-	-	+/-			-	+/-	-		-	-	-	4	0	
HFF-iG-12	-	-	-			-	-	-		-	-	-	4	0	

a; iM indicates clones from OKS-iM RNA replicon., iG indicates clones from OKS-iG RNA replicon.
b; regions for RT-PCR, R1; nsP2, R2; nsP4, R3; Oct4-T2A-Klf4 (OK), c; transfection on plate (PL)
before passaging to feeder cells, d; transfection after passaging to feeder cells (FD). +; positive
band detected, +/-; faint band detected, -; no band detected. ND; not done

[00149] To further characterize the established iPS cell clones, the expression of human ES marker genes by qRT-PCR was analyzed. Consistent with expression levels in human HUES9 ES cells, iPS clones generated from both parental BJ and HFF fibroblasts with either the OKS-iM or OKS-iG VEE-RF RNA protocol expressed robust levels of endogenous OCT4, SOX2, NANOG, LIN28, TDGF1, DNMT3B and TERT, in contrast to low or no expression levels in starting parental BJ and HFF fibroblasts (Fig. 3a). A hallmark of induced pluripotency is reduced DNA methylation of CpG dinucleotides in the OCT4 and NANOG promoter regions. Bisulfite genomic sequencing of

both the OCT4 and NANOG promoter regions showed extensive demethylation in iPS cell clones compared to parental fibroblasts (Fig. 3b). To investigate genome-wide mRNA expression profiles in iPS cell clones, whole genome RNA sequencing (RNA-seq) was performed of OKS-iM and OKS-iG VEE-RF RNA generated iPS cell clones, parental BJ and HUES-9 ES cell controls. All four iPS cell clones analyzed by RNA-seq showed unsupervised hierarchical clustering and expression signatures characteristic of human HUES9 ES cells that were highly divergent from parental human fibroblasts (Fig. 3c,d). Lastly, the *in vivo* pluripotency of human iPS cell clones were tested for their ability to differentiate into cells of all three germ layers by teratoma formation in immunocompromised mice. All of the VEE-RF RNA iPS clones analyzed formed teratomas containing representative cell types from the three germ layers, detected by H&E staining that were confirmed by immunohistochemistry staining (Fig. 3e; Fig. 6). Collectively, these observations confirm the ability of both OKS-iM and OKS-iG VEE RF-RNA replicons to efficiently generate pluripotent human iPS cells.

[00150] The generation of iPS cells has great potential for the development of personalized stem cell therapies; however, a straightforward and consistent RNA-based method to generate iPS cells has remained elusive. The disclosure provides a simple, highly reproducible RNA-based approach to generate iPS cells by transfection of a single, synthetic VEE-RF RNA replicon that expresses one, two, three, four or more independent reprogramming factors. VEE-RF RNA generated iPS cells acquired full pluripotency by rigorous *in vivo* biological and molecular criterion that paralleled human ES cells. The generation of the VEE RF-RNA transcript utilizes a standard SP6 *in vitro* transcription kit that does not require special conditions and thereby, further simplifies the approach for broad use. By expressing the four RFs at consistent, high levels over time in the same cell combined with replication of the VEE-RF RNA for a limited number of multiple cell generations, the VEE-RF RNA approach solves both of the major inefficiency problems associated with attempting to generate iPS cells by daily repeated daily transfections for >14 days of four

individual RF mRNAs. Importantly, the VEE-RF RNA is an ectopic hit-and-run approach that does not utilize a DNA intermediate and therefore, there is no opportunity for integrative mutation that can occur with DNA vector-based iPS cell approaches. Moreover, the timing of VEE-RF RNA replicon loss by degradation can be regulated by B18R withdrawal from the media. Using the VEE-RF RNA approach, >100 independent iPS cell clones were generated from both OCT4/KLF4/SOX2/c-MYC and OCT4/KLF4/SOX2/GLIS1 VEE-RF RNA protocols from two independent parental human fibroblast populations. In addition, the VEE-RF RNA approach can be engineered to express alternative RF combinations and/or insertion of additional RF ORFs into the RF-RNA backbone for refining iPS cell generation from specific cell types or for use in driving transdifferentiation. In summary, the VEE-RF RNA replicon approach has broad applicability for the efficient generation of human iPS cells for ultimate use in human stem cell therapies and regenerative medicine.

[00151] ACCESSION NUMBERS. RNA-Seq data have been submitted and can be accessed by the Gene Expression Omnibus (GEO) accession number GSE38265.

[00152] Table 3: iPS Cell Generation with VEE-RF RNA Replicon

RNA Replicon	Cell	Tfx Days	Puromycin selection	AP+ Colonies per starting well
OKS-iM	BJ	d1,	d2-d7	6
OKS-iM	BJ	d1, 2	d2-d7	32
OKS-iM	BJ	d1, 2, 3	d2-d7	221
OKS-iM	BJ	d1, 4, 7, 10	d2-d7	140
OKS-iM	BJ	d1	none	6
OKS-iM	BJ	d1, 2	none	12
OKS-iM	BJ	d1, 2, 3	none	8
OKS-iM	HFF	d1, 5, 9	d2-d10	179
OKS-iM	HFF	d1, 4, 7, 10	d2-d4	189
OKS-iM	HFF	d1, 4, 7, 10	d2-d7	308
OKS-iM	HFF	d1, 4, 7, 10	d2-d10	338
OKS-iG	BJ	d1, 4, 7, 10	d2-d7	282
OKS-iG	BJ	d1, 4, 7, 10	d2-d10	122
OKS-iG	HFF	d1, 4, 7, 10	d2-d7	267
OKS-iG	HFF	d1, 4, 7, 10	d2-d10	248

[00153] Although a number of embodiments and features have been described above, it will be understood by those skilled in the art that modifications and variations of the described embodiments and features may be made without departing from the teachings of the disclosure or the scope of the disclosure as defined by the appended claims.

WHAT IS CLAIMED IS:

1. A purified synthetic RNA replicon comprising:
at least one non-structural replicase domain from an alphavirus and at least one non-alphavirus heterologous sequence encoding one or more factors for inducing the generation of pluripotent stem cells when expressed in a somatic cell.
2. The purified synthetic RNA replicon of claim 1, wherein the replicon comprises sequences obtained from an alphavirus selected from the group consisting of Eastern Equine Encephalitis virus (EEE), Venezuelan Equine Encephalitis virus (VEE), Everglades virus, Mucambo virus, Pixuna virus and Western Equine Encephalitis virus (WEE).
3. The purified synthetic RNA replicon of claim 1, wherein the replicon comprises sequences obtained from an alphavirus selected from the group consisting of Sindbis virus, Semliki Forest virus, Middelburg virus, Chikungunya virus, O'nyong-nyong virus, Ross River virus, Barmah Forest virus, Getah virus, Sagiya virus, Bebaru virus, Mayaro virus, Una virus, Aura virus, Whataroa virus, Babanki virus, Kyzylagach virus, Highlands J virus, Fort Morgan virus, Ndumu virus and Buggy Creek virus.
4. The purified synthetic RNA replicon of claim 1, wherein the at least one non-alphavirus heterologous sequence comprises at least 2, 3, 4 or 5 non-alphavirus heterologous sequences.
5. The purified synthetic RNA replicon of any one of claims 1-4, wherein the non-alphavirus heterologous sequence is selected from a polynucleotide encoding a KLF polypeptide, a SOX-2 polypeptide, a OCT-3/4 polypeptide, a c-MYC or n-MYC or L-MYC polypeptide, a GLIS1 polypeptide, a NANOG polypeptide and any combination thereof.
6. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the KLF polypeptide encodes a KLF

polypeptide having at least 95% identity to a sequence of SEQ ID NO:8.

7. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the KLF polypeptide encodes a KLF polypeptide having a sequence of SEQ ID NO:8.

8. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the KLF polypeptide comprises a sequence as set forth in SEQ ID NO:7, wherein "T" is "U".

9. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the SOX-2 polypeptide encodes a SOX-2 polypeptide having at least 95% identity to a sequence of SEQ ID NO:6.

10. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the SOX-2 polypeptide encodes a SOX-2 polypeptide having a sequence of SEQ ID NO:6.

11. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the Sox-2 polypeptide comprises a sequence as set forth in SEQ ID NO:5, wherein "T" is "U".

12. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the OCT-4 polypeptide encodes a OCT-4 polypeptide having at least 95% identity to a sequence of SEQ ID NO:4.

13. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the OCT-4 polypeptide encodes a OCT-4 polypeptide having a sequence of SEQ ID NO:4.

14. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the OCT-4 polypeptide comprises a sequence as set forth in SEQ ID NO:3, wherein "T" is "U".

15. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the c-MYC polypeptide encodes a c-MYC polypeptide having at least 95% identity to a sequence of SEQ ID NO:10.

16. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the c-MYC polypeptide encodes a c-MYC polypeptide having a sequence of SEQ ID NO:10.

17. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the c-MYC polypeptide comprises a sequence as set forth in SEQ ID NO:9, wherein "T" is "U".

18. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the GLIS1 polypeptide encodes a GLIS1 polypeptide having at least 95% identity to a sequence of SEQ ID NO:34.

19. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the GLIS1 polypeptide encodes a GLIS1 polypeptide having a sequence of SEQ ID NO:34.

20. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the GLIS1 polypeptide comprises a sequence as set forth in SEQ ID NO:33, wherein "T" is "U".

21. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the NANOG polypeptide encodes a NANOG polypeptide having at least 95% identity to a sequence of SEQ ID NO:2.

22. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the NANOG polypeptide encodes a NANOG polypeptide having a sequence of SEQ ID NO:2.

23. The purified synthetic RNA replicon of claim 5, wherein the polynucleotide encoding the NANOG polypeptide comprises a sequence as set forth in SEQ ID NO:1, wherein "T" is "U".

24. The purified synthetic RNA replicon of claim 1, wherein the replicon comprises from 5' to 3': (VEE RNA replicases)-(promoter)-(RF₁)-(self cleaving peptide)-(RF₂)-(self cleaving peptide)-(RF₃)-(IRES or core promoter)-(RF₄)-(IRES or optional promoter)-(optional selectable marker)-(VEE 3'UTR and polyA tail)-(optional selectable marker)-promoter; wherein RF₁₋₄ are factors that induce de-differentiation of a somatic cell to a pluripotent cells, wherein RF₂₋₃ are optional, RF₃₋₄ are optional, or RF₄ is optional; wherein RF₁₋₄ are selected from the group consisting of Oct-4, Klf4, Sox-2, c-Myc, Nanog, and Glis1.

25. The purified synthetic RNA replicon of claim 1 or 24, wherein the replicon comprise a sequence that is 90%, 95%, 98%, 99% or 100% identical to SEQ ID NO:29, 30, 31, or 32 from about position 1 to about position 7561 wherein "T" of the sequence is substituted with "U", followed by one or more RFs, followed by a 3'UTR and polyA tail, wherein the one or more RFs are selected from the group consisting of Oct-3/4, Sox-2, Klf4, c-Myc, Nanog, and Glis1; wherein when more than one RF is present, the coding sequences may be separated by an internal ribosome entry site (IRES) or a small promoter.

26. The purified synthetic RNA replicon of claim 25, wherein the replicon comprises a sequence that is at least 95%, 98%, 99% or 100% identical to a sequence selected from the group consisting of SEQ ID NO:29, 30, 31, or 32, wherein "T" is "U".

27. A composition comprising human cells transfected with a replicon of any one of claims 1-26.

28. The composition of claim 27, further comprising B18R conditioned media.

29. The composition of claim 27, wherein the human cells are somatic cells.

30. The composition of claim 29, wherein the human cells are fibroblasts.

31. An *in vitro* method comprising:

transfecting a human somatic cell with a purified synthetic ectopic self-replicating RNA replicon derived from an alphavirus, comprising polynucleotides encoding at least the following four de-differentiation factors (i) KLF4, (ii) OCT4, (iii) SOX2, (iv) c-MYC or n-MYC or L-MYC, and optionally (v) GLIS1 and (vi) NANOG,

culturing the somatic cell to express the de-differentiation factor;

selecting cells that display a stem cell morphology and/or stem cell markers; and

subculturing the cells to obtain a population of induced stem cells.

32. The method of claim 31, wherein the cells are selected by detecting expression of a Tumor Rejection Antigen 1-60 and/or 1-81.

33. A vector system for producing human stem cells, comprising:

at least one purified synthetic self-replicating RNA replicon derived from an alphavirus, comprising one or more polynucleotides encoding de-differentiation factors selected from the group consisting of a KLF4, OCT4, SOX2, c-MYC or n-MYC or L-MYC, GLIS1, NANOG or any combination thereof.

34. The vector system of claim 33, wherein the replicon comprises:

(a) Oct4, Sox2, Klf4, and c-Myc, or

(b)Oct4, Sox2, Klf4, and Glis1.

35. The vector system of claim 33, wherein the one self-replicating RNA vector is derived from an alphavirus.

36. The vector system of claim 35, wherein the alphavirus is VEE.

37. An isolated human somatic cell comprising a purified synthetic ectopic RNA replicon derived from an alphavirus comprising one or more de-differentiation polynucleotide sequences.

38. The human somatic cell of claim 37, wherein the somatic cell de-differentiates when cultured under conditions to express at least the following four de-differentiation factors selected from the group consisting of a (i) KLF4, (ii) OCT4, (iii) SOX2, (iv) c-MYC or n-MYC or L-MYC, and optionally (v) GLIS1 and (vi) NANOG from the ectopic RNA replicon.

39. A cell population comprising the cells of claim 38.

40. A cell population of claims 39 or method of claim 31, wherein the cells or method further comprise culturing the cells in a medium conditioned with B18R so that the degradation of the ectopic RNA replicon is inhibited.

41. The cell population or method of claim 40, wherein the B18R conditioned media is produced by transfection of B18R mRNA into primary human fibroblasts.