

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

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 October 2011 (20.10.2011)

PCT

(10) International Publication Number  
**WO 2011/130402 A2**

(51) International Patent Classification:  
C12N 5/0735 (2010.01)

(21) International Application Number:  
PCT/US2011/032309

(22) International Filing Date:  
13 April 2011 (13.04.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
61/323,689 13 April 2010 (13.04.2010) US

(71) Applicant (for all designated States except US): **CELLULAR DYNAMICS INTERNATIONAL, INC.** [US/US]; University Research Park, 525 Science Drive, Suite 200, Madison, WI 53711 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **YU, Junying** [—/US]; University Research Park, 525 Science Drive, Suite 200, Madison, WI 53711 (US). **CHAU, Fongching, Kevin** [—/US]; University Research Park, 525 Science Drive, Suite 200, Madison, WI 53711 (US). **JIANG, Jinlan** [—/US]; University Research Park, 525 Science Drive, Suite 200, Madison, WI 53711 (US). **JIANG, Yong** [—/US]; University Research Park, 525 Science Drive, Suite 200, Madison, WI 53711 (US). **VODYANYK, Maksym, A.** [—/US]; University Research Park, 525 Science Drive, Suite 200, Madison, WI 53711 (US).

(74) Agent: **PARKER, David, L.**; Fulbright & Jaworski L.L.P., 98 San Jacinto Blvd., Suite 1100, Austin, TX 78701 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))



WO 2011/130402 A2

(54) Title: HEPATOCYTE PRODUCTION BY FORWARD PROGRAMMING

(57) Abstract: The invention generally features methods for providing hepatocytes from a variety of cell sources, particularly pluripotent stem cells, therapeutic compositions featuring such cells, and methods of using them for the treatment of subjects.

## DESCRIPTION

### HEPATOCYTE PRODUCTION BY FORWARD PROGRAMMING

#### BACKGROUND OF THE INVENTION

5 [0001] The present application claims the priority benefit of United States provisional application number 61/323,689, filed April 13, 2010, the entire contents of which are incorporated herein by reference.

##### **1. Field of the Invention**

10 [0002] The present invention relates generally to the field of molecular biology, stem cells and differentiated cells. More particularly, it concerns programming of somatic cells and undifferentiated cells toward specific cell lineages, particularly hepatic lineage cells.

##### **2. Description of Related Art**

15 [0003] In addition to their use in the transplantation therapies to treat various liver diseases, human hepatocytes are in high demand for drug toxicity screening and development due to their critical functions in the detoxification of drugs or other xenobiotics as well as endogenous substrates. Human primary hepatocytes, however, quickly lose their functions when cultured *in vitro*. Moreover, the drug metabolic ability of human primary hepatocytes exhibits significant difference between different individuals. The availability of an unlimited supply of patient-specific functional hepatocytes would greatly facilitate both the drug development and the eventual clinical application of hepatocyte transplantation.

20 [0004] Therefore, there is a need for production of hepatic lineage cells in therapeutic and research use, especially, human hepatocytes.

#### SUMMARY OF THE INVENTION

25 [0005] The present invention overcomes a major deficiency in the art in providing hepatocytes by forward programming to provide an unlimited supply of patient-specific hepatocytes. In a first embodiment there is provided a method of providing hepatocytes by forward programming of a variety of cell types, including somatic cells or stem cells. Forward programming into hepatocytes may comprise increasing the expression level of a sufficient number of hepatocyte programming factor genes capable of causing forward programming of non-hepatocytes to hepatocytes.

[0006] In another embodiment, there may also be provided a method of directly programming non-hepatocytes, such as differentiation of pluripotent stem cells, into hepatocytes, comprising increasing expression of a sufficient number of hepatocyte programming factor genes (*e.g.*, genes in **Table 1** and variants and isoforms thereof) capable of causing forward programming to a hepatic lineage or to hepatocyte cells, therefore directly programming the cells into hepatocytes.

[0007] “Forward programming,” as used herein, refers to a process having essentially no requirement to culture cells through intermediate cellular stages using culture conditions that are adapted for each such stage and/or, optionally, having no need to add different growth factors during different time points between the starting cell source and the desired end cell product, *e.g.*, hepatocytes, as exemplified in the upper part of **FIG. 1**. The terms “direct programming” or “direct differentiation,” as used in the priority application provisional application 61/323,689, are intended to be commensurate with the term “forward programming,” as used in the present application. “Forward programming” may include programming of a multipotent or pluripotent cell, as opposed to a differentiated somatic cell that has lost multipotency or pluripotency, by artificially increasing the expression of one or more specific lineage-determining genes in a multipotent or pluripotent cell. For example, forward programming may describe the process of programming embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs) to hepatocyte-like cells or other differentiated precursor or somatic cells. In certain other aspects, “forward programming” may refer to “trans-differentiation,” in which differentiated cells are programmed directly into another differentiated cell type without passing through an intermediate pluripotency stage.

[0008] On the other hand, the bottom part of **FIG. 1** demonstrates various developmental stages present in a step-wise differentiation process and the need to add different growth factors at different times during the process, which costs more labor, time and expenses than methods described in certain aspects of the current invention. Therefore, the methods of forward programming in certain aspects of the present invention are advantageous by avoiding the need to add different growth factors at different stages of programming or differentiation to improve efficiency. For example, the medium for culturing the cells to be programmed or progeny cells thereof may be essentially free of one or more of fibroblast growth factors (FGFs), epidermal growth factors (EGFs), and

nicotinamides, which are normally required for progressive programming (*i.e.*, directed differentiation as defined below) along different developmental stages.

[0009] Forward programming as used in certain aspects of the present invention may be different from directed differentiation. In directed differentiation, growth factors or small molecules are added to the culture medium, thereby indirectly causing an increased expression of the endogenous genes. In directed differentiation, the added growth factors or small molecules signal through cell surface proteins and surface protein-mediated signaling to activate endogenous pathways toward the lineage desired. To the contrary, in forward programming, the programming factors that normally are only intra-cellular (*e.g.*, transcription factors) are forced to have an increased expression by introducing or inducing the gene expression cassette or by being added directly (*e.g.*, in the form of polypeptides or RNAs), thereby directly activating the programming factor genes for differentiation directly and by-passing the cell surface proteins and surface protein-mediated signaling pathways. These means for increasing the expression of programming factors may be defined as “artificial,” and may be different from the directed differentiation which comprises adding growth factors or small molecules to the medium thereby indirectly causing increased expression of endogenous programming factor genes.

[0010] Sources of cells suitable for hepatic forward programming may include any stem cells or non-hepatocyte somatic cells. For example, the stem cells may be pluripotent stem cells or any non-pluripotent stem cells. The pluripotent stem cells may be induced pluripotent stem cells, embryonic stem cells, or pluripotent stem cells derived by nuclear transfer or cell fusion. The stem cells may also include multipotent stem cells, oligopotent stem cells, or unipotent stem cells. The stem cells may also include fetal stem cells or adult stem cells, such as hematopoietic stem cells, mesenchymal stem cells, neural stem cells, epithelial stem cells, skin stem cells. In certain aspects, the stem cells may be isolated from umbilical, placenta, amniotic fluid, chorion villi, blastocysts, bone marrow, adipose tissue, brain, peripheral blood, cord blood, menstrual blood, blood vessels, skeletal muscle, skin and liver.

[0011] In other aspects, hepatocytes may be produced by transdifferentiation of non-hepatocyte somatic cells. The somatic cells for hepatic lineage programming can be any cells forming the body of an organism other than hepatocytes. In some embodiments, the somatic cells are human somatic cells such as skin fibroblasts, adipose tissue-derived cells and human

umbilical vein endothelial cells (HUVEC). In a particular aspect, the somatic cells may be immortalized to provide an unlimited supply of cells, for example, by increasing the level of telomerase reverse transcriptase (TERT). This can be effected by increasing the transcription of TERT from the endogenous gene, or by introducing a transgene through any gene delivery method or system.

[0012] Hepatocyte programming factor genes include any genes that, alone or in combination, directly impose hepatic fate upon non-hepatocytes, especially transcription factor genes or genes that are important in hepatic differentiation or hepatic function when expressed in cells. For example, one, two, three, four, five, six, seven, eight, nine, ten, or more of the exemplary genes and isoforms or variants thereof as listed in **Table 1** may be used in certain aspects of the invention. Many of these genes have different isoforms, which might have similar functions and therefore are contemplated for use in certain aspects of the invention.

[0013] In a particular embodiment, the hepatocyte programming factor genes used herein may comprise one, two, three, four, five, or six of Forkhead box protein A1 (FOXA1), forkhead box A2 (FOXA2) (*e.g.*, FOXA2-2; NM\_153675.2), hematopoietically-expressed homeobox protein (HHEX), hepatocyte nuclear factor 1 homeobox A (HNF1A), hepatocyte nuclear factor 4 alpha (HNF4A) (*e.g.*, HNF4A-2; NM\_000457.3), GATA binding protein 4 (GATA4), NR0B2 (nuclear receptor subfamily 0, group B, member 2), sex comb on midleg-like 1 (SCML1), and T-box transcription factor (TBX3) (*e.g.*, TBX3-1; NM\_005996.3). In a more particular aspect, the hepatocyte programming factor genes include FOXA2, HHEX, HNF1A, and HNF4A. For example, the hepatocyte programming factor genes may be a combination of FOXA1, FOXA2, HHEX, HNF1A, HNF4A and TBX3. In another example, the hepatocyte programming factor genes may be a combination of FOXA2, HHEX, HNF4A, GATA4, NR0B2 and SCML1.

[0014] In certain aspects, there is provided a method of providing hepatocytes by forward programming of pluripotent stem cells, comprising: providing the hepatocytes by culturing the pluripotent stem cells under conditions to increasing the expression level of a sufficient number of hepatocyte programming factor genes capable of causing forward programming of the stem cells (*e.g.*, pluripotent stem cells) to hepatocytes, thereby causing the pluripotent stem cells to directly differentiate into hepatocytes.

[0015] The skilled artisan will understand that methods for increasing the expression of the hepatocyte programming factor genes in the cells to be programmed into hepatocytes may include any method known in the art, for example, by induction of expression of one or more expression cassettes previously introduced into the cells, or by introduction of nucleic acids such as DNA or RNA, polypeptides, or small molecules to the cells. Increasing the expression of certain endogenous but transcriptionally repressed programming factor genes may also reverse the silencing or inhibitory effect on the expression of these programming factor genes by regulating the upstream transcription factor expression or epigenetic modulation.

[0016] In one aspect, the cells for hepatic lineage programming may comprise at least one exogenous expression cassette, wherein the expression cassette comprises the hepatocyte programming factor genes in a sufficient number to cause forward programming or transdifferentiation of non-hepatocytes to hepatocytes. The exogenous expression cassette may comprise an externally inducible transcriptional regulatory element for inducible expression of the hepatocyte programming factor genes, such as an inducible promoter comprising a tetracycline response element.

[0017] In a further aspect, one or more of the exogenous expression cassette for hepatocyte programming may be comprised in a gene delivery system. Non-limiting examples of a gene delivery system may include a transposon system, a viral gene delivery system, or an episomal gene delivery system. The viral gene delivery system may be an RNA-based or DNA-based viral vector. The episomal gene delivery system may be a plasmid, an Epstein-Barr virus (EBV)-based episomal vector, a yeast-based vector, an adenovirus-based vector, a simian virus 40 (SV40)-based episomal vector, a bovine papilloma virus (BPV)-based vector, or the like.

[0018] In another aspect, the cells for hepatic lineage programming may be contacted with hepatocyte programming factors in an amount sufficient to cause forward programming of the stem cells to hepatocytes. The hepatocyte programming factors may comprise gene products of the hepatocyte programming factor genes. The gene products may be polypeptides or RNA transcripts of the hepatocyte programming factor genes. In a further aspect, the hepatocyte programming factors may comprise one or more protein transduction domains to facilitate their intracellular entry and/or nuclear entry. Such protein transduction domains are well known in the art, such as an HIV TAT protein transduction domain, HSV

VP22 protein transduction domain, *Drosophila* Antennapedia homeodomain or variants thereof.

[0019] The method may further comprise a selection or enrichment step for the hepatocytes provided from forward programming or transdifferentiation. To aid selection or enrichment, the cells for programming, such as the pluripotent stem cells or progeny cells thereof, may comprise a selectable or screenable reporter expression cassette comprising a reporter gene. The reporter expression cassette may comprise a hepatocyte-specific transcriptional regulatory element operably linked to a reporter gene. Non-limiting examples of hepatocyte-specific transcriptional regulatory element include a promoter of albumin,  $\alpha$ -1-antitrypsin (AAT), cytochrome p450 3A4 (CYP3A4), apolipoprotein A-I, or apoE. Mature hepatocyte-specific transcriptional regulatory element may comprise a promoter of albumin,  $\alpha$ 1-antitrypsin, asialoglycoprotein receptor, cytokeratin 8 (CK8), cytokeratin 18 (CK18), CYP3A4, fumaryl acetoacetate hydrolase (FAH), glucose-6-phosphates, tyrosine aminotransferase, phosphoenolpyruvate carboxykinase, and tryptophan 2,3-dioxygenase.

[0020] Characteristics of the hepatocytes provided in certain aspects of the invention include, but are not limited to one or more of: (i) expression of one or more hepatocyte markers including glucose-6-phosphatase, albumin,  $\alpha$ -1-antitrypsin (AAT), cytokeratin 8 (CK8), cytokeratin 18 (CK18), asialoglycoprotein receptor (ASGR), alcohol dehydrogenase 1, arginase Type I, cytochrome p450 3A4 (CYP3A4), liver-specific organic anion transporter (LST-1), or a combination thereof; (ii) activity of liver-specific enzymes such as glucose-6-phosphatase or CYP3A4, production of by-products such as bile and urea or bile secretion, or xenobiotic detoxification; (iii) hepatocyte morphological features; or (iv) *in vivo* liver engraftment in an immunodeficient subject.

[0021] For selection or enrichment of the hepatocytes, there may be further provided a step by identifying hepatocytes comprising expression of a hepatic reporter gene or one or more hepatocyte characteristics as described herein.

[0022] In particular aspects, the hepatocytes provided herein may be mature hepatocytes. The mature hepatocytes may be selected or enriched by using a screenable or selectable reporter expression cassette comprising a mature hepatocyte-specific transcriptional regulatory element operably linked to a reporter gene, or magnetic cell sorting using antibody against hepatocyte-specific cell surface antigens such as ASGR, or by

assessing characteristic specific for mature hepatocytes as known in the art. For example, mature hepatocytes can be identified by one or more of: the presence of hepatocyte growth factor receptor, albumin,  $\alpha$ 1-antitrypsin, asialoglycoprotein receptor, cytokeratin 8 (CK8), cytokeratin 18 (CK18), CYP3A4, fumaryl acetoacetate hydrolase (FAH), glucose-6-phosphates, tyrosine aminotransferase, phosphoenolpyruvate carboxykinase, and tryptophan 2,3-dioxygenase, and the absence of intracellular pancreas-associated insulin or proinsulin production. In further aspects, hepatocyte-like cells provided herein may be further forward programmed into mature hepatocytes by the artificially increased expression of genes detailed in Table 1.

10           [0023] For production of more mature hepatocytes, the starting cell population may be cultured in a medium comprising one or more growth factors such as Oncostain M (OSM), or further comprising hepatocyte growth factor (HGF). The culturing may be prior to, during, or after the effected expression of hepatocyte programming factors. Hepatocytes may be provided at least, about or up to 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20  
15 days (or any range derivable therein) after the increased expression or culturing in the presence or absence of growth factors.

          [0024] In a further embodiment, a hepatocyte may be produced by any of the methods set forth herein. In certain aspects, there may also be provided a tissue engineered liver comprising the hepatocytes provided by the methods described herein. In another aspect,  
20 there may be provided a hepatocyte-based bio-artificial liver (BAL) comprising the hepatocytes.

          [0025] In certain aspects, the invention provides a cell comprising one or more exogenous expression cassettes comprising one or more hepatocyte programming factor genes (*e.g.*, genes in **Table 1** and isoforms or variants thereof). The exogenous expression  
25 cassettes may comprise two, three, four, five or six of the hepatocyte programming factor gene. For example, the exogenous expression cassettes may comprise FOXA2, HNF4A and one or more additional hepatocyte programming factor genes selected from the group consisting of HHEX, HNF1A, FOXA1, TBX3-1, GATA4, NR0B2, SCML1, CEBPB, HLF, HLX, NR1H3, NR1H4, NR1I2, NR1I3, NR5A2, SEBOX, ZNF391. In a particular example,  
30 the exogenous expression cassettes comprise FOXA1, FOXA2, HHEX, HNF1A, HNF4A, and TBX3. In another example, the hepatocyte programming factor genes may be a combination of FOXA2, HHEX, HNF4A, GATA4, NR0B2 and SCML1.



[0026] For inducible expression of the hepatocyte programming factor genes, at least one of the exogenous expression cassettes may comprise an externally inducible transcriptional regulatory element. In particular aspects, there may be provided a cell comprising one or more exogenous expression cassettes, wherein the one or more exogenous expression cassettes comprise FOXA2, HNF4A and one or more additional hepatocyte programming factor genes selected from the group consisting of HHEX, HNF1A, FOXA1, TBX3-1, GATA4, NR0B2, SCML1, CEBPB, HLF, HLX, NR1H3, NR1H4, NR1I2, NR1I3, NR5A2, SEBOX, ZNF391, and at least one of the exogenous expression cassettes is operably linked to an externally inducible transcriptional regulatory element.

10 [0027] The exogenous expression cassettes may be comprised in one or more gene delivery systems. The gene delivery system may be a transposon system; a viral gene delivery system; an episomal gene delivery system; or a homologous recombination system such as utilizing a zinc finger nuclease, a transcription activator-like effector (TALE) nuclease, or a meganuclease, or the like. The cell may further comprise a screenable or  
15 selectable reporter expression cassette comprising a hepatocyte-specific promoter operably linked to a reporter gene. The hepatocyte-specific transcriptional regulatory element may be a promoter of albumin,  $\alpha$ -1-antitrypsin (AAT), cytochrome p450 3A4 (CYP3A4), apolipoprotein A-I, apoE, or any other hepatocyte-specific promoter or enhancer in the art.

[0028] In one aspect, the cell may be a stem cell or a progeny cell thereof. The stem  
20 cell may be a pluripotent stem cell or any non-pluripotent stem cell. The pluripotent stem cell may be an induced pluripotent stem cell, an embryonic stem cell, or a pluripotent stem cell derived by nuclear transfer or cell fusion. The stem cell may also be a multipotent stem cell, oligopotent stem cell, or unipotent stem cell. The stem cell may also be a fetal stem cell or an  
25 adult stem cell, for example, a hematopoietic stem cell, a mesenchymal stem cell, a neural stem cell, an epithelial stem cell or a skin stem cell. In another aspect, the cell may be a somatic cell, either immortalized or not. The cell may also be a hepatocyte, more particularly, a mature hepatocyte or an immature hepatocyte (*e.g.*, hepatocyte-like cell).

[0029] There may also be provided a composition comprising a cell population comprising two cell types, *i.e.*, the cells to be programmed to hepatocytes and hepatocytes,  
30 and essentially free of other intermediate cell types. For example, such a cell population may have two cell types including the stem cells and hepatocytes but essentially free of other cells types in the intermediate developmental stages along the hepatic differentiation process. In

particular, a composition comprising a cell population consisting of stem cells and hepatocytes may be provided. The stem cells may be particularly pluripotent stem cells, *e.g.*, induced pluripotent stem cells. Hepatocytes may be at least, about, or up to 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.9% (or any intermediate ranges) of the cell population, or any range derivable therein.

[0030] There may be also provided a cell population comprising hepatocytes, wherein at least 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.9% (or any intermediate ranges) of the hematopoietic precursor cells comprise one or more expression cassettes that comprise FOXA2, HNF4A and one or more additional hepatocyte programming factor genes selected from the group consisting of HHEX, HNF1A, FOXA1, TBX3-1, GATA4, NR0B2, SCML1, CEBPB, HLF, HLX, NR1H3, NR1H4, NR1I2, NR1I3, NR5A2, SEBOX, ZNF391.

[0031] The hepatocytes provided herein may be used in any methods and applications currently known in the art for hepatocytes. For example, a method of assessing a compound may be provided, comprising assaying a pharmacological or toxicological property of the compound on the hepatocyte or tissue engineered liver provided herein. There may also be provided a method of assessing a compound for an effect on a hepatocyte, comprising: a) contacting the hepatocyte provided herein with the compound; and b) assaying an effect of the compound on the hepatocyte.

[0032] In a further aspect, there may also be provided a method for treating a subject having or at risk of a liver dysfunction comprising administering to the subject with a therapeutically effective amount of hepatocytes or hepatocyte-containing cell population provided herein.

[0033] Embodiments discussed in the context of methods and/or compositions of the invention may be employed with respect to any other method or composition described herein. Thus, an embodiment pertaining to one method or composition may be applied to other methods and compositions of the invention as well.

[0034] As used herein the terms “encode” or “encoding” with reference to a nucleic acid are used to make the invention readily understandable by the skilled artisan however these terms may be used interchangeably with “comprise” or “comprising” respectively.

[0035] As used herein the specification, "a" or "an" may mean one or more. As used herein in the claim(s), when used in conjunction with the word "comprising", the words "a" or "an" may mean one or more than one.

5 [0036] The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and "and/or." As used herein "another" may mean at least a second or more.

10 [0037] Throughout this application, the term "about" is used to indicate that a value includes the inherent variation of error for the device, the method being employed to determine the value, or the variation that exists among the study subjects.

15 [0038] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

20 [0039] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0040] **FIG. 1.** Alternative approaches for hepatocyte differentiation from human ESC/iPSCs.

25 [0041] **FIG. 2.** The strategy employed for identifying transgenes that could forward program human ESC/iPSCs to mature hepatocytes.

[0042] **FIG. 3.** The establishment of human ESC/iPSC reporter/inducible (R/I) lines for hepatocyte differentiation.

[0043] **FIGS. 4A-4B.** Confirmation of restricted marker gene (mOrange) expression in hepatocytes during normal human ESC differentiation. (**FIG. 4A**) Cellular morphological changes during hepatic differentiation of human ESC R/I lines, and restricted expression of mOrange in maturing hepatocytes. (**FIG. 4B**) Flow cytometric analysis of hepatic marker Albumin and ASGPR1 (marker for mature hepatocytes) in d20 differentiated culture.

[0044] **FIGS. 5A-5C.** Confirmation of the Tet-On inducible gene expression in human H1 ESC R/I lines. (**FIG. 5A**) a two-vector *PiggyBac* stable gene expression system. Ptight: an rtTET-responsive inducible promoter; pEF: the eukaryotic elongation factor 1 $\alpha$  promoter; hPBase: the coding region for the *PiggyBac* transposase with codons optimized for expression in human cells. (**FIG. 5B**) EGFP induction in human ESC R/I lines. The EGFP driven by the Ptight promoter was introduced into human ESC R/I lines using Fugene HD-mediated transfection of both vectors in (**FIG. 5A**). Human ESCs with stable *PiggyBac* transposon integration were selected with geneticin (100  $\mu$ g/ml). Images are shown with human ESC R/I lines after 2 days induction with or without Doxycycline (1  $\mu$ g/ml). (**FIG. 5C**) Flow cytometric analysis of EGFP expression in human ESC R/I lines after 4 days induction with or without Doxycycline (1  $\mu$ g/ml). Gray lines: Human ESC R/I lines without the transfection of the EGFP vector (negative control). Black lines: Human ESC R/I lines with stable *PiggyBac* transposon integration after 4 days induction with or without Doxycycline.

[0045] **FIG. 6.** Direct induction of hepatocytes from human ESC R/I lines through transgene expression.

[0046] **FIGS. 7A-7C.** Forward programming of human ESC R/I lines to hepatocyte-like cells by transgene expression. Among genes that are either implicated in hepatic differentiation during normal mammalian development or enriched in adult hepatocytes (**Table 1**), a combination of genes (FOXA2, HHEX, HNF4A, GATA4, NR0B2 and SCML1) were identified that are sufficient to convert human ESCs directly into hepatocyte-like cells.

[0047] **FIG. 8.** Examples of additional combinations (**Table 2**) that could induce hepatocyte-like cells from human ESC R/I lines via forward programming. AFP, ALB, ASGPR1 expression of cells by forward programming using combinations 4, 5, 11, 12 (C4, C5, C11, C12) in **Table 2** were shown.

## DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0048] The instant invention overcomes several major problems with current technologies by providing methods and compositions for hepatocyte production by programming. In contrast to previous methods using step-wise differentiation protocols, certain aspects of these methods increase the level of hepatocyte programming transcription factors in non-hepatocytes to provide hepatocytes by forward programming. The extra steps including adding different growth factors during various intermediate developmental stages may be unnecessary in certain aspects the present methods. Therefore, certain aspects of the present methods may be more time- and cost-efficient and may enable manufacture of hepatocytes for therapeutics from a renewable source, for example, stem cells. Further embodiments and advantages of the invention are described below.

### **I. Definitions**

[0049] "Programming" is a process that changes a cell to form progeny of at least one new cell type, either in culture or *in vivo*, than it would have under the same conditions without programming. This means that after sufficient proliferation, a measurable proportion of progeny having phenotypic characteristics of the new cell type if essentially no such progeny could form before programming; alternatively, the proportion having characteristics of the new cell type is measurably more than before programming. This process includes differentiation, dedifferentiation and transdifferentiation. "Differentiation" is the process by which a less specialized cell becomes a more specialized cell type. "Dedifferentiation" is a cellular process in which a partially or terminally differentiated cell reverts to an earlier developmental stage, such as pluripotency or multipotency. "Transdifferentiation" is a process of transforming one differentiated cell type into another differentiated cell type. Under certain conditions, the proportion of progeny with characteristics of the new cell type may be at least about 1%, 5%, 25% or more in the in order of increasing preference.

[0050] The term "exogenous," when used in relation to a protein, gene, nucleic acid, or polynucleotide in a cell or organism refers to a protein, gene, nucleic acid, or polynucleotide which has been introduced into the cell or organism by artificial means, or in relation a cell refers to a cell which was isolated and subsequently introduced to other cells or to an organism by artificial means. An exogenous nucleic acid may be from a different organism or cell, or it may be one or more additional copies of a nucleic acid which occurs naturally within the organism or cell. An exogenous cell may be from a different organism, or

it may be from the same organism. By way of a non-limiting example, an exogenous nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature.

5 [0051] By "expression construct" or "expression cassette" is meant a nucleic acid molecule that is capable of directing transcription. An expression construct includes, at the least, one or more transcriptional control elements (such as promoters, enhancers or a structure functionally equivalent thereof) that direct gene expression in one or more desired cell types, tissues or organs. Additional elements, such as a transcription termination signal, may also be included.

10 [0052] A "vector" or "construct" (sometimes referred to as gene delivery system or gene transfer "vehicle") refers to a macromolecule or complex of molecules comprising a polynucleotide to be delivered to a host cell, either *in vitro* or *in vivo*.

15 [0053] A "plasmid", a common type of a vector, is an extra-chromosomal DNA molecule separate from the chromosomal DNA which is capable of replicating independently of the chromosomal DNA. In certain cases, it is circular and double-stranded.

20 [0054] An "origin of replication" ("ori") or "replication origin" is a DNA sequence, *e.g.*, in a lymphotropic herpes virus, that when present in a plasmid in a cell is capable of maintaining linked sequences in the plasmid, and/or a site at or near where DNA synthesis initiates. An ori for EBV includes FR sequences (20 imperfect copies of a 30 bp repeat), and preferably DS sequences, however, other sites in EBV bind EBNA-1, *e.g.*, Rep\* sequences can substitute for DS as an origin of replication (Kirshmaier and Sugden, 1998). Thus, a replication origin of EBV includes FR, DS or Rep\* sequences or any functionally equivalent sequences through nucleic acid modifications or synthetic combination derived therefrom. For example, the present invention may also use genetically engineered replication origin of  
25 EBV, such as by insertion or mutation of individual elements, as specifically described in Lindner, *et. al.*, 2008.

[0055] The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (*i.e.*, is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to a  
30 reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a

reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

[0056] A "gene," "polynucleotide," "coding region," "sequence," "segment," "fragment," or "transgene" which "encodes" a particular protein, is a nucleic acid molecule which is transcribed and optionally also translated into a gene product, *e.g.*, a polypeptide, *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The coding region may be present in either a cDNA, genomic DNA, or RNA form. When present in a DNA form, the nucleic acid molecule may be single-stranded (*i.e.*, the sense strand) or double-stranded. The boundaries of a coding region are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A gene can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the gene sequence.

[0057] The term "control elements" refers collectively to promoter regions, polyadenylation signals, transcription termination sequences, upstream regulatory domains, origins of replication, internal ribosome entry sites ("IRES"), enhancers, splice junctions, and the like, which collectively provide for the replication, transcription, post-transcriptional processing and translation of a coding sequence in a recipient cell. Not all of these control elements need always be present so long as the selected coding sequence is capable of being replicated, transcribed and translated in an appropriate host cell.

[0058] The term "promoter" is used herein in its ordinary sense to refer to a nucleotide region comprising a DNA regulatory sequence, wherein the regulatory sequence is derived from a gene which is capable of binding RNA polymerase and initiating transcription of a downstream (3' direction) coding sequence.

[0059] By "enhancer" is meant a nucleic acid sequence that, when positioned proximate to a promoter, confers increased transcription activity relative to the transcription activity resulting from the promoter in the absence of the enhancer domain.

[0060] By "operably linked" with reference to nucleic acid molecules is meant that two or more nucleic acid molecules (*e.g.*, a nucleic acid molecule to be transcribed, a promoter, and an enhancer element) are connected in such a way as to permit transcription of

the nucleic acid molecule. "Operably linked" with reference to peptide and/or polypeptide molecules is meant that two or more peptide and/or polypeptide molecules are connected in such a way as to yield a single polypeptide chain, *i.e.*, a fusion polypeptide, having at least one property of each peptide and/or polypeptide component of the fusion. The fusion  
5 polypeptide is preferably chimeric, *i.e.*, composed of heterologous molecules.

[0061] "Homology" refers to the percent of identity between two polynucleotides or two polypeptides. The correspondence between one sequence and to another can be determined by techniques known in the art. For example, homology can be determined by a direct comparison of the sequence information between two polypeptide molecules by  
10 aligning the sequence information and using readily available computer programs. Alternatively, homology can be determined by hybridization of polynucleotides under conditions which form stable duplexes between homologous regions, followed by digestion with single strand-specific nuclease(s), and size determination of the digested fragments. Two DNA, or two polypeptide, sequences are "substantially homologous" to each other when at  
15 least about 80%, preferably at least about 90%, and most preferably at least about 95% of the nucleotides, or amino acids, respectively match over a defined length of the molecules, as determined using the methods above.

[0062] The term "cell" is herein used in its broadest sense in the art and refers to a living body which is a structural unit of tissue of a multicellular organism, is surrounded by a  
20 membrane structure which isolates it from the outside, has the capability of self replicating, and has genetic information and a mechanism for expressing it. Cells used herein may be naturally-occurring cells or artificially modified cells (*e.g.*, fusion cells, genetically modified cells, *etc.*).

[0063] As used herein, the term "stem cell" refers to a cell capable of giving rising to  
25 at least one type of a more specialized cell. A stem cells has the ability to self-renew, *i.e.*, to go through numerous cycles of cell division while maintaining the undifferentiated state, and has potency, *i.e.*, the capacity to differentiate into specialized cell types. Typically, stem cells can regenerate an injured tissue. Stem cells herein may be, but are not limited to, embryonic stem (ES) cells, induced pluripotent stem cells, or tissue stem cells (also called tissue-specific  
30 stem cell, or somatic stem cell). Any artificially produced cell which can have the above-described abilities (*e.g.*, fusion cells, reprogrammed cells, or the like used herein) may be a stem cell.



[0064] “Embryonic stem (ES) cells” are pluripotent stem cells derived from early embryos. An ES cell was first established in 1981, which has also been applied to production of knockout mice since 1989. In 1998, a human ES cell was established, which is currently becoming available for regenerative medicine.

5 [0065] Unlike ES cells, tissue stem cells have a limited differentiation potential. Tissue stem cells are present at particular locations in tissues and have an undifferentiated intracellular structure. Therefore, the pluripotency of tissue stem cells is typically low. Tissue stem cells have a higher nucleus/cytoplasm ratio and have few intracellular organelles. Most tissue stem cells have low pluripotency, a long cell cycle, and proliferative ability beyond the  
10 life of the individual. Tissue stem cells are separated into categories, based on the sites from which the cells are derived, such as the dermal system, the digestive system, the bone marrow system, the nervous system, and the like. Tissue stem cells in the dermal system include epidermal stem cells, hair follicle stem cells, and the like. Tissue stem cells in the digestive system include pancreatic (common) stem cells, liver stem cells, and the like. Tissue stem  
15 cells in the bone marrow system include hematopoietic stem cells, mesenchymal stem cells, and the like. Tissue stem cells in the nervous system include neural stem cells, retinal stem cells, and the like.

[0066] “Induced pluripotent stem cells,” commonly abbreviated as iPS cells or iPSCs, refer to a type of pluripotent stem cell artificially prepared from a non-pluripotent cell,  
20 typically an adult somatic cell, or terminally differentiated cell, such as fibroblast, a hematopoietic cell, a myocyte, a neuron, an epidermal cell, or the like, by inserting certain genes, referred to as reprogramming factors.

[0067] “Pluripotency” refers to a stem cell that has the potential to differentiate into all cells constituting one or more tissues or organs, or preferably, any of the three germ  
25 layers: endoderm (interior stomach lining, gastrointestinal tract, the lungs), mesoderm (muscle, bone, blood, urogenital), or ectoderm (epidermal tissues and nervous system). “Pluripotent stem cells” used herein refer to cells that can differentiate into cells derived from any of the three germ layers, for example, direct descendants of totipotent stem cells or induced pluripotent stem cells.

30 [0068] As used herein “totipotent stem cells” refers to cells has the ability to differentiate into all cells constituting an organism, such as cells that are produced from the

fusion of an egg and sperm cell. Cells produced by the first few divisions of the fertilized egg are also totipotent. These cells can differentiate into embryonic and extraembryonic cell types. Pluripotent stem cells can give rise to any fetal or adult cell type. However, alone they cannot develop into a fetal or adult animal because they lack the potential to contribute to  
5 extraembryonic tissue, such as the placenta.

[0069] In contrast, many progenitor cells are multipotent stem cells, *i.e.*, they are capable of differentiating into a limited number of cell fates. Multipotent progenitor cells can give rise to several other cell types, but those types are limited in number. An example of a multipotent stem cell is a hematopoietic cell — a blood stem cell that can develop into  
10 several types of blood cells, but cannot develop into brain cells or other types of cells. At the end of the long series of cell divisions that form the embryo are cells that are terminally differentiated, or that are considered to be permanently committed to a specific function.

[0070] As used herein, the term "somatic cell" refers to any cell other than germ cells, such as an egg, a sperm, or the like, which does not directly transfer its DNA to the next  
15 generation. Typically, somatic cells have limited or no pluripotency. Somatic cells used herein may be naturally-occurring or genetically modified.

[0071] Cells are "substantially free" of certain undesired cell types, as used herein, when they have less than 10% of the undesired cell types, and are "essentially free" of certain cell types when they have less than 1% of the undesired cell types. However, even more  
20 desirable are cell populations wherein less than 0.5% or less than 0.1% of the total cell population comprise the undesired cell types. Thus, cell populations wherein less than 0.1% to 1% (including all intermediate percentages) of the cells of the population comprise undesirable cell types are essentially free of these cell types. A medium may be "essentially free" of certain reagents, as used herein, when there is no externally addition of such agents.  
25 More preferably, these agents are absent or present at a undetectable amount.

[0072] The term "hepatocyte" as used herein is meant to include hepatocyte-like cells that exhibit some but not all characteristics of mature hepatocytes, as well as mature and fully functional hepatocytes. The cells produced by this method may be as at least as functional as the hepatocytes produced by directed differentiation to date. This technique may, as it is  
30 further improved, enable the production of completely fully functional hepatocytes, which

have all characteristics of hepatocytes as determined by morphology, marker expression, *in vitro* and *in vivo* functional assays.

## II. Cells involved in hepatocyte programming

5 [0073] In certain embodiments of the invention, there are disclosed methods and compositions for producing hepatocytes by forward programming of cells which are not hepatocytes. There may be also provided cells that comprise exogenous expression cassettes including one or more hepatocyte programming factor genes and/or reporter expression cassettes specific for hepatocyte identification. In some embodiments, the cells may be stem cells, including but are not limited to, embryonic stem cells, fetal stem cells, or adult stem  
10 cells. In further embodiments, the cells may be any somatic cells.

### A. Stem Cells

[0074] Stem cells are cells found in most, if not all, multi-cellular organisms. They are characterized by the ability to renew themselves through mitotic cell division and differentiating into a diverse range of specialized cell types. The two broad types of  
15 mammalian stem cells are: embryonic stem cells that are found in blastocysts, and adult stem cells that are found in adult tissues. In a developing embryo, stem cells can differentiate into all of the specialized embryonic tissues. In adult organisms, stem cells and progenitor cells act as a repair system for the body, replenishing specialized cells, but also maintain the normal turnover of regenerative organs, such as blood, skin or intestinal tissues.

20 [0075] Human embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSC) are capable of long-term proliferation *in vitro*, while retaining the potential to differentiate into all cell types of the body, including hepatocytes. Thus these cells could potentially provide an unlimited supply of patient-specific functional hepatocytes for both drug development and transplantation therapies. The differentiation of human ESC/iPSCs to  
25 hepatocytes *in vitro* recapitulates normal *in vivo* development, *i.e.* they undergo the following sequential developmental stages: definitive endoderm, hepatic specification, immature hepatocyte and mature hepatocyte (**FIG. 1**). This requires the addition of different growth factors at different stages of differentiation, and generally requires over 20 days of differentiation (**FIG. 4**). More importantly, the human ESC/iPSC-derived hepatocytes  
30 generally are yet to exhibit the full functional spectrum of human primary adult hepatocytes. Certain aspects of the invention provided that hepatocytes such as hepatocyte-like cells or

fully functional hepatocytes could be induced directly from human ESC/iPSCs via expression of a combination of transcription factors important for hepatocyte differentiation/function, similar to the generation of iPSCs, bypassing most, if not all, normal developmental stages (FIG. 1). This approach could be more time- and cost-efficient, and generate hepatocytes with functions highly similar, if not identical, to human primary adult hepatocytes. In addition, human ESC/iPSCs, with their unlimited proliferation ability, have a unique advantage over somatic cells as the starting cell population for hepatocyte differentiation.

### 1. Embryonic stem cells

[0076] Embryonic stem cell lines (ES cell lines) are cultures of cells derived from the epiblast tissue of the inner cell mass (ICM) of a blastocyst or earlier morula stage embryos. A blastocyst is an early stage embryo—approximately four to five days old in humans and consisting of 50–150 cells. ES cells are pluripotent and give rise during development to all derivatives of the three primary germ layers: ectoderm, endoderm and mesoderm. In other words, they can develop into each of the more than 200 cell types of the adult body when given sufficient and necessary stimulation for a specific cell type. They do not contribute to the extra-embryonic membranes or the placenta.

[0077] Nearly all research to date has taken place using mouse embryonic stem cells (mES) or human embryonic stem cells (hES). Both have the essential stem cell characteristics, yet they require very different environments in order to maintain an undifferentiated state. Mouse ES cells may be grown on a layer of gelatin and require the presence of Leukemia Inhibitory Factor (LIF). Human ES cells could be grown on a feeder layer of mouse embryonic fibroblasts (MEFs) and often require the presence of basic Fibroblast Growth Factor (bFGF or FGF-2). Without optimal culture conditions or genetic manipulation (Chambers *et al.*, 2003), embryonic stem cells will rapidly differentiate.

[0078] A human embryonic stem cell may be also defined by the presence of several transcription factors and cell surface proteins. The transcription factors Oct-4, Nanog, and Sox-2 form the core regulatory network that ensures the suppression of genes that lead to differentiation and the maintenance of pluripotency (Boyer *et al.*, 2005). The cell surface antigens most commonly used to identify hES cells include the glycolipids SSEA3 and SSEA4 and the keratan sulfate antigens Tra-1-60 and Tra-1-81.

[0079] Methods for obtaining mouse ES cells are well known. In one method, a preimplantation blastocyst from the 129 strain of mice is treated with mouse antiserum to remove the trophoectoderm, and the inner cell mass is cultured on a feeder cell layer of chemically inactivated mouse embryonic fibroblasts in medium containing fetal calf serum. Colonies of undifferentiated ES cells that develop are subcultured on mouse embryonic fibroblast feeder layers in the presence of fetal calf serum to produce populations of ES cells. In some methods, mouse ES cells can be grown in the absence of a feeder layer by adding the cytokine leukemia inhibitory factor (LIF) to serum-containing culture medium (Smith, 2000). In other methods, mouse ES cells can be grown in serum-free medium in the presence of bone morphogenetic protein and LIF (Ying *et al.*, 2003).

[0080] Human ES cells can be obtained from blastocysts using previously described methods (Thomson *et al.*, 1995; Thomson *et al.*, 1998; Thomson and Marshall, 1998; Reubinoff *et al.*, 2000.) In one method, day-5 human blastocysts are exposed to rabbit anti-human spleen cell antiserum, then exposed to a 1:5 dilution of Guinea pig complement to lyse trophectoderm cells. After removing the lysed trophectoderm cells from the intact inner cell mass, the inner cell mass is cultured on a feeder layer of gamma-inactivated mouse embryonic fibroblasts and in the presence of fetal bovine serum. After 9 to 15 days, clumps of cells derived from the inner cell mass can be chemically (*i.e.* exposed to trypsin) or mechanically dissociated and replated in fresh medium containing fetal bovine serum and a feeder layer of mouse embryonic fibroblasts. Upon further proliferation, colonies having undifferentiated morphology are selected by micropipette, mechanically dissociated into clumps, and replated (see U.S. Patent No. 6,833,269). ES-like morphology is characterized as compact colonies with apparently high nucleus to cytoplasm ratio and prominent nucleoli. Resulting ES cells can be routinely passaged by brief trypsinization or by selection of individual colonies by micropipette. In some methods, human ES cells can be grown without serum by culturing the ES cells on a feeder layer of fibroblasts in the presence of basic fibroblast growth factor (Amit *et al.*, 2000). In other methods, human ES cells can be grown without a feeder cell layer by culturing the cells on a protein matrix such as Matrigel<sup>TM</sup> or laminin in the presence of "conditioned" medium containing basic fibroblast growth factor (Xu *et al.*, 2001). The medium is previously conditioned by coculturing with fibroblasts.

[0081] Methods for the isolation of rhesus monkey and common marmoset ES cells are also known (Thomson, and Marshall, 1998; Thomson *et al.*, 1995; Thomson and Odorico, 2000).

[0082] Another source of ES cells are established ES cell lines. Various mouse cell lines and human ES cell lines are known and conditions for their growth and propagation have been defined. For example, the mouse CGR8 cell line was established from the inner cell mass of mouse strain 129 embryos, and cultures of CGR8 cells can be grown in the presence of LIF without feeder layers. As a further example, human ES cell lines H1, H7, H9, H13 and H14 were established by Thompson *et al.* In addition, subclones H9.1 and H9.2 of the H9 line have been developed. It is anticipated that virtually any ES or stem cell line known in the art and may be used with the present invention, such as, *e.g.*, those described in Yu and Thompson, 2008, which is incorporated herein by reference.

[0083] The source of ES cells for use in connection with the present invention can be a blastocyst, cells derived from culturing the inner cell mass of a blastocyst, or cells obtained from cultures of established cell lines. Thus, as used herein, the term “ES cells” can refer to inner cell mass cells of a blastocyst, ES cells obtained from cultures of inner mass cells, and ES cells obtained from cultures of ES cell lines.

## 2. Induced pluripotent stem cells

[0084] Induced pluripotent stem (iPS) cells are cells which have the characteristics of ES cells but are obtained by the reprogramming of differentiated somatic cells. Induced pluripotent stem cells have been obtained by various methods. In one method, adult human dermal fibroblasts are transfected with transcription factors Oct4, Sox2, c-Myc and Klf4 using retroviral transduction (Takahashi *et al.*, 2007). The transfected cells are plated on SNL feeder cells (a mouse cell fibroblast cell line that produces LIF) in medium supplemented with basic fibroblast growth factor (bFGF). After approximately 25 days, colonies resembling human ES cell colonies appear in culture. The ES cell-like colonies are picked and expanded on feeder cells in the presence of bFGF.

[0085] Based on cell characteristics, cells of the ES cell-like colonies are induced pluripotent stem cells. The induced pluripotent stem cells are morphologically similar to human ES cells, and express various human ES cell markers. Also, when grown under conditions that are known to result in differentiation of human ES cells, the induced

pluripotent stem cells differentiate accordingly. For example, the induced pluripotent stem cells can differentiate into cells having neuronal structures and neuronal markers. It is anticipated that virtually any iPS cells or cell lines may be used with the present invention, including, *e.g.*, those described in Yu and Thompson, 2008.

5           **[0086]** In another method, human fetal or newborn fibroblasts are transfected with four genes, Oct4, Sox2, Nanog and Lin28 using lentivirus transduction (Yu *et al.*, 2007). At 12-20 days post infection, colonies with human ES cell morphology become visible. The colonies are picked and expanded. The induced pluripotent stem cells making up the colonies are morphologically similar to human ES cells, express various human ES cell  
10 markers, and form teratomas having neural tissue, cartilage and gut epithelium after injection into mice.

**[0087]** Methods of preparing induced pluripotent stem cells from mouse are also known (Takahashi and Yamanaka, 2006). Induction of iPS cells typically require the expression of or exposure to at least one member from Sox family and at least one member  
15 from Oct family. Sox and Oct are thought to be central to the transcriptional regulatory hierarchy that specifies ES cell identity. For example, Sox may be Sox-1, Sox-2, Sox-3, Sox-15, or Sox-18; Oct may be Oct-4. Additional factors may increase the reprogramming efficiency, like Nanog, Lin28, Klf4, or c-Myc; specific sets of reprogramming factors may be a set comprising Sox-2, Oct-4, Nanog and, optionally, Lin-28; or comprising Sox-2, Oct4,  
20 Klf and, optionally, c-Myc.

**[0088]** iPS cells, like ES cells, have characteristic antigens that can be identified or confirmed by immunohistochemistry or flow cytometry, using antibodies for SSEA-1, SSEA-3 and SSEA-4 (Developmental Studies Hybridoma Bank, National Institute of Child Health and Human Development, Bethesda Md.), and TRA-1-60 and TRA-1-81 (Andrews *et al.*,  
25 1987). Pluripotency of embryonic stem cells can be confirmed by injecting approximately  $0.5-10 \times 10^6$  cells into the rear leg muscles of 8-12 week old male SCID mice. Teratomas develop that demonstrate at least one cell type of each of the three germ layers.

**[0089]** In certain aspects of the present invention, iPS cells are made from reprogramming somatic cells using reprogramming factors comprising an Oct family member  
30 and a Sox family member, such as Oct4 and Sox2 in combination with Klf or Nanog as described above. The somatic cell for reprogramming may be any somatic cell that can be

induced to pluripotency, such as a fibroblast, a keratinocyte, a hematopoietic cell, a mesenchymal cell, a liver cell, a stomach cell, or a  $\beta$  cell. In a certain aspect, T cells may also be used as source of somatic cells for reprogramming (see U.S. Application No. 61/184,546, incorporated herein by reference).

5           **[0090]** Reprogramming factors may be expressed from expression cassettes comprised in one or more vectors, such as an integrating vector or an episomal vector, *e.g.*, an EBV element-based system (see U.S. Application No. 61/058,858, incorporated herein by reference; Yu *et al.*, 2009). In a further aspect, reprogramming proteins or RNA (such as mRNA or miRNA) could be introduced directly into somatic cells by protein transduction or  
10 RNA transfection (see U.S. Application No. 61/172,079, incorporated herein by reference; Yakubov *et al.*, 2010).

### 3.       **Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer**

**[0091]** Pluripotent stem cells can be prepared by means of somatic cell nuclear transfer, in which a donor nucleus is transferred into a spindle-free oocyte. Stem cells  
15 produced by nuclear transfer are genetically identical to the donor nuclei. In one method, donor fibroblast nuclei from skin fibroblasts of a rhesus macaque are introduced into the cytoplasm of spindle-free, mature metaphase II rhesus macaque oocytes by electrofusion (Byrne *et al.*, 2007). The fused oocytes are activated by exposure to ionomycin, then incubated until the blastocyst stage. The inner cell mass of selected blastocysts are then  
20 cultured to produce embryonic stem cell lines. The embryonic stem cell lines show normal ES cell morphology, express various ES cell markers, and differentiate into multiple cell types both *in vitro* and *in vivo*. As used herein, the term “ES cells” refers to embryonic stem cells derived from embryos containing fertilized nuclei. ES cells are distinguished from embryonic stem cells produced by nuclear transfer, which are referred to as “embryonic stem  
25 cells derived by somatic cell nuclear transfer.”

### 4.       **Other stem cells**

**[0092]** Fetal stem cells are cells with self-renewal capability and pluripotent differentiation potential. They can be isolated and expanded from fetal cytotrophoblast cells (European Patent EP0412700) and chorionic villi, amniotic fluid and the placenta  
30 (WO/2003/042405). These are hereby incorporated by reference in their entirety. Cell surface markers of fetal stem cells include CD117/c-kit<sup>+</sup>, SSEA3<sup>+</sup>, SSEA4<sup>+</sup> and SSEA1.



[0093] Somatic stem cells have been identified in most organ tissues. The best characterized is the hematopoietic stem cell. This is a mesoderm-derived cell that has been purified based on cell surface markers and functional characteristics. The hematopoietic stem cell, isolated from bone marrow, blood, cord blood, fetal liver and yolk sac, is the progenitor cell that reinitiates hematopoiesis for the life of a recipient and generates multiple hematopoietic lineages (see U.S. Pat. No. 5,635,387; 5,460,964; 5,677,136; 5,750,397; 5,759,793; 5,681,599; 5,716,827; Hill *et al.*, 1996). These are hereby incorporated by reference in their entirety. When transplanted into lethally irradiated animals or humans, hematopoietic stem cells can repopulate the erythroid, neutrophil-macrophage, megakaryocyte and lymphoid hematopoietic cell pool. *In vitro*, hematopoietic stem cells can be induced to undergo at least some self-renewing cell divisions and can be induced to differentiate to the same lineages as is seen *in vivo*. Therefore, this cell fulfills the criteria of a stem cell.

[0094] The next best characterized is the mesenchymal stem cells (MSC), originally derived from the embryonic mesoderm and isolated from adult bone marrow, can differentiate to form muscle, bone, cartilage, fat, marrow stroma, and tendon. During embryogenesis, the mesoderm develops into limb-bud mesoderm, tissue that generates bone, cartilage, fat, skeletal muscle and possibly endothelium. Mesoderm also differentiates to visceral mesoderm, which can give rise to cardiac muscle, smooth muscle, or blood islands consisting of endothelium and hematopoietic progenitor cells. Primitive mesodermal or mesenchymal stem cells, therefore, could provide a source for a number of cell and tissue types. A number of mesenchymal stem cells have been isolated (see, for example, U.S. Pat. No. 5,486,359; 5,827,735; 5,811,094; 5,736,396; U.S. Pat. No. 5,837,539; 5,837,670; 5,827,740; Jaiswal *et al.*, 1997; Cassiede *et al.*, 1996; Johnstone *et al.*, 1998; Yoo *et al.*, 1998; Gronthos, 1994; Makino *et al.*, 1999). These are hereby incorporated by reference in their entirety. Of the many mesenchymal stem cells that have been described, all have demonstrated limited differentiation to form only those differentiated cells generally considered to be of mesenchymal origin. To date, the most multipotent mesenchymal stem cell expresses the SH2<sup>+</sup> SH4<sup>+</sup> CD29<sup>+</sup> CD44<sup>+</sup> CD71<sup>+</sup> CD90<sup>+</sup> CD106<sup>+</sup> CD120a<sup>+</sup> CD124<sup>+</sup> CD14<sup>-</sup> CD34<sup>-</sup> CD45<sup>-</sup> phenotype.

[0095] Other stem cells have been identified, including gastrointestinal stem cells, epidermal stem cells, neural and hepatic stem cells, also termed oval cells (Potten, 1998; Watt, 1997; Alison *et al*, 1998).

5 [0096] In some embodiments, the stem cells useful for the method described herein include but not limited to embryonic stem cells, induced pluripotent stem cells, mesenchymal stem cells, bone-marrow derived stem cells, hematopoietic stem cells, chondrocyte progenitor cells, epidermal stem cells, gastrointestinal stem cells, neural stem cells, hepatic stem cells, adipose-derived mesenchymal stem cells, pancreatic progenitor cells, hair follicular stem cells, endothelial progenitor cells and smooth muscle progenitor cells.

10 [0097] In some embodiments, the stem cells used for the method described herein is isolated from umbilical cord, placenta, amniotic fluid, chorion villi, blastocysts, bone marrow, adipose tissue, brain, peripheral blood, the gastrointestinal tract, cord blood, blood vessels, skeletal muscle, skin, liver and menstrual blood. Stem cells prepared in the menstrual blood are called endometrial regenerative cells (Medistem Inc.).

15 [0098] One ordinary skilled artisan in the art can locate, isolate and expand such stem cells. The detailed procedures for the isolation of human stem cells from various sources are described in Current Protocols in Stem Cell Biology (2007) and it is hereby incorporated by reference in its entirety. Alternatively, commercial kits and isolation systems can be used. For example, the BD FACS Aria cell sorting system, BD IMag magnetic cell separation system,  
20 and BD IMag mouse hematopoietic progenitor cell enrichment set from BD Biosciences. Methods of isolating and culturing stem cells from various sources are also described in 5,486,359, 6,991,897, 7,015,037, 7,422,736, 7,410,798, 7,410,773, 7,399,632 and these are hereby incorporated by reference in their entirety.

### **B. Somatic cells**

25 [0099] In certain aspects of the invention, there may also be provided methods of transdifferentiation, *i.e.*, the direct conversion of one somatic cell type into another, *e.g.*, deriving hepatocytes from other somatic cells. Transdifferentiation may involve the use of hepatocyte programming factor genes or gene products to increase expression levels of such genes in somatic cells for production of hepatocytes.

[00100] However, the human somatic cells may be limited in supply, especially those from living donors. In certain aspects to provide a unlimited supply of starting cells for programming, somatic cells may be immortalized by introduction of immortalizing genes or proteins, such as hTERT or oncoogenes. The immortalization of cells may be reversible (*e.g.*,  
5 using removable expression cassettes) or inducible (*e.g.*, using inducible promoters).

[00101] Somatic cells in certain aspects of the invention may be primary cells (non-immortalized cells), such as those freshly isolated from an animal, or may be derived from a cell line (immortalized cells). The cells may be maintained in cell culture following their isolation from a subject. In certain embodiments the cells are passaged once or more  
10 than once (*e.g.*, between 2-5, 5-10, 10-20, 20-50, 50-100 times, or more) prior to their use in a method of the invention. In some embodiments the cells will have been passaged no more than 1, 2, 5, 10, 20, or 50 times prior to their use in a method of the invention. They may be frozen, thawed, *etc.*

[00102] The somatic cells used or described herein may be native somatic cells,  
15 or engineered somatic cells, *i.e.*, somatic cells which have been genetically altered. Somatic cells of the present invention are typically mammalian cells, such as, for example, human cells, primate cells or mouse cells. They may be obtained by well-known methods and can be obtained from any organ or tissue containing live somatic cells, *e.g.*, blood, bone marrow, skin, lung, pancreas, liver, stomach, intestine, heart, reproductive organs, bladder, kidney,  
20 urethra and other urinary organs, *etc.*

[00103] Mammalian somatic cells useful in the present invention include, but are not limited to, Sertoli cells, endothelial cells, granulosa epithelial cells, neurons, pancreatic islet cells, epidermal cells, epithelial cells, hepatocytes, hair follicle cells, keratinocytes, hematopoietic cells, melanocytes, chondrocytes, lymphocytes (B and T  
25 lymphocytes), erythrocytes, macrophages, monocytes, mononuclear cells, cardiac muscle cells, and other muscle cells, *etc.*

[00104] In some embodiments cells are selected based on their expression of an endogenous marker known to be expressed only or primarily in a desired cell type. For example, vimentin is a fibroblast marker. Other useful markers include various keratins, cell  
30 adhesion molecules such as cadherins, fibronectin, CD molecules, *etc.* The population of somatic cells may have an average cell cycle time of between 18 and 96 hours, *e.g.*, between

24-48 hours, between 48-72 hours, *etc.* In some embodiments, at least 90%, 95%, 98%, 99%, or more of the cells would be expected to divide within a predetermined time such as 24, 48, 72, or 96 hours.

[00105] Methods described herein may be used to program one or more  
5 somatic cells, *e.g.*, colonies or populations of somatic cells into hepatocytes. In some  
embodiments a population of cells of the present invention is substantially uniform in that at  
least 90% of the cells display a phenotype or characteristic of interest. In some embodiments  
at least 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9, 99.95% or more of the cells display  
a phenotype or characteristic of interest. In certain embodiments of the invention the somatic  
10 cells have the capacity to divide, *i.e.*, the somatic cells are not post-mitotic.

[00106] Somatic cells may be partially or completely differentiated.  
Differentiation is the process by which a less specialized cell becomes a more specialized cell  
type. Cell differentiation can involve changes in the size, shape, polarity, metabolic activity,  
gene expression and/or responsiveness to signals of the cell. For example, hematopoietic stem  
15 cells differentiate to give rise to all the blood cell types including myeloid (monocytes and  
macrophages, neutrophils, basophils, eosinophils, erythrocytes, megakaryocytes/platelets,  
dendritic cells) and lymphoid lineages (T-cells, B-cells, NK-cells). During progression along  
the path of differentiation, the ultimate fate of a cell becomes more fixed. As described  
herein, both partially differentiated somatic cells and fully differentiated somatic cells can be  
20 programmed as described herein to produce desired cell types such as hepatocytes.

### III. Hepatocyte programming factors

[00107] Certain aspects of the invention provide hepatocyte programming  
factors for hepatocyte forward programming. The hepatocytes could be produced directly  
from other cell sources by increasing the level of hepatocyte programming factors in cells.  
25 The numerous functions of hepatocytes could be controlled at the transcriptional level by the  
concerted actions of a limited number of hepatocyte-enriched transcription factors. Any  
transcription factors important for hepatocyte differentiation or function may be used herein,  
like hepatocyte-enriched transcription factors, particularly the genes thereof listed in **Table 1**.  
All the isoforms and variants of the genes listed in **Table 1** may be included in this invention,  
30 and non-limiting examples of accession numbers for certain isoforms or variants are  
provided.

[00108] For example, by effecting expression of a combination of transcription factors in **Table 1**, forward programming into hepatocytes from pluripotent stem cells may bypass most, if not all, normal developmental stages. Examples shown are a combination of the following transcription factors.: FOXA1, FOXA2, HHEX, HNF1A, HNF4A and TBX3, or a combination of FOXA2, HHEX, HNF4A, GATA4, NR0B2 and SCML1.

**Table 1.** A list of genes for forward programming to hepatocytes.

#	Symbol	Entrez Gene ID	Exemplary Accession No.	SEQ ID NO:	Name
1	SOX17	64321	NM_022454.3	9	SRY (sex determining region Y)-box 17 [Homo sapiens]
2	FOXA1	3169	NM_004496.2	10	forkhead box A1 [Homo sapiens]
3	FOXA2	3170	NM_021784.4 NM_153675.2	11 12	forkhead box A2 [Homo sapiens]
4	HHEX	3087	NM_002729.4	13	hematopoietically expressed homeobox [Homo sapiens]
5	GATA4	2626	NM_002052.3	14	GATA binding protein 4 [Homo sapiens]
6	GATA6	2627	NM_005257.3	15	GATA binding protein 6 [Homo sapiens]
7	PROX1	5629	NM_002763.3	16	prospero homeobox 1 [Homo sapiens]
8	TBX3	6926	NM_005996.3 NM_016569.3	17 18	T-box 3 [Homo sapiens]
9	HLX	3142	NM_021958.3	19	H2.0-like homeobox [Homo sapiens]
10	ONECUT1	3175	NM_004498.1	20	one cut homeobox 1 [Homo sapiens]
11	ONECUT2	9480	NM_004852.2	21	one cut homeobox 2 [Homo sapiens]
12	MYC	4609	NM_002467.4	22	v-myc myelocytomatosis viral oncogene homolog (avian) [Homo sapiens]
13	FOXA3	3171	NM_004497.2	23	forkhead box A3 [Homo sapiens]
14	HNF4A	3172	NM_000457.3	24	hepatocyte nuclear factor 4, alpha [Homo sapiens]
15	HNF1A	6927	NM_000545.5	25	HNF1 homeobox A [Homo sapiens]
16	HNF1B	6928	NM_000458.2	26	HNF1 homeobox B [Homo sapiens]
17	CEBPA	1050	NM_004364.3	27	CCAAT/enhancer binding protein (C/EBP), alpha [Homo sapiens]
18	CEBPB	1051	NM_005194.2	28	CCAAT/enhancer binding protein (C/EBP), beta [Homo sapiens]
19	DBP	1628	NM_001352.3	29	D site of albumin promoter (albumin D-box) binding protein [Homo sapiens]
20	ZBTB20	26137	NM_001164342.1 NM_001164343.1	30 31	zinc finger and BTB domain containing 20 [Homo sapiens]
21	NR1I3	9970			nuclear receptor subfamily 1, group I, member 3 [Homo sapiens]
22	NR1I2	8856	NM_003889.3 NM_022002.2	32 33	nuclear receptor subfamily 1, group I, member 2 [Homo sapiens]
23	NR1H4	9971	NM_005123.2	34	nuclear receptor subfamily 1, group H, member 4 [Homo sapiens]
24	ATF5	22809	NM_012068.5	35	activating transcription factor 5 [Homo sapiens]
25	NR5A2	2494	NM_003822.3	36	nuclear receptor subfamily 5, group A, member 2 [Homo sapiens]
26	NR1H3	10062	NM_005693.2	37	nuclear receptor subfamily 1, group H, member 3 [Homo sapiens]

27	CREB3L3	84699	NM_032607.1	38	cAMP responsive element binding protein 3-like 3 [Homo sapiens]
28	NKX2-8	26257	NM_014360.2	39	NK2 homeobox 8 [ <i>Homo sapiens</i> ]
29	CEBPD	1052	NM_005195.3	40	CCAAT/enhancer binding protein (C/EBP), delta [Homo sapiens]
30	HLF	3131	NM_002126.4	41	hepatic leukemia factor [ <i>Homo sapiens</i> ]
31	NR0B2	8431	NM_021969.2	42	nuclear receptor subfamily 0, group B, member 2 [Homo sapiens]
32	ABLIM3	22885	NM_014945.2	43	actin binding LIM protein family, member 3 [ <i>Homo sapiens</i> ]
33	ATOH8	84913	NM_032827.6	44	atonal homolog 8 ( <i>Drosophila</i> ) [Homo sapiens]
34	C14orf39	317761	NM_174978.2	45	chromosome 14 open reading frame 39 [Homo sapiens]
35	SCML1	6322	NM_001037540.1 NM_006746.4	46 47	sex comb on midleg-like 1 ( <i>Drosophila</i> ) [Homo sapiens]
36	SEBOX	645832	NM_001083896.1	48	SEBOX homeobox [Homo sapiens]
37	ZBED3	84327	NM_032527.3	49	zinc finger, BED-type containing 3 [Homo sapiens]
38	ZGPAT	84619	NM_032527.3 NM_181485.2	49 50	zinc finger, CCCH-type with G patch domain [Homo sapiens]
39	ZNF391	346157	NM_001076781.1	51	zinc finger protein 391 [Homo sapiens]
40	ZNF426	79088	NM_024106.1	52	zinc finger protein 426 [Homo sapiens]
41	ZNF517	340385	NM_213605.2	53	zinc finger protein 517 [Homo sapiens]

[00109] The hepatocyte-enriched transcription factors include, but are not limited to, hepatocyte nuclear factor 1- $\alpha$  (HNF-1 $\alpha$ ), -1 $\beta$ , -3 $\alpha$ , -3 $\beta$ , -3 $\gamma$ , -4 $\alpha$ , and -6 and members of the *c/ebp* family). Hepatocyte nuclear factors (HNFs) are a group of phylogenetically unrelated transcription factors that regulate the transcription of a diverse group of genes into proteins. These proteins include blood clotting factors and in addition, enzymes and transporters involved with glucose, cholesterol, and fatty acid transport and metabolism. Of these, HNF4A (also known as HNF4 $\alpha$  or nuclear receptor 2A1 or (NR2A1)) and HNF1A (*i.e.*, HNF1 $\alpha$ ) appear to be correlated with the differentiated phenotype of cultured hepatoma cells. HNF1A-null mice are viable, indicating that this factor is not an absolute requirement for the formation of an active hepatic parenchyma. In contrast, HNF4A-null mice die during embryogenesis. HNF4A is expressed early in development, visible by *in situ* hybridization in the mouse visceral endoderm at embryonic day 4.5, long before liver development. Whereas HNF4A appears to be essential in the visceral endoderm it may not be necessary for the earliest steps in the development of the fetal liver (Li *et al.*, 2000). HNF-4A is both essential for hepatocyte differentiation during mammalian liver development and also crucial for metabolic regulation and proper liver function (Hayhurst *et al.*, 2001). HNF-4A is also known as TCF; HNF4; MODY; MODY1; NR2A1; TCF14; HNF4a7; HNF4a8; HNF4a9; NR2A21; and FLJ39654. Six transcriptional variants or isoforms are produced from the

genomic gene, isoforms a, b c, d, d, e, and f (Genbank Accession NOs: NM\_000457.3, NM\_001030003.1, NM\_001030004.1, NMJ75914.3, NM\_178849.1, and NM\_178850.1). All isoforms contain a zinc finger, C4 type DNA binding domain and ligand-binding domain. The encoded protein is a nuclear transcription factor which binds DNA as a homodimer and controls the expression of several genes, including HNF1A, a transcription factor which in turns regulates the expression of several hepatic genes. Over 55 distinct target genes have been identified for HNF4A. Since many of those genes contain more than one HNF4A binding site, the total number of distinct, non species redundant HNF4A binding sites is now 74. These genes can be grouped into several different categories, according to function, such as nutrient transport and metabolism, blood maintenance, immune function, liver differentiation and growth factors. The best characterized HNF-4A target genes are those involved in lipid transport (*e.g.*, apolipoprotein genes) and glucose metabolism (*e.g.*, L-PK and PEPCK). Nearly all of the target genes identified thus far are expressed primarily in the liver; several are expressed in other organs as well, such as the pancreas.

**[00110]** HNF1A is also known as HNF1, LFBI, TCF1, and M0DY3. HNF1A is a transcription factor that is highly expressed in the liver and is involved in the regulation of the expression of several liver specific genes such as the human class I alcohol dehydrogenase. HNF-1A (Genbank Accession No: NM 000545.4) belongs to the homeobox gene family for it contains a homeobox DNA binding domain. A homeobox is a DNA sequence that binds DNA. The translated homeobox is a highly conserved stretch of 60 amino acid residues.

**[00111]** Forkhead box A2 (FOXA2), is also known as HNF-3 $\beta$ , HNF3B, TCF3B and MGC19807. FOXA2 is a member of the forkhead class of DNA-binding proteins. The forkhead box is a sequence of 80 to 100 amino acids that form a motif that binds to DNA. This forkhead motif is also known as the winged helix due to the butterfly-like appearance of the loops in the protein structure of the domain. These hepatocyte nuclear factors are transcriptional activators for liver-specific genes such as albumin and transthyretin, and they also interact with chromatin. Similar family members in mice have roles in the regulation of metabolism and in the differentiation of the pancreas and liver. This gene has been linked to sporadic cases of maturity-onset diabetes of the young. Transcript variants encoding different isoforms, isoform 1 and 2, have been identified for this gene (Genbank Accession Nos: NM 021784.4; FOXA2-1) and NM\_153675.2; FOXA2-2).

[00112] CCAAT/enhancer binding protein (C/EBP) alpha is a CCAAT/enhancer-binding protein. C/EBPs are a family of transcription factors that are critical for cellular differentiation, terminal functions and inflammatory response. Six members of the family have been characterized (C/EBP alpha, C/EBP beta, C/EBP delta, C/EBP epsilon, C/EBP gamma and C/EBP zeta) and are distributed in a variety of tissues.

[00113] Hematopoietically-expressed homeobox protein HHEX is a protein that in humans is encoded by the HHEX gene. This gene encodes a member of the homeobox family of transcription factors, many of which are involved in developmental processes. HHEX is required for early development of the liver. A null mutation of HHEX results in a failure to form the liver bud and embryonic lethality.

[00114] T-box transcription factor TBX3 is a protein that in humans is encoded by the TBX3 gene. This gene is a member of a phylogenetically conserved family of genes that share a common DNA-binding domain, the T-box. T-box genes encode transcription factors involved in the regulation of developmental processes. This protein is a transcriptional repressor and is thought to play a role in the anterior/posterior axis of the tetrapod forelimb. Mutations in this gene cause ulnar-mammary syndrome, affecting limb, apocrine gland, tooth, hair, and genital development. Alternative splicing of this gene results in three transcript variants encoding different isoforms.

[00115] GATA4 gene encodes a member of the GATA family of zinc finger transcription factors. Members of this family recognize the GATA motif which is present in the promoters of many genes. GATA4 protein is thought to regulate genes involved in embryogenesis and in myocardial differentiation and function. Mutations in this gene have been associated with cardiac septal defects as well as reproductive defects.

[00116] NR0B2 gene (nuclear receptor subfamily 0, group B, member 1; previous name is dosage-sensitive sex reversal (DAX1)) encodes a protein that contains a DNA-binding domain. The encoded protein acts as a dominant-negative regulator of transcription which is mediated by the retinoic acid receptor. This protein also functions as an anti-testis gene by acting antagonistically to Sry. Mutations in this gene result in both X-linked congenital adrenal hypoplasia and hypogonadotropic hypogonadism. The encoded protein plays an important role in the normal development of several hormone-producing tissues. These tissues include the adrenal glands, the pituitary gland and hypothalamus which



are located in the brain and the male and female reproductive structures (the testes and ovaries). The encoded protein controls the activity of certain genes in the cells that form these tissues during embryonic development. Proteins that control the activity of other genes are known as transcription factors. The encoded protein also plays a role in regulating hormone production in these tissues after they have been formed.

[00117] SCML1 (Sex comb on midleg-like protein 1) encodes a putative Polycomb group (PcG) protein. PcG proteins act by forming multiprotein complexes, which are required to maintain the transcriptionally repressive state of homeotic genes throughout development. The encoded protein may be involved in spermatogenesis during sexual maturation

#### IV. Delivery of gene or gene products

[00118] In certain embodiments, vectors for delivery of nucleic acids encoding hepatic lineage programming or differentiation factors could be constructed to express these factors in cells. Details of components of these vectors and delivery methods are disclosed below. In addition, protein transduction compositions or methods may be also used to effect expression of the hepatocyte programming factors.

[00119] In a further aspect, the following systems and methods may also be used in delivery of reporter expression cassette for identification of desired cell types, such as hepatocytes. In particular, a hepatocyte-specific regulatory element may be used to drive expression of a reporter gene, therefore hepatocytes derived from forward programming may be characterized, selected or enriched.

##### A. Nucleic acid delivery systems

[00120] One of skill in the art would be well equipped to construct a vector through standard recombinant techniques (see, for example, Sambrook *et al.*, 2001 and Ausubel *et al.*, 1996, both incorporated herein by reference). Vectors include but are not limited to, plasmids, cosmids, viruses (bacteriophage, animal viruses, and plant viruses), and artificial chromosomes (*e.g.*, YACs), such as retroviral vectors (*e.g.* derived from Moloney murine leukemia virus vectors (MoMLV), MSCV, SFFV, MPSV, SNV, *etc.*), lentiviral vectors (*e.g.* derived from HIV-1, HIV-2, SIV, BIV, FIV *etc.*), adenoviral (Ad) vectors including replication competent, replication deficient and gutless forms thereof, adeno-associated viral (AAV) vectors, simian virus 40 (SV-40) vectors, bovine papilloma virus

vectors, Epstein-Barr virus, herpes virus vectors, vaccinia virus vectors, Harvey murine sarcoma virus vectors, murine mammary tumor virus vectors, Rous sarcoma virus vectors.

## 1. Viral Vectors

[00121] In generating recombinant viral vectors, non-essential genes are typically replaced with a gene or coding sequence for a heterologous (or non-native) protein. Viral vectors are a kind of expression construct that utilizes viral sequences to introduce nucleic acid and possibly proteins into a cell. The ability of certain viruses to infect cells or enter cells *via* receptor-mediated endocytosis, and to integrate into host cell genome and express viral genes stably and efficiently have made them attractive candidates for the transfer of foreign nucleic acids into cells (*e.g.*, mammalian cells). Non-limiting examples of virus vectors that may be used to deliver a nucleic acid of certain aspects of the present invention are described below.

[00122] Retroviruses have promise as gene delivery vectors due to their ability to integrate their genes into the host genome, transferring a large amount of foreign genetic material, infecting a broad spectrum of species and cell types and of being packaged in special cell lines (Miller, 1992).

[00123] In order to construct a retroviral vector, a nucleic acid is inserted into the viral genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the *gag*, *pol*, and *env* genes but without the LTR and packaging components is constructed (Mann *et al.*, 1983). When a recombinant plasmid containing a cDNA, together with the retroviral LTR and packaging sequences is introduced into a special cell line (*e.g.*, by calcium phosphate precipitation for example), the packaging sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). The media containing the recombinant retroviruses is then collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression require the division of host cells (Paskind *et al.*, 1975).

[00124] Lentiviruses are complex retroviruses, which, in addition to the common retroviral genes *gag*, *pol*, and *env*, contain other genes with regulatory or structural

function. Lentiviral vectors are well known in the art (see, for example, Naldini *et al.*, 1996; Zufferey *et al.*, 1997; Blomer *et al.*, 1997; U.S. Patents 6,013,516 and 5,994,136).

[00125] Recombinant lentiviral vectors are capable of infecting non-dividing cells and can be used for both *in vivo* and *ex vivo* gene transfer and expression of nucleic acid sequences. For example, recombinant lentivirus capable of infecting a non-dividing cell wherein a suitable host cell is transfected with two or more vectors carrying the packaging functions, namely gag, pol and env, as well as rev and tat is described in U.S. Patent 5,994,136, incorporated herein by reference.

## 2. Episomal Vectors

10 [00126] The use of plasmid- or liposome-based extra-chromosomal (*i.e.*, episomal) vectors may be also provided in certain aspects of the invention. Such episomal vectors may include, *e.g.*, oriP-based vectors, and/or vectors encoding a derivative of EBNA-1. These vectors may permit large fragments of DNA to be introduced to a cell and maintained extra-chromosomally, replicated once per cell cycle, partitioned to daughter cells  
15 efficiently, and elicit substantially no immune response.

[00127] In particular, EBNA-1, the only viral protein required for the replication of the oriP-based expression vector, does not elicit a cellular immune response because it has developed an efficient mechanism to bypass the processing required for presentation of its antigens on MHC class I molecules (Levitskaya *et al.*, 1997). Further,  
20 EBNA-1 can act in *trans* to enhance expression of the cloned gene, inducing expression of a cloned gene up to 100-fold in some cell lines (Langle-Rouault *et al.*, 1998; Evans *et al.*, 1997). Finally, the manufacture of such oriP-based expression vectors is inexpensive.

[00128] Other extra-chromosomal vectors include other lymphotropic herpes virus-based vectors. Lymphotropic herpes virus is a herpes virus that replicates in a lymphoblast (*e.g.*, a human B lymphoblast) and becomes a plasmid for a part of its natural  
25 life-cycle. Herpes simplex virus (HSV) is not a "lymphotropic" herpes virus. Exemplary lymphotropic herpes viruses include, but are not limited to EBV, Kaposi's sarcoma herpes virus (KSHV); Herpes virus saimiri (HS) and Marek's disease virus (MDV). Also other sources of episome-base vectors are contemplated, such as yeast ARS, adenovirus, SV40, or  
30 BPV.

[00129] One of skill in the art would be well equipped to construct a vector through standard recombinant techniques (see, for example, Maniatis *et al.*, 1988 and Ausubel *et al.*, 1994, both incorporated herein by reference).

5 [00130] Vectors can also comprise other components or functionalities that further modulate gene delivery and/or gene expression, or that otherwise provide beneficial properties to the targeted cells. Such other components include, for example, components that influence binding or targeting to cells (including components that mediate cell-type or tissue-specific binding); components that influence uptake of the vector nucleic acid by the cell; components that influence localization of the polynucleotide within the cell after uptake (such  
10 as agents mediating nuclear localization); and components that influence expression of the polynucleotide.

[00131] Such components also might include markers, such as detectable and/or selection markers that can be used to detect or select for cells that have taken up and are expressing the nucleic acid delivered by the vector. Such components can be provided as a  
15 natural feature of the vector (such as the use of certain viral vectors which have components or functionalities mediating binding and uptake), or vectors can be modified to provide such functionalities. A large variety of such vectors are known in the art and are generally available. When a vector is maintained in a host cell, the vector can either be stably replicated by the cells during mitosis as an autonomous structure, incorporated within the genome of the  
20 host cell, or maintained in the host cell's nucleus or cytoplasm.

### 3. Transposon-based system

[00132] According to a particular embodiment the introduction of nucleic acids may use a transposon - transposase system. The used transposon - transposase system could be the well known Sleeping Beauty, the Frog Prince transposon - transposase system (for the  
25 description of the latter see *e.g.* EP1507865), or the TTAA-specific transposon piggyBac system.

[00133] Transposons are sequences of DNA that can move around to different positions within the genome of a single cell, a process called transposition. In the process, they can cause mutations and change the amount of DNA in the genome. Transposons were  
30 also once called jumping genes, and are examples of mobile genetic elements.

[00134] There are a variety of mobile genetic elements, and they can be grouped based on their mechanism of transposition. Class I mobile genetic elements, or retrotransposons, copy themselves by first being transcribed to RNA, then reverse transcribed back to DNA by reverse transcriptase, and then being inserted at another position in the genome. Class II mobile genetic elements move directly from one position to another using a transposase to "cut and paste" them within the genome.

#### 4. Homologous Recombination Nuclease-based Systems

[00135] Homologous recombination (HR) is a targeted genome modification technique that has been the standard method for genome engineering in mammalian cells since the mid 1980s. The efficiency of standard HR in mammalian cells is only  $10^{-6}$  to  $10^{-9}$  of cells treated (Capecchi, 1990). The use of meganucleases, or homing endonucleases, such as I-SceI have been used to increase the efficiency of HR. Both natural meganucleases as well as engineered meganucleases with modified targeting specificities have been utilized to increase HR efficiency (Pingoud and Silva, 2007; Chevalier *et al.*, 2002). Another path toward increasing the efficiency of HR has been to engineer chimeric endonucleases with programmable DNA specificity domains (Silva *et al.*, 2011). Zinc-finger nucleases (ZFN) are one example of such a chimeric molecule in which Zinc-finger DNA binding domains are fused with the catalytic domain of a Type IIS restriction endonuclease such as FokI (as reviewed in Durai *et al.*, 2005; PCT/US2004/030606). Another class of such specificity molecules includes Transcription Activator Like Effector (TALE) DNA binding domains fused to the catalytic domain of a Type IIS restriction endonuclease such as FokI (Miller *et al.*, 2011; PCT/IB2010/000154).

#### B. Regulatory Elements:

[00136] Eukaryotic expression cassettes included in the vectors preferably contain (in a 5'-to-3' direction) a eukaryotic transcriptional promoter operably linked to a protein-coding sequence, splice signals including intervening sequences, and a transcriptional termination/polyadenylation sequence.

##### 1. Promoter/Enhancers

[00137] A "promoter" is a control sequence that is a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind, such as RNA polymerase and

other transcription factors, to initiate the specific transcription a nucleic acid sequence. The phrases “operatively positioned,” “operatively linked,” “under control,” and “under transcriptional control” mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional initiation and/or expression of that sequence.

[00138] A promoter generally comprises a sequence that functions to position the start site for RNA synthesis. The best known example of this is the TATA box, but in some promoters lacking a TATA box, such as, for example, the promoter for the mammalian terminal deoxynucleotidyl transferase gene and the promoter for the SV40 late genes, a discrete element overlying the start site itself helps to fix the place of initiation. Additional promoter elements regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have been shown to contain functional elements downstream of the start site as well. To bring a coding sequence "under the control of" a promoter, one positions the 5' end of the transcription initiation site of the transcriptional reading frame "downstream" of (*i.e.*, 3' of) the chosen promoter. The "upstream" promoter stimulates transcription of the DNA and promotes expression of the encoded RNA.

[00139] The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the tk promoter, the spacing between promoter elements can be increased to 50 bp apart before activity begins to decline. Depending on the promoter, it appears that individual elements can function either cooperatively or independently to activate transcription. A promoter may or may not be used in conjunction with an “enhancer,” which refers to a cis-acting regulatory sequence involved in the transcriptional activation of a nucleic acid sequence.

[00140] A promoter may be one naturally associated with a nucleic acid sequence, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment and/or exon. Such a promoter can be referred to as “endogenous.” Similarly, an enhancer may be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence. Alternatively, certain advantages will be gained by positioning the coding nucleic acid segment under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with a

nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers also to an enhancer not normally associated with a nucleic acid sequence in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes, and promoters or enhancers isolated from any other virus, or prokaryotic or eukaryotic cell, and promoters or enhancers not “naturally occurring,” *i.e.*, containing different elements of different transcriptional regulatory regions, and/or mutations that alter expression. For example, promoters that are most commonly used in recombinant DNA construction include the  $\beta$ -lactamase (penicillinase), lactose and tryptophan (*trp*) promoter systems. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR™, in connection with the compositions disclosed herein (see U.S. Patent Nos. 4,683,202 and 5,928,906, each incorporated herein by reference). Furthermore, it is contemplated the control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts, and the like, can be employed as well.

**[00141]** Naturally, it will be important to employ a promoter and/or enhancer that effectively directs the expression of the DNA segment in the organelle, cell type, tissue, organ, or organism chosen for expression. Those of skill in the art of molecular biology generally know the use of promoters, enhancers, and cell type combinations for protein expression, (see, for example Sambrook *et al.* 1989, incorporated herein by reference). The promoters employed may be constitutive, tissue-specific, inducible, and/or useful under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins and/or peptides. The promoter may be heterologous or endogenous.

**[00142]** Additionally any promoter/enhancer combination (as per, for example, the Eukaryotic Promoter Data Base EPDB, through world wide web at [epd.isb-sib.ch/](http://epd.isb-sib.ch/)) could also be used to drive expression. Use of a T3, T7 or SP6 cytoplasmic expression system is another possible embodiment. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial polymerase is provided, either as part of the delivery complex or as an additional genetic expression construct.

**[00143]** Non-limiting examples of promoters include early or late viral promoters, such as, SV40 early or late promoters, cytomegalovirus (CMV) immediate early promoters,

Rous Sarcoma Virus (RSV) early promoters; eukaryotic cell promoters, such as, *e. g.*, beta actin promoter (Ng, 1989; Quitsche *et al.*, 1989), GADPH promoter (Alexander *et al.*, 1988, Ercolani *et al.*, 1988), metallothionein promoter (Karin *et al.*, 1989; Richards *et al.*, 1984); and concatenated response element promoters, such as cyclic AMP response element promoters (cre), serum response element promoter (sre), phorbol ester promoter (TPA) and response element promoters (tre) near a minimal TATA box. It is also possible to use human growth hormone promoter sequences (*e.g.*, the human growth hormone minimal promoter described at Genbank, accession no. X05244, nucleotide 283-341) or a mouse mammary tumor promoter (available from the ATCC, Cat. No. ATCC 45007). A specific example could be a phosphoglycerate kinase (PGK) promoter.

[00144] Tissue-specific transgene expression, especially for reporter gene expression (such as antibiotic resistant gene expression) in hepatocytes produced from forward programming, is desirable as a way to identify produced hepatocytes. To increase both specificity and activity, the use of cis-acting regulatory elements has been contemplated. For example, a hepatocyte-specific promoter may be used, such as a promoter of albumin,  $\alpha$ -1-antitrypsin (AAT), cytochrome p450 3A4 (CYP3A4), apolipoprotein A-I, or APOE.

[00145] In certain aspects, this also concerns enhancer sequences, *i.e.* nucleic acid sequences that increase a promoter's activity and that have the potential to act in cis, and regardless of their orientation, even over relatively long distances (up to several kilobases away from the target promoter). However, enhancer function is not necessarily restricted to such long distances as they may also function in close proximity to a given promoter. For the liver, numerous approaches to incorporate such organ-specific regulatory sequences into retroviral, lentiviral, adenoviral and adeno-associated viral vectors or non-viral vectors (often in addition to house-keeping hepatocyte-specific cellular promoters) have been reported so far (Ferry *et al.*, 1998; Ghosh *et al.*, 2000; Miao *et al.*, 2000; Follenzi *et al.*, 2002).

[00146] Several enhancer sequences for liver-specific genes have been documented. , WO2009130208 describes several liver-specific regulatory enhancer sequences. WO95/011308 describes a gene therapy vector comprising a hepatocyte-specific control region (HCR) enhancer linked to a promoter and a transgene. The human apolipoprotein E-Hepatocyte Control Region (ApoE-HCR) is a locus control region (LCR) for liver-specific expression of the apolipoprotein E (ApoE) gene. The ApoE-HCR is located in the ApoE/CI/CII locus, has a total length of 771 bp and is important in expression of the



genes ApoE and ApoC-1 in the liver (Simonet *et al.*, 1993). In WO01/098482, the combination of this specific ApoE enhancer sequence or a truncated version thereof with hepatic promoters is suggested. It was shown that vector constructs combining the (non-truncated) ApoE-HCR enhancer with a human alpha-antitrypsin (AAT) promoter were able to produce the highest level of therapeutic protein *in vivo* (Miao *et al.*, 2000) and may confer sustained expression when used in conjunction with a heterologous transgene (Miao *et al.*, 2001 ).

[00147] This ApoE-HCR-AAT expression cassette as used, *e.g.*, in the pAAV-ApoHCR-AAT-FIXIA construct (VandenDriessche *et al.*, 2007) is one of the most potent liver-specific FIX expression constructs known, and has been successfully applied in a phase 1/2 dose- escalation clinical study in humans with severe hemophilia B (Manno *et al.*, 2006). The expression of this hFIX minigene is driven from an ApoE-HCR joined to the human AAT promoter. The 5'-flanking sequence of the human AAT gene contains multiple cis-regulatory elements, including a distal enhancer and proximal sequences, with a total length of around 1.2 kb. It was shown to be sufficient to confer tissue specificity *in vivo* by driving gene expression primarily in the liver and also, to a lesser extent, in other tissues known to express AAT (Shen *et al.*, 1989). A 347 bp fragment of this 1.2 kb region in combination with the ApoE enhancer is capable of achieving long-term liver-specific gene expression *in vivo* (Le *et al.*, 1997). Interestingly, this shorter promoter targets expression to the liver with a greater specificity than that reported for larger AAT promoter fragments (Yull *et al.*, 1995).

[00148] Other chimeric liver-specific constructs have also been proposed in the literature, *e.g.*, with the AAT promoter and the albumin or hepatitis B enhancers (Kramer *et al.*, 2003), or the alcohol dehydrogenase 6 (ADH6) basal promoter linked to two tandem copies of the apolipoprotein E enhancer element (Gehrke *et al.*, 2003). The authors of the latter publication stress the importance of the relatively small size (1068 bp) of this enhancer-promoter combination.

## 2. Initiation Signals and Internal Ribosome Binding Sites

[00149] A specific initiation signal also may be used for efficient translation of coding sequences. These signals include the ATG initiation codon or adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be "in-frame"

with the reading frame of the desired coding sequence to ensure translation of the entire insert. The exogenous translational control signals and initiation codons can be either natural or synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements.

5           **[00150]** In certain embodiments of the invention, the use of internal ribosome entry sites (IRES) elements are used to create multigene, or polycistronic, messages. IRES elements are able to bypass the ribosome scanning model of 5' methylated Cap dependent translation and begin translation at internal sites (Pelletier and Sonenberg, 1988). IRES elements from two members of the picornavirus family (polio and encephalomyocarditis) 10 have been described (Pelletier and Sonenberg, 1988), as well an IRES from a mammalian message (Macejak and Sarnow, 1991). IRES elements can be linked to heterologous open reading frames. Multiple open reading frames can be transcribed together, each separated by an IRES, creating polycistronic messages. By virtue of the IRES element, each open reading frame is accessible to ribosomes for efficient translation. Multiple genes can be efficiently 15 expressed using a single promoter/enhancer to transcribe a single message (see U.S. Patent Nos. 5,925,565 and 5,935,819, each herein incorporated by reference).

### 3. Origins of Replication

**[00151]** In order to propagate a vector in a host cell, it may contain one or more origins of replication sites (often termed "ori"), for example, a nucleic acid sequence 20 corresponding to oriP of EBV as described above or a genetically engineered oriP with a similar or elevated function in programming, which is a specific nucleic acid sequence at which replication is initiated. Alternatively a replication origin of other extra-chromosomally replicating virus as described above or an autonomously replicating sequence (ARS) can be employed.

### 25           4. Selection and Screenable Markers

**[00152]** In certain embodiments of the invention, cells containing a nucleic acid construct of the present invention may be identified *in vitro* or *in vivo* by including a marker in the expression vector. Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression vector. Generally, a selection 30 marker is one that confers a property that allows for selection. A positive selection marker is one in which the presence of the marker allows for its selection, while a negative selection

marker is one in which its presence prevents its selection. An example of a positive selection marker is a drug resistance marker.

[00153] Usually the inclusion of a drug selection marker aids in the cloning and identification of transformants, for example, genes that confer resistance to neomycin, puromycin, hygromycin, DHFR, GPT, zeocin and histidinol are useful selection markers. In addition to markers conferring a phenotype that allows for the discrimination of transformants based on the implementation of conditions, other types of markers including screenable markers such as GFP, whose basis is colorimetric analysis, are also contemplated. Alternatively, screenable enzymes as negative selection markers such as herpes simplex virus thymidine kinase (*tk*) or chloramphenicol acetyltransferase (CAT) may be utilized. One of skill in the art would also know how to employ immunologic markers, possibly in conjunction with FACS analysis. The marker used is not believed to be important, so long as it is capable of being expressed simultaneously with the nucleic acid encoding a gene product. Further examples of selection and screenable markers are well known to one of skill in the art. One feature of the present invention includes using selection and screenable markers to select for hepatocytes after the programming factors have effected a desired programming change in those cells.

### C. Nucleic acid Delivery

[00154] Introduction of a nucleic acid, such as DNA or RNA, into cells to be programmed with the current invention may use any suitable methods for nucleic acid delivery for transformation of a cell., as described herein or as would be known to one of ordinary skill in the art. Such methods include, but are not limited to, direct delivery of DNA such as by *ex vivo* transfection (Wilson *et al.*, 1989, Nabel *et al.*, 1989), by injection (U.S. Patent Nos. 5,994,624, 5,981,274, 5,945,100, 5,780,448, 5,736,524, 5,702,932, 5,656,610, 5,589,466 and 5,580,859, each incorporated herein by reference), including microinjection (Harland and Weintraub, 1985; U.S. Patent No. 5,789,215, incorporated herein by reference); by electroporation (U.S. Patent No. 5,384,253, incorporated herein by reference; Tur-Kaspa *et al.*, 1986; Potter *et al.*, 1984); by calcium phosphate precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe *et al.*, 1990); by using DEAE-dextran followed by polyethylene glycol (Gopal, 1985); by direct sonic loading (Fechheimer *et al.*, 1987); by liposome mediated transfection (Nicolau and Sene, 1982; Fraley *et al.*, 1979; Nicolau *et al.*, 1987; Wong *et al.*, 1980; Kaneda *et al.*, 1989; Kato *et al.*, 1991) and receptor-

mediated transfection (Wu and Wu, 1987; Wu and Wu, 1988); by microprojectile bombardment (PCT Application Nos. WO 94/09699 and 95/06128; U.S. Patent Nos. 5,610,042; 5,322,783 5,563,055, 5,550,318, 5,538,877 and 5,538,880, and each incorporated herein by reference); by agitation with silicon carbide fibers (Kaeppler *et al.*, 1990; U.S. Patent Nos. 5,302,523 and 5,464,765, each incorporated herein by reference); by *Agrobacterium*-mediated transformation (U.S. Patent Nos. 5,591,616 and 5,563,055, each incorporated herein by reference); by desiccation/inhibition-mediated DNA uptake (Potrykus *et al.*, 1985), and any combination of such methods. Through the application of techniques such as these, organelle(s), cell(s), tissue(s) or organism(s) may be stably or transiently transformed.

### 1. Liposome-Mediated Transfection

[00155] In a certain embodiment of the invention, a nucleic acid may be entrapped in a lipid complex such as, for example, a liposome. Liposomes are vesicular structures characterized by a phospholipid bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers (Ghosh and Bachhawat, 1991). Also contemplated is an nucleic acid complexed with Lipofectamine (Gibco BRL) or Superfect (Qiagen). The amount of liposomes used may vary upon the nature of the liposome as well as the cell used, for example, about 5 to about 20  $\mu\text{g}$  vector DNA per 1 to 10 million of cells may be contemplated.

[00156] Liposome-mediated nucleic acid delivery and expression of foreign DNA *in vitro* has been very successful (Nicolau and Sene, 1982; Fraley *et al.*, 1979; Nicolau *et al.*, 1987). The feasibility of liposome-mediated delivery and expression of foreign DNA in cultured chick embryo, HeLa and hepatoma cells has also been demonstrated (Wong *et al.*, 1980).

[00157] In certain embodiments of the invention, a liposome may be complexed with a hemagglutinating virus (HVJ). This has been shown to facilitate fusion with the cell membrane and promote cell entry of liposome-encapsulated DNA (Kaneda *et al.*, 1989). In other embodiments, a liposome may be complexed or employed in conjunction with nuclear non-histone chromosomal proteins (HMG-1) (Kato *et al.*, 1991). In yet further embodiments,

a liposome may be complexed or employed in conjunction with both HVJ and HMG-1. In other embodiments, a delivery vehicle may comprise a ligand and a liposome.

## 2. Electroporation

[00158] In certain embodiments of the present invention, a nucleic acid is introduced into an organelle, a cell, a tissue or an organism *via* electroporation. Electroporation involves the exposure of a suspension of cells and DNA to a high-voltage electric discharge. Recipient cells can be made more susceptible to transformation by mechanical wounding. Also the amount of vectors used may vary upon the nature of the cells used, for example, about 5 to about 20  $\mu\text{g}$  vector DNA per 1 to 10 million of cells may be contemplated.

[00159] Transfection of eukaryotic cells using electroporation has been quite successful. Mouse pre-B lymphocytes have been transfected with human kappa-immunoglobulin genes (Potter *et al.*, 1984), and rat hepatocytes have been transfected with the chloramphenicol acetyltransferase gene (Tur-Kaspa *et al.*, 1986) in this manner.

## 3. Calcium Phosphate

[00160] In other embodiments of the present invention, a nucleic acid is introduced to the cells using calcium phosphate precipitation. Human KB cells have been transfected with adenovirus 5 DNA (Graham and Van Der Eb, 1973) using this technique. Also in this manner, mouse L (A9), mouse C127, CHO, CV-1, BHK, NIH3T3 and HeLa cells were transfected with a neomycin marker gene (Chen and Okayama, 1987), and rat hepatocytes were transfected with a variety of marker genes (Rippe *et al.*, 1990).

## 4. DEAE-Dextran

[00161] In another embodiment, a nucleic acid is delivered into a cell using DEAE-dextran followed by polyethylene glycol. In this manner, reporter plasmids were introduced into mouse myeloma and erythroleukemia cells (Gopal, 1985).

## 5. Sonication Loading

[00162] Additional embodiments of the present invention include the introduction of a nucleic acid by direct sonic loading. LTK<sup>-</sup> fibroblasts have been transfected with the thymidine kinase gene by sonication loading (Fechheimer *et al.*, 1987).

## 6. Microprojectile Bombardment

[00163] Microprojectile bombardment techniques can be used to introduce a nucleic acid into at least one, organelle, cell, tissue or organism (U.S. Patent No. 5,550,318; U.S. Patent No. 5,538,880; U.S. Patent No. 5,610,042; and PCT Application WO 94/09699; each of which is incorporated herein by reference). This method depends on the ability to accelerate DNA-coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter cells without killing them (Klein *et al.*, 1987). There are a wide variety of microprojectile bombardment techniques known in the art, many of which are applicable to the invention.

[00164] In this microprojectile bombardment, one or more particles may be coated with at least one nucleic acid and delivered into cells by a propelling force. Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force (Yang *et al.*, 1990). The microprojectiles used have consisted of biologically inert substances such as tungsten or gold particles or beads. Exemplary particles include those comprised of tungsten, platinum, and preferably, gold. It is contemplated that in some instances DNA precipitation onto metal particles would not be necessary for DNA delivery to a recipient cell using microprojectile bombardment. However, it is contemplated that particles may contain DNA rather than be coated with DNA. DNA-coated particles may increase the level of DNA delivery via particle bombardment but are not, in and of themselves, necessary.

[00165] For the bombardment, cells in suspension are concentrated on filters or solid culture medium. Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate.

### B. Protein Transduction

[00166] In certain aspects of the present invention, the cells to be programmed into hepatocytes may be contacted with hepatocyte programming factors comprising polypeptides of hepatocyte transcription factor genes at a sufficient amount for forward programming. Protein transduction has been used as a method for enhancing the delivery of macromolecules into cells. Protein transduction domains may be used to introduce hepatocyte programming polypeptides or functional fragments thereof directly into cells. Research by

many groups has shown that a region of the TAT protein which is derived from the HIV Tat protein can be fused to a target protein allowing the entry of the target protein into the cell. A particular exemplary protein sequence of this domain is RKKRRQRRR (SEQ ID NO:1) where R encodes Arginine, K encodes Lysine and Q encodes Glutamine. This sequence has  
5 been shown to enable the entry of a protein fusion both as an N-terminal or C-terminal fusion. The mechanism of TAT mediated entry is thought to be by macropinocytosis (Gump and Dowdy).

[00167] A "protein transduction domain" or "PTD" is an amino acid sequence that can cross a biological membrane, particularly a cell membrane. When attached to a  
10 heterologous polypeptide, a PTD can enhance the translocation of the heterologous polypeptide across a biological membrane. The PTD is typically covalently attached (*e.g.*, by a peptide bond) to the heterologous DNA binding domain. For example, the PTD and the heterologous DNA binding domain can be encoded by a single nucleic acid, *e.g.*, in a common open reading frame or in one or more exons of a common gene. An exemplary PTD  
15 can include between 10-30 amino acids and may form an amphipathic helix. Many PTD's are basic in character. For example, a basic PTD can include at least 4, 5, 6 or 8 basic residues (*e.g.*, arginine or lysine). A PTD may be able to enhance the translocation of a polypeptide into a cell that lacks a cell wall or a cell from a particular species, *e.g.*, a mammalian cell, such as a human, simian, murine, bovine, equine, feline, or ovine cell.

[00168] A PTD can be linked to an artificial transcription factor, for example, using a flexible linker. Flexible linkers can include one or more glycine residues to allow for free rotation. For example, the PTD can be spaced from a DNA binding domain of the transcription factor by at least 10, 20, or 50 amino acids. A PTD can be located N- or C-terminal relative to a DNA binding domain. Being located N- or C-terminal to a particular  
25 domain does not require being adjacent to that particular domain. For example, a PTD N-terminal to a DNA binding domain can be separated from the DNA binding domain by a spacer and/or other types of domains. A PTD can be chemically synthesized then conjugated chemically to separately prepared DNA binding domain with or without linker peptide. An artificial transcription factor can also include a plurality of PTD's, *e.g.*, a plurality of different  
30 PTD's or at least two copies of one PTD.

[00169] Several proteins and small peptides have the ability to transduce or travel through biological membranes independent of classical receptor- or endocytosis-

mediated pathways. Examples of these proteins include the HIV-1 TAT protein, the herpes simplex virus 1 (HSV-1) DNA-binding protein VP22, and the Drosophila Antennapedia (Antp) homeotic transcription factor. The small protein transduction domains (PTDs) from these proteins can be fused to other macromolecules, peptides or proteins to successfully transport them into a cell. Sequence alignments of the transduction domains from these proteins show a high basic amino acid content (Lys and Arg) which may facilitate interaction of these regions with negatively charged lipids in the membrane. Secondary structure analyses show no consistent structure between all three domains.

[00170] The advantages of using fusions of these transduction domains is that protein entry is rapid, concentration-dependent and appears to work with difficult cell types.

[00171] The Tat protein from human immunodeficiency virus type I (HIV-1) has the remarkable capacity to enter cells when added exogenously (Frankel and Pabo, 1988; Mann and Frankel, 1991; Fawell *et al.*, 1994). A particular example of Tat PTD may include residues 47-57 of the human immunodeficiency virus Tat protein: YGRKKRRQRRR (SEQ ID NO:2). This peptide sequence is referred to as "TAT" herein. This peptide has been shown to successfully mediate the introduction of heterologous peptides and proteins in excess of 100 kDa into mammalian cells *in vitro* and *in vivo* (Ho *et al.*, 2001). Schwarze *et al.* showed that when the 120 kDa  $\beta$ -galactosidase protein fused with TAT was injected into mouse intraperitoneally, the fusion proteins were found in all types of cells and tissues even including brain, which has been thought to be difficult because of the blood-brain-barrier (Schwarze *et al.*, 1999).

[00172] The antennapedia homeodomain also includes a peptide that is a PTD (Derossi *et al.*, 1994). This peptide, also referred to as "Penetratin", includes the amino acid sequence: AKIWFQNR RMKWK KENN (SEQ ID NO:3).

[00173] The HSV VP22 protein also includes a PTD. This PTD is located at the VP22 C-terminal 34 amino acid residues: DAATATRGRSAASRPTERPRAPARSASRPRRPVE\_(SEQ ID NO:4). See, *e.g.*, Elliott and O'Hare (1997) and U.S. Pat. No. 6,184,038.

[00174] In one embodiment, the PTD is obtained from a human or other mammalian protein. Exemplary mammalian PTD's are described in WO 03/059940 (human SIM-2) and WO 03/059941 (Mph). In certain embodiments, the PTD could be a synthetic



PTD. The minimal Tat PTD (aa 47-57) was modified to optimize protein transduction potential (Ho *et al.*, 2001). A FITC coupled with series of synthetic PTD's was tested with cultured T lymphocytes. Some synthetic PTD's showed enhanced protein transduction compared to Tat PTD. These PTD include: YARKARRQARR\_(SEQ ID NO:5);  
5 YARAARRAARR (SEQ ID NO:6); YARAARRAARA (SEQ ID NO:7); YARAAARQARA (SEQ ID NO:8). Especially, the FITC conjugated with synthetic PTD YARAAARQARA (SEQ ID NO:8); showed enhanced uptake by whole blood cells when the mice were i.p. injected.

[00175] The poly-arginine peptides composed of about 6-12 arginine residues  
10 also can mediate protein transduction in some cases. For additional information about poly-arginine, see, *e.g.*, Rothbard *et al.* (2000); Wender *et al.* (2000).

[00176] For additional information about PTD's, see also U.S. 2003/0082561; U.S. 2002/0102265; U.S. 2003/0040038; Schwarze *et al.* (1999); Derossi *et al.* (1996); Hancock *et al.* (1991); Buss *et al.* (1988); Derossi *et al.* (1998); Lindgren *et al.* (2000); Kilic  
15 *et al.* (2003); Asoh *et al.* (2002); and Tanaka *et al.* (2003).

[00177] In addition to PTD's, cellular uptake signals can be used. Such signals include amino acid sequences which are specifically recognized by cellular receptors or other surface proteins. Interaction between the cellular uptake signal and the cell cause internalization of the artificial transcription factor that includes the cellular uptake signal.  
20 Some PTD's may also function by interaction with cellular receptors or other surface proteins.

[00178] A number of assays are available to determine if an amino acid sequence can function as a PTD. For example, the amino acid sequence can be fused to a reporter protein such as  $\beta$ -galactosidase to form a fusion protein. This fusion protein is  
25 contacted with culture cells. The cells are washed and then assayed for reporter activity. Another assay detects the presence of a fusion protein that includes the amino acid sequence in question and another detectable sequence, *e.g.*, an epitope tag. This fusion protein is contacted with culture cells. The cells are washed and then analyzed by Western or immunofluorescence to detect presence of the detectable sequence in cells. Still other assays  
30 can be used to detect transcriptional regulatory activity of a fusion protein that includes the putative PTD, a DNA binding domain, and optionally an effector domain. For example, cells

contacted with such fusion proteins can be assayed for the presence or level of mRNA or protein, *e.g.*, using microarrays, mass spectroscopy, and high-throughput techniques.

## V. Cell culturing

[00179] Generally, cells of the present invention are cultured in a culture medium, which is a nutrient-rich buffered solution capable of sustaining cell growth.

[00180] Culture media suitable for isolating, expanding and differentiating stem cells into hepatocytes according to the method described herein include but not limited to high glucose Dulbecco's Modified Eagle's Medium (DMEM), DMEM/F-15, Liebovitz L-15, RPMI 1640, Iscove's modified Dulbecco's media (IMDM), and Opti-MEM SFM (Invitrogen Inc.). Chemically Defined Medium comprises a minimum essential medium such as Iscove's Modified Dulbecco's Medium (IMDM) (Gibco), supplemented with human serum albumin, human Ex Cyte lipoprotein, transferrin, insulin, vitamins, essential and non essential amino acids, sodium pyruvate, glutamine and a mitogen is also suitable. As used herein, a mitogen refers to an agent that stimulates cell division of a cell. An agent can be a chemical, usually some form of a protein that encourages a cell to commence cell division, triggering mitosis. In one embodiment, serum free media such as those described in U.S. Ser. No. 08/464,599 and WO96/39487, and the "complete media" as described in U.S. Pat. No. 5,486,359 are contemplated for use with the method described herein. In some embodiments, the culture medium is supplemented with 10% Fetal Bovine Serum (FBS), human autologous serum, human AB serum or platelet rich plasma supplemented with heparin (2U/ml). Cell cultures may be maintained in a CO<sub>2</sub> atmosphere, *e.g.*, 5% to 12%, to maintain pH of the culture fluid, incubated at 37°C in a humid atmosphere and passaged to maintain a confluence below 85%.

[00181] Pluripotent stem cells to be differentiated into hepatocytes may be cultured in a medium sufficient to maintain the pluripotency. Culturing of induced pluripotent stem (iPS) cells generated in certain aspects of this invention can use various medium and techniques developed to culture primate pluripotent stem cells, more specially, embryonic stem cells, as described in U.S. Pat. App. 20070238170 and U.S. Pat. App. 20030211603. For example, like human embryonic stem (hES) cells, iPS cells can be maintained in 80% DMEM (Gibco #10829-018 or #11965-092), 20% defined fetal bovine serum (FBS) not heat inactivated, 1% non-essential amino acids, 1 mM L-glutamine, and 0.1 mM .beta.-mercaptoethanol. Alternatively, ES cells can be maintained in serum-free medium, made with 80% Knock-Out DMEM (Gibco #10829-018), 20% serum replacement (Gibco #10828-028),

1% non-essential amino acids, 1 mM L-glutamine, and 0.1 mM .beta.-mercaptoethanol. Just before use, human bFGF may be added to a final concentration of .about 4 ng/mL (WO 99/20741).

5 [00182] Hepatocytes of this invention can be made by culturing pluripotent stem cells or other non-hepatocytes in a medium under conditions that increase the intracellular level of hepatocyte programming factors to be sufficient to promote programming of the cells into hepatocytes. The medium may also contain one or more hepatocyte differentiation and maturation agents, like various kinds of growth factors. However, by increasing the intracellular level of hepatocyte programming transcription factors, aspects of the present  
10 invention bypass most stages toward mature hepatocytes without the need to change the medium for each of the stages. Therefore, in view of the advantages provided by the present invention, in particular aspects, the medium for culturing cells under hepatocyte programming may be essentially free of one or more of the hepatocyte differentiation and maturation agents, or may not undergo serial change with media containing different  
15 combination of such agents.

[00183] These agents may either help induce cells to commit to a more mature phenotype—or preferentially promote survival of the mature cells—or have a combination of both these effects. Hepatocyte differentiation and maturation agents illustrated in this disclosure may include soluble growth factors (peptide hormones, cytokines, ligand-receptor  
20 complexes, and other compounds) that are capable of promoting the growth of cells of the hepatocyte lineage. Non-limiting examples of such agents include but are not limited to epidermal growth factor (EGF), insulin, TGF- $\alpha$ , TGF- $\beta$ , fibroblast growth factor (FGF), heparin, hepatocyte growth factor (HGF), Oncostatin M (OSM), IL-1, IL-6, insulin-like growth factors I and II (IGF-I, IGF-2), heparin binding growth factor 1 (HBGF-1), and  
25 glucagon. The skilled reader will already appreciate that Oncostatin M is structurally related to Leukemia inhibitory factor (LIF), Interleukin-6 (IL-6), and ciliary neurotrophic factor (CNTF).

[00184] An additional examples is n-butyrate, as described in previous patent disclosures (U.S. Pat. No. 6,458,589, U.S. Pat. No. 6,506,574; WO 01/81549). Homologs of  
30 n-butyrate can readily be identified that have a similar effect, and can be used as substitutes in the practice of this invention. Some homologs have similar structural and physicochemical properties to those of n-butyrate: acidic hydrocarbons comprising 3-10 carbon atoms, and a

conjugate base selected from the group consisting of a carboxylate, a sulfonate, a phosphonate, and other proton donors. Examples include isobutyric acid, butenoic acid, propanoic acid, other short-chain fatty acids, and dimethylbutyrate. Also included are isoteric hydrocarbon sulfonates or phosphonates, such as propanesulfonic acid and  
5 propanephosphonic acid, and conjugates such as amides, saccharides, piperazine and cyclic derivatives. A further class of butyrate homologs is inhibitors of histone deacetylase. Non-limiting examples include trichostatin A, 5-azacytidine, trapoxin A, oxamflatin, FR901228, cisplatin, and MS-27-275. Another class of agents is organic solvents like DMSO. Alternatives with similar properties include but are not limited to dimethylacetamide (DMA),  
10 hexmethylene bisacetamide, and other polymethylene bisacetamides. Solvents in this class are related, in part, by the property of increasing membrane permeability of cells. Also of interest are solutes such as nicotinamide.

## **VI. Hepatocytes characteristics**

[00185] Cells can be characterized according to a number of phenotypic  
15 criteria. The criteria include but are not limited to the detection or quantitation of expressed cell markers, enzymatic activity, and the characterization of morphological features and intercellular signaling. In other aspects, cells to be programmed may comprise reporter gene expression cassette comprising tissue- or cell-specific transcriptional regulatory element, like hepatocyte-specific promoters for hepatocyte identification.

[00186] Hepatocytes embodied in certain aspects of this invention have  
20 morphological features characteristic of hepatocytes in the nature, such as primary hepatocytes from organ sources. The features are readily appreciated by those skilled in evaluating such things, and include any or all of the following: a polygonal cell shape, a binucleate phenotype, the presence of rough endoplasmic reticulum for synthesis of secreted  
25 protein, the presence of Golgi-endoplasmic reticulum lysosome complex for intracellular protein sorting, the presence of peroxisomes and glycogen granules, relatively abundant mitochondria, and the ability to form tight intercellular junctions resulting in creation of bile canalicular spaces. A number of these features present in a single cell are consistent with the cell being a member of the hepatocyte lineage. Unbiased determination of whether cells have  
30 morphologic features characteristic of hepatocytes can be made by coding micrographs of programming progeny cells, adult or fetal hepatocytes, and one or more negative control cells, such as a fibroblast, or RPE (Retinal pigment epithelial) cells—then evaluating the

micrographs in a blinded fashion, and breaking the code to determine if the cells produced from forward programming are accurately identified.

[00187] Cells of this invention can also be characterized according to whether they express phenotypic markers characteristic of cells of the hepatocyte lineage. Non-limiting examples of cell markers useful in distinguishing hepatocytes include albumin, 5 asialoglycoprotein receptor,  $\alpha$ 1-antitrypsin,  $\alpha$ -fetoprotein, apoE, arginase I, apoAI, apoAII, apoB, apoCIII, apoCII, aldolase B, alcohol dehydrogenase 1, catalase, CYP3A4, glucokinase, glucose-6-phosphatase, insulin growth factors 1 and 2, IGF-1 receptor, insulin receptor, leptin, liver-specific organic anion transporter (LST-1), L-type fatty acid binding protein, 10 phenylalanine hydroxylase, transferrin, retinol binding protein, and erythropoietin (EPO). Mature hepatocyte markers include, but are limited to, albumin,  $\alpha$ 1-antitrypsin, asialoglycoprotein receptor, cytokeratin 8 (CK8), cytokeratin 18 (CK18), CYP3A4, fumaryl acetoacetate hydrolase (FAH), glucose-6-phosphates, tyrosine aminotransferase, phosphoenolpyruvate carboxykinase, and tryptophan 2,3-dioxygenase.

15 [00188] Assessment of the level of expression of such markers can be determined in comparison with other cells. Positive controls for the markers of mature hepatocytes include adult hepatocytes of the species of interest, and established hepatocyte cell lines. The reader is cautioned that permanent cell lines or long-term liver cell cultures may be metabolically altered, and fail to express certain characteristics of primary 20 hepatocytes. Negative controls include cells of a separate lineage, such as an adult fibroblast cell line, or retinal pigment epithelial (RPE) cells. Undifferentiated stem cells are positive for some of the markers listed above, but negative for markers of mature hepatocytes, as illustrated in the examples below.

[00189] Tissue-specific (*e.g.*, hepatocyte-specific) protein and oligosaccharide 25 determinants listed in this disclosure can be detected using any suitable immunological technique—such as flow immunocytochemistry for cell-surface markers, immunohistochemistry (for example, of fixed cells or tissue sections) for intracellular or cell-surface markers, Western blot analysis of cellular extracts, and enzyme-linked immunoassay, for cellular extracts or products secreted into the medium. Expression of an antigen by a cell 30 is said to be “antibody-detectable” if a significantly detectable amount of antibody will bind to the antigen in a standard immunocytochemistry or flow cytometry assay, optionally after

fixation of the cells, and optionally using a labeled secondary antibody or other conjugate (such as a biotin-avidin conjugate) to amplify labeling.

[00190] The expression of tissue-specific (*e.g.*, hepatocyte-specific) markers can also be detected at the mRNA level by Northern blot analysis, dot-blot hybridization  
5 analysis, or by real time polymerase chain reaction (RT-PCR) using sequence-specific primers in standard amplification methods (U.S. Pat. No. 5,843,780). Sequence data for the particular markers listed in this disclosure can be obtained from public databases such as GenBank. Expression at the mRNA level is said to be “detectable” according to one of the assays described in this disclosure if the performance of the assay on cell samples according  
10 to standard procedures in a typical controlled experiment results in clearly discernable hybridization or amplification product within a standard time window. Unless otherwise required, expression of a particular marker is indicated if the corresponding mRNA is detectable by RT-PCR. Expression of tissue-specific markers as detected at the protein or mRNA level is considered positive if the level is at least 2-fold, and preferably more than 10-  
15 or 50-fold above that of a control cell, such as an undifferentiated pluripotent stem cell, a fibroblast, or other unrelated cell type.

[00191] Cells can also be characterized according to whether they display enzymatic activity that is characteristic of cells of the hepatocyte lineage. For example, assays for glucose-6-phosphatase activity are described by Bublitz (1991); Yasmineh *et al.*  
20 (1992); and Ockerman (1968). Assays for alkaline phosphatase (ALP) and 5'-nucleotidase (5'-Nase) in liver cells are described by Shiojiri (1981). A number of laboratories that serve the research and health care sectors provide assays for liver enzymes as a commercial service.

[00192] In other embodiments, cells of the invention are assayed for activity indicative of xenobiotic detoxification. Cytochrome p450 is a key catalytic component of the  
25 mono-oxygenase system. It constitutes a family of hemoproteins responsible for the oxidative metabolism of xenobiotics (administered drugs), and many endogenous compounds. Different cytochromes present characteristic and overlapping substrate specificity. Most of the biotransforming ability is attributable by the cytochromes designated 1A2, 2A6, 2B6, 3A4, 2C 9 -11, 2D6, and 2E1 (Gomes-Lechon *et al.*, 1997).

[00193] A number of assays are known in the art for measuring xenobiotic  
30 detoxification by cytochrome p450 enzyme activity. Detoxification by CYP3 A4 is

demonstrated using the P450-Glo™ CYP3A4 DMSO-tolerance assay (Luciferin-PPXE) and the P450-Glo™ CYP3A4 cell-based/biochemical assay (Luciferin-PFBE) (Promega Inc, # V8911 and # V8901). Detoxification by CYP1A1 and or CYP1B1 is demonstrated using the P450-Glo™ assay (Luciferin-CEE) (Promega Inc., # V8762) . Detoxification by CYP1A2 and or CYP4A is demonstrated using the P450-Glo™ assay (Luciferin-ME) (Promega Inc., # V8772) Detoxification by CYP2C9 is demonstrated using the P450-Glo™ CYP2C9 assay (Luciferin-H) (Promega Inc., # V8791).

[00194] In another aspect, the biological function of a hepatocyte cell provided by programming is evaluated, for example, by analysing glycogen storage. Glycogen storage is characterized by assaying Periodic Acid Schiff (PAS) functional staining for glycogen granules. The hepatocyte-like cells are first oxidized by periodic acid. The oxidative process results in the formation of aldehyde groupings through carbon-to- carbon bond cleavage. Free hydroxyl groups should be present for oxidation to take place. Oxidation is completed when it reaches the aldehyde stage. The aldehyde groups are detected by the Schiff reagent. A colorless, unstable dialdehyde compound is formed and then transformed to the colored final product by restoration of the quinoid chromophoric grouping (Thompson, 1966; Sheehan and Hrapchak, 1987). PAS staining can be performed according the protocol described at [http://www.jhu.edu/~iic/PDF\\_protocols/LM/Glycogen\\_Staining.pdf](http://www.jhu.edu/~iic/PDF_protocols/LM/Glycogen_Staining.pdf) and <http://library.med.utah.edu/WebPath/HISTHTML/MANUALS/PAS.PDF> with some modifications for an *in vitro* culture of hepatocyte-like cells. One of ordinary skill in the art should be able to make the appropriate modifications.

[00195] In another aspect, a hepatocyte cell produced by forward programming in certain aspects of the invention is characterized for urea production. Urea production can be assayed colorimetrically using kits from Sigma Diagnostic (Miyoshi *et al*, 1998) based on the biochemical reaction of urease reduction to urea and ammonia and the subsequent reaction with 2-oxoglutarate to form glutamate and NAD.

[00196] In another aspect, bile secretion is analyzed. Biliary secretion can be determined by fluorescein diacetate time lapse assay. Briefly, monolayer cultures of hepatocyte-like cells are rinsed with phosphate buffered saline (PBS) three times and incubated with serum- free hepatocyte growth media supplemented with doxycycline and fluorescein diacetate (20 µg/ml) (Sigma-Aldrich) at 37°C for 35 minutes. The cells are washed with PBS three times and fluorescence imaging is carried out. Fluorescein diacetate is

a non fluorescent precursor of fluorescein. The image is evaluated to determine that the compound had been taken up and metabolized in the hepatocyte-like cell to fluorescein. In some embodiments, the compound is secreted into intercellular clefts of the monolayer of cells. Alternatively, bile secretion is determined by a method using sodium fluorescein  
5 described by Gebhart and Wang (1982).

[00197] In yet another aspect, lipid synthesis is analyzed. Lipid synthesis in the hepatocyte-like cell can be determined by oil red O staining. Oil Red O (Solvent Red 27, Sudan Red 5B, C.I. 26125, C<sub>26</sub>H<sub>24</sub>N<sub>4</sub>O) is a lysochrome (fat-soluble dye) diazo dye used for staining of neutral triglycerides and lipids on frozen sections and some lipoproteins on  
10 paraffin sections. It has the appearance of a red powder with maximum absorption at 518(359) nm. Oil Red O is one of the dyes used for Sudan staining. Similar dyes include Sudan III, Sudan IV, and Sudan Black B. The staining has to be performed on fresh samples and/or formalin fixed samples. Hepatocyte-like cells are cultured on microscope slides, rinsed in PBS three times, the slides are air dried for 30-60 minutes at room temperature, fixed in ice  
15 cold 10% formalin for 5-10 minutes, and then rinse immediately in 3 changes of distilled water. The slide is then placed in absolute propylene glycol for 2-5 minutes to avoid carrying water into Oil Red O and stained in pre-warmed Oil Red O solution for 8 minutes in 600 °C oven. The slide is then placed in 85% propylene glycol solution for 2-5 minutes and rinsed in  
20 2 changes of distilled water. Oil red O staining can also be performed according the protocol described at [library.med.utah.edu/WebPath/HISTHTML/MANUALS/OILRED.PDF](http://library.med.utah.edu/WebPath/HISTHTML/MANUALS/OILRED.PDF) with some modifications for an *in vitro* culture of hepatocyte-like cell by one of ordinary skill in the art.

[00198] In still another aspect, the cells are assayed for glycogen synthesis. Glycogen assays are well known to one of ordinary skill in the art, for example, in  
25 Passonneau and Lauderdale (1974). Alternatively, commercial glycogen assays can be used, for example, from BioVision, Inc. catalog # K646-100.

[00199] Cells of the hepatocyte lineage can also be evaluated by their ability to store glycogen. A suitable assay uses Periodic Acid Schiff (PAS) stain, which does not react with mono- and disaccharides, but stains long-chain polymers such as glycogen and dextran.  
30 PAS reaction provides quantitative estimations of complex carbohydrates as well as soluble and membrane-bound carbohydrate compounds. Kirkeby *et al.* (1992) describe a quantitative PAS assay of carbohydrate compounds and detergents. van der Laarse *et al.* (1992) describe a



microdensitometric histochemical assay for glycogen using the PAS reaction. Evidence of glycogen storage is determined if the cells are PAS-positive at a level that is at least 2-fold, and preferably more than 10-fold above that of a control cell, such as a fibroblast. The cells can also be characterized by karyotyping according to standard methods.

5           **[00200]**        Assays are also available for enzymes involved in the conjugation, metabolism, or detoxification of small molecule drugs. For example, cells can be characterized by an ability to conjugate bilirubin, bile acids, and small molecule drugs, for excretion through the urinary or biliary tract. Cells are contacted with a suitable substrate, incubated for a suitable period, and then the medium is analyzed (by GCMS or other suitable  
10            technique) to determine whether a conjugation product has been formed. Drug metabolizing enzyme activities include de-ethylation, dealkylation, hydroxylation, demethylation, oxidation, glucuroconjugation, sulfoconjugation, glutathione conjugation, and N-acetyl transferase activity (A. Guillouzo, pp 411-431 in *In vitro* Methods in Pharmaceutical Research, Academic Press, 1997). Assays include penacetin de-ethylation, procainamide N-  
15            acetylation, paracetamol sulfoconjugation, and paracetamol glucuronidation (Chesne *et al.*, 1988).

**[00201]**        A further feature of certain cell populations of this invention is that they are susceptible under appropriate circumstances to pathogenic agents that are tropic for primate liver cells. Such agents include hepatitis A, B, C, and delta, Epstein-Barr virus  
20            (EBV), cytomegalovirus (CMV), tuberculosis, and malaria. For example, infectivity by hepatitis B can be determined by combining cultured forward programming-derived hepatocytes with a source of infectious hepatitis B particles (such as serum from a human HBV carrier). The liver cells can then be tested for synthesis of viral core antigen (HBcAg) by immunohistochemistry or RT-PCR.

25            **[00202]**        The skilled reader will readily appreciate that an advantage of forward programming-derived hepatocytes is that they will be essentially free of other cell types that typically contaminate primary hepatocyte cultures isolated from adult or fetal liver tissue. Markers characteristic of sinusoidal endothelial cells include Von Willebrand factor, CD4, CD14, and CD32. Markers characteristic of bile duct epithelial cells include cytokeratin-7,  
30            cytokeratin-19, and  $\gamma$ -glutamyl transpeptidase. Markers characteristic of stellate cells include  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA), vimentin, synaptophysin, glial fibrillary acidic protein (GFAP), neural-cell adhesion molecule (N-CAM), and presence of lipid droplets (detectable

by autofluorescence or staining by oil red O). Markers characteristic of Kupffer cells include CD68, certain lectins, and markers for cells of the macrophage lineage (such as HLA Class II, and mediators of phagocytosis). Forward programming-derived hepatocytes can be characterized as essentially free of some or all of these cell types if less than 0.1% (preferably  
5 less than 100 or 10 ppm) bear markers or other features of the undesired cell type, as determined by immunostaining and fluorescence-activated quantitation, or other appropriate technique.

[00203] Hepatocytes provided by forward programming according to certain aspects of this invention can have a number of the features of the stage of cell they are  
10 intended to represent. The more of these features that are present in a particular cell, the more it can be characterized as a cell of the hepatocyte lineage. Cells having at least 2, 3, 5, 7, or 9 of these features are increasingly more preferred. In reference to a particular cell population as may be present in a culture vessel or a preparation for administration, uniformity between cells in the expression of these features is often advantageous. In this circumstance,  
15 populations in which at least about 40%, 60%, 80%, 90%, 95%, or 98% of the cells have the desired features are increasingly more preferred.

[00204] Other desirable features of hepatocytes provided in certain aspects of this invention are an ability to act as target cells in drug screening assays, and an ability to reconstitute liver function, both *in vivo*, and as part of an extracorporeal device. These  
20 features are described further in sections that follow.

## VII. Use of hepatocytes

[00205] The hepatocytes provided by methods and compositions of certain aspects of the invention can be used in a variety of applications. These include but not limited to transplantation or implantation of the hepatocytes *in vivo*; screening cytotoxic compounds,  
25 carcinogens, mutagens growth/regulatory factors, pharmaceutical compounds, *etc.*, *in vitro*; elucidating the mechanism of liver diseases and infections; studying the mechanism by which drugs and/or growth factors operate; diagnosing and monitoring cancer in a patient; gene therapy; and the production of biologically active products, to name but a few.

### A. Test compound screening

[00206] Forward programming-derived hepatocytes of this invention can be  
30 used to screen for factors (such as solvents, small molecule drugs, peptides, and

polynucleotides) or environmental conditions (such as culture conditions or manipulation) that affect the characteristics of hepatocytes provided herein.

[00207] In some applications, stem cells (differentiated or undifferentiated) are used to screen factors that promote maturation of cells along the hepatocyte lineage, or promote proliferation and maintenance of such cells in long-term culture. For example, candidate hepatocyte maturation factors or growth factors are tested by adding them to stem cells in different wells, and then determining any phenotypic change that results, according to desirable criteria for further culture and use of the cells.

[00208] Particular screening applications of this invention relate to the testing of pharmaceutical compounds in drug research. The reader is referred generally to the standard textbook *In vitro Methods in Pharmaceutical Research*, Academic Press, 1997, and U.S. Pat. No. 5,030,015). In certain aspects of this invention, cell programmed to the hepatocyte lineage play the role of test cells for standard drug screening and toxicity assays, as have been previously performed on hepatocyte cell lines or primary hepatocytes in short-term culture. Assessment of the activity of candidate pharmaceutical compounds generally involves combining the hepatocytes provided in certain aspects of this invention with the candidate compound, determining any change in the morphology, marker phenotype, or metabolic activity of the cells that is attributable to the compound (compared with untreated cells or cells treated with an inert compound), and then correlating the effect of the compound with the observed change. The screening may be done either because the compound is designed to have a pharmacological effect on liver cells, or because a compound designed to have effects elsewhere may have unintended hepatic side effects. Two or more drugs can be tested in combination (by combining with the cells either simultaneously or sequentially), to detect possible drug-drug interaction effects.

[00209] In some applications, compounds are screened initially for potential hepatotoxicity (Castell *et al.*, 1997). Cytotoxicity can be determined in the first instance by the effect on cell viability, survival, morphology, and leakage of enzymes into the culture medium. More detailed analysis is conducted to determine whether compounds affect cell function (such as gluconeogenesis, ureogenesis, and plasma protein synthesis) without causing toxicity. Lactate dehydrogenase (LDH) is a good marker because the hepatic isoenzyme (type V) is stable in culture conditions, allowing reproducible measurements in culture supernatants after 12-24 h incubation. Leakage of enzymes such as mitochondrial

glutamate oxaloacetate transaminase and glutamate pyruvate transaminase can also be used. Gomez-Lechon *et al.* (1996) describes a microassay for measuring glycogen, which can be used to measure the effect of pharmaceutical compounds on hepatocyte gluconeogenesis.

[00210] Other current methods to evaluate hepatotoxicity include determination  
5 of the synthesis and secretion of albumin, cholesterol, and lipoproteins; transport of  
conjugated bile acids and bilirubin; ureagenesis; cytochrome p450 levels and activities;  
glutathione levels; release of  $\alpha$ -glutathione s-transferase; ATP, ADP, and AMP metabolism;  
intracellular  $K^+$  and  $Ca^{2+}$  concentrations; the release of nuclear matrix proteins or  
oligonucleosomes; and induction of apoptosis (indicated by cell rounding, condensation of  
10 chromatin, and nuclear fragmentation). DNA synthesis can be measured as [ $^3H$ ]-thymidine or  
BrdU incorporation. Effects of a drug on DNA synthesis or structure can be determined by  
measuring DNA synthesis or repair. [ $^3H$ ]-thymidine or BrdU incorporation, especially at  
unscheduled times in the cell cycle, or above the level required for cell replication, is  
consistent with a drug effect. Unwanted effects can also include unusual rates of sister  
15 chromatid exchange, determined by metaphase spread. The reader is referred to Vickers  
(1997) for further elaboration.

#### **B. Liver therapy and transplantation**

[00211] This invention also provides for the use of hepatocytes provided herein  
to restore a degree of liver function to a subject needing such therapy, perhaps due to an  
20 acute, chronic, or inherited impairment of liver function.

[00212] To determine the suitability of hepatocytes provided herein for  
therapeutic applications, the cells can first be tested in a suitable animal model. At one level,  
cells are assessed for their ability to survive and maintain their phenotype *in vivo*.  
Hepatocytes provided herein are administered to immunodeficient animals (such as SCID  
25 mice, or animals rendered immunodeficient chemically or by irradiation) at a site amenable  
for further observation, such as under the kidney capsule, into the spleen, or into a liver  
lobule. Tissues are harvested after a period of a few days to several weeks or more, and  
assessed as to whether starting cell types such as pluripotent stem cells are still present. This  
can be performed by providing the administered cells with a detectable label (such as green  
30 fluorescent protein, or  $\beta$ -galactosidase); or by measuring a constitutive marker specific for the  
administered cells. Where hepatocytes provided herein are being tested in a rodent model, the  
presence and phenotype of the administered cells can be assessed by immunohistochemistry

or ELISA using human-specific antibody, or by RT-PCR analysis using primers and hybridization conditions that cause amplification to be specific for human polynucleotide sequences. Suitable markers for assessing gene expression at the mRNA or protein level are provided in elsewhere in this disclosure. General descriptions for determining the fate of hepatocyte-like cells in animal models is provided in Grompe *et al.* (1999); Peeters *et al.*, (1997); and Ohashi *et al.* (2000).

[00213] At another level, hepatocytes provided herein are assessed for their ability to restore liver function in an animal lacking full liver function. Braun *et al.* (2000) outline a model for toxin-induced liver disease in mice transgenic for the HSV-tk gene. Rhim *et al.* (1995) and Lieber *et al.* (1995) outline models for liver disease by expression of urokinase. Mignon *et al.* (1998) outline liver disease induced by antibody to the cell-surface marker Fas. Overturf *et al.* (1998) have developed a model for Hereditary Tyrosinemia Type I in mice by targeted disruption of the *Fah* gene. The animals can be rescued from the deficiency by providing a supply of 2-(2-nitro-4-fluoro-methyl-benzyol)-1,3-cyclohexanedione (NTBC), but they develop liver disease when NTBC is withdrawn. Acute liver disease can be modeled by 90% hepatectomy (Kobayashi *et al.*, 2000). Acute liver disease can also be modeled by treating animals with a hepatotoxin such as galactosamine, CCl<sub>4</sub>, or thioacetamide.

[00214] Chronic liver diseases such as cirrhosis can be modeled by treating animals with a sub-lethal dose of a hepatotoxin long enough to induce fibrosis (Rudolph *et al.*, 2000). Assessing the ability of hepatocytes provided herein to reconstitute liver function involves administering the cells to such animals, and then determining survival over a 1 to 8 week period or more, while monitoring the animals for progress of the condition. Effects on hepatic function can be determined by evaluating markers expressed in liver tissue, cytochrome p450 activity, and blood indicators, such as alkaline phosphatase activity, bilirubin conjugation, and prothrombin time), and survival of the host. Any improvement in survival, disease progression, or maintenance of hepatic function according to any of these criteria relates to effectiveness of the therapy, and can lead to further optimization.

[00215] Hepatocytes provided in certain aspects of this invention that demonstrate desirable functional characteristics according to their profile of metabolic enzymes, or efficacy in animal models, may also be suitable for direct administration to human subjects with impaired liver function. For purposes of hemostasis, the cells can be

administered at any site that has adequate access to the circulation, typically within the abdominal cavity. For some metabolic and detoxification functions, it is advantageous for the cells to have access to the biliary tract. Accordingly, the cells are administered near the liver (e.g., in the treatment of chronic liver disease) or the spleen (e.g., in the treatment of fulminant hepatic failure). In one method, the cells administered into the hepatic circulation either through the hepatic artery, or through the portal vein, by infusion through an indwelling catheter. A catheter in the portal vein can be manipulated so that the cells flow principally into the spleen, or the liver, or a combination of both. In another method, the cells are administered by placing a bolus in a cavity near the target organ, typically in an excipient or matrix that will keep the bolus in place. In another method, the cells are injected directly into a lobe of the liver or the spleen.

**[00216]** The hepatocytes provided in certain aspects of this invention can be used for therapy of any subject in need of having hepatic function restored or supplemented. Human conditions that may be appropriate for such therapy include fulminant hepatic failure due to any cause, viral hepatitis, drug-induced liver injury, cirrhosis, inherited hepatic insufficiency (such as Wilson's disease, Gilbert's syndrome, or  $\alpha_1$ -antitrypsin deficiency), hepatobiliary carcinoma, autoimmune liver disease (such as autoimmune chronic hepatitis or primary biliary cirrhosis), and any other condition that results in impaired hepatic function. For human therapy, the dose is generally between about  $10^9$  and  $10^{12}$  cells, and typically between about  $5 \times 10^9$  and  $5 \times 10^{10}$  cells, making adjustments for the body weight of the subject, nature and severity of the affliction, and the replicative capacity of the administered cells. The ultimate responsibility for determining the mode of treatment and the appropriate dose lies with the managing clinician.

### C. Use in a liver assist device

**[00217]** Certain aspects of this invention include hepatocytes provided herein that are encapsulated or part of a bioartificial liver device. Various forms of encapsulation are described in *Cell Encapsulation Technology and Therapeutics*, 1999. Hepatocytes provided in certain aspects of this invention can be encapsulated according to such methods for use either *in vitro* or *in vivo*.

**[00218]** Bioartificial organs for clinical use are designed to support an individual with impaired liver function—either as a part of long-term therapy, or to bridge the time between a fulminant hepatic failure and hepatic reconstitution or liver transplant.

Bioartificial liver devices are reviewed by Macdonald *et al.*, pp. 252-286 of “Cell Encapsulation Technology and Therapeutics”, *op cit.*, and exemplified in U.S. Pat. Nos. 5,290,684, 5,624,840, 5,837,234, 5,853,717, and 5,935,849. Suspension-type bioartificial livers comprise cells suspended in plate dialysers, microencapsulated in a suitable substrate, or attached to microcarrier beads coated with extracellular matrix. Alternatively, hepatocytes can be placed on a solid support in a packed bed, in a multiplate flat bed, on a microchannel screen, or surrounding hollow fiber capillaries. The device has an inlet and outlet through which the subject's blood is passed, and sometimes a separate set of ports for supplying nutrients to the cells.

10           [00219]       Hepatocytes are prepared according to the methods described earlier, and then plated into the device on a suitable substrate, such as a matrix of Matrigel® or collagen. The efficacy of the device can be assessed by comparing the composition of blood in the afferent channel with that in the efferent channel—in terms of metabolites removed from the afferent flow, and newly synthesized proteins in the efferent flow.

15           [00220]       Devices of this kind can be used to detoxify a fluid such as blood, wherein the fluid comes into contact with the hepatocytes provided in certain aspects of this invention under conditions that permit the cell to remove or modify a toxin in the fluid. The detoxification will involve removing or altering at least one ligand, metabolite, or other compound (either natural and synthetic) that is usually processed by the liver. Such compounds include but are not limited to bilirubin, bile acids, urea, heme, lipoprotein, carbohydrates, transferrin, hemopexin, asialoglycoproteins, hormones like insulin and glucagon, and a variety of small molecule drugs. The device can also be used to enrich the efferent fluid with synthesized proteins such as albumin, acute phase reactants, and unloaded carrier proteins. The device can be optimized so that a variety of these functions is performed, thereby restoring as many hepatic functions as are needed. In the context of therapeutic care, the device processes blood flowing from a patient in hepatocyte failure, and then the blood is returned to the patient.

#### **D.       Distribution for Commercial, Therapeutic, and Research Purposes**

30           [00221]       For purposes of manufacture, distribution, and use, the hepatocyte lineage cells of this invention are typically supplied in the form of a cell culture or suspension in an isotonic excipient or culture medium, optionally frozen to facilitate transportation or storage.

[00222] This invention also includes different reagent systems, comprising a set or combination of cells that exist at any time during manufacture, distribution, or use. The cell sets comprise any combination of two or more cell populations described in this disclosure, exemplified but not limited to programming-derived cells (hepatocyte lineage  
5 cells, their precursors and subtypes), in combination with undifferentiated stem cells, somatic cell-derived hepatocytes, or other differentiated cell types. The cell populations in the set sometimes share the same genome or a genetically modified form thereof. Each cell type in the set may be packaged together, or in separate containers in the same facility, or at different locations, at the same or different times, under control of the same entity or different entities  
10 sharing a business relationship.

### VIII. Cells and methods for testing candidate gene in programming

[00223] The ability of a particular candidate gene or a combination of candidate genes to act as forward programming factors for a specific cell type, such as hepatocytes, can be tested using the methods and cells provided in this disclosure. Efficacy of  
15 particular candidate genes or combinations of candidate genes in forward programming can be assessed by their effect on cell morphology, marker expression, enzymatic activity, proliferative capacity, or other features of interest, which is then determined in comparison with parallel cultures that did not include the candidate genes or combinations. Candidate genes may be transcription factors important for differentiation into desired cell types or for  
20 function of the desired cell types.

[00224] In certain embodiments, starting cells, such as pluripotent stem cells, comprising at least one expression cassette for expression of a candidate gene or a combination of candidate genes may be provided. The expression cassette may comprise an externally controllable transcriptional regulatory element, such as an inducible promoter. The  
25 activity of these promoters may be induced by the presence or absence of biotic or abiotic factors. Inducible promoters are a very powerful tool in genetic engineering because the expression of genes operably linked to them can be turned on or off at certain stages of development of an organism or in a particular tissue. Tet-On and Tet-Off inducible gene expression systems based on the essential regulatory components of the *E. coli* tetracycline-resistance operon may be used. Once established in the starting cells, the inducer doxycycline  
30 (Dox, a tetracycline derivative) could controls the expression system in a dose-dependent manner, allowing to precisely modulate the expression levels of candidate genes.



[00225] To aid identification of desired cell types, the starting cells may further comprise a cell-specific or tissue-specific reporter expression cassette. The reporter expression cassette may comprise a reporter gene operably linked to a transcriptional regulatory element specific for the desired cell types. For example, the reporter expression cassette may comprise a hepatocyte-specific promoter for hepatocyte production, isolation, selection, or enrichment. The reporter gene may be any selectable or screenable marker gene known in the art and exemplified in the preceding disclosure.

### VIII. Examples

[00226] The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

#### Example 1 - Forward Programming into hepatocytes

[00227] Alternative approaches for hepatocyte differentiation from human ESC/iPSCs are shown in **FIG. 1**. Hepatic lineage cells such as mature hepatocytes can likely be efficiently induced from human ESC/iPSCs via expression of appropriate transgene combination (top box), bypassing most, if not all, developmental stages required during normal differentiation (bottom box).

[00228] The strategy employed for identifying transgenes that could directly convert human ESC/iPSCs to hepatic lineage cells including mature hepatocytes is shown in **FIG. 2**. Human ESC/iPSCs were engineered to carry reporters under the control of a hepatocyte-specific promoter, and to constitutively express rtTET protein for inducible gene expression. Transgenes under the control of the inducible promoter Ptight will be introduced into the engineered hESC/iPSCs either by lipid-mediated transfection or electroporation. Upon doxycycline (Dox) addition, transgene expression will be induced, and hepatocyte differentiation will be monitored by the expression of reporters and hepatocyte-specific marker genes, and additional hepatocyte-specific function analyses. Once the right transgene

combination is identified, the hepatocytes will be purified for both *in vitro* and *in vivo* functional assays.

[00229] Human ESC/iPSC reporter/inducible (R/I) lines were established for hepatocyte differentiation (**FIG. 3**). The human Rosa26 locus on chromosome 3 was selected to allow the expression of both hepatocyte-specific reporter and rtTET, while minimizing the chromosome location-dependent silencing effect. First, the LoxP recombination sites (LOX71 and LOX2272) were introduced into a site between exon 1 and exon 2 of human ROSA 26 gene via homologous recombination. The targeting construct (KI construct) used the phosphoglycerate kinase promoter (PGK)-driven expression of diphtheria toxin A fragment gene (DTA) for negative selection, and contains a ~ 2.0 kb 5' arm and a 4.5 kb 3' arm. A splicing acceptor signal from human BCL2 gene (SA) was placed in front of LOX71 site to allow the expression of selection markers from the endogenous human ROSA26 promoter. The coding region for thymidine kinase (TK) was included to enable negative selection against incorrect Cre/LoxP recombination events at step 2 using ganciclovir. The neomycin phosphotransferase (Neo) was used for positive selection during homologous recombination (step 1). The foot-and-mouth disease virus peptide (F2A) was used to co-express the TK and Neo genes from the endogenous human ROSA26 promoter. BGHpA is polyadenylation signal derived from bovine growth hormone gene. The homologous recombination yielded parental human ESC/iPSC lines for efficient cassette exchange via Cre/LoxP recombination. To establish reporter/inducible cell lines for hepatocyte differentiation, F2A peptide linked marker gene mOrange and Blasticidin S deaminase (BSD) (driven by a hepatocyte-specific promoter ApoE4pAAT) and rtTET (driven by the constitutively active eukaryotic elongation factor 1 $\alpha$  promoter – pEF) was introduced into the Rosa 26 locus by lipid-mediated cotransfection of the recombination mediated cassette exchange (RMCE) vector and a Cre-expressing plasmid. The puromycin N-acetyltransferase (Puro) was used to select for recombination events. The correctly recombined R/I cells are resistant to puromycin (Puro<sup>+</sup>) and ganciclovir (TK<sup>-</sup>), and sensitive to geneticin selection (Neo<sup>-</sup>).

[00230] Restricted marker gene (mOrange) expression in hepatocytes during normal human ESC differentiation were confirmed (**FIGS. 4A-4B**). Human H1 ESC R/I lines were routinely maintained in MEF-conditioned human ES cell medium supplemented with 100 ng/ml bFGF (CM100) on matrigel (Growth Factor Reduced; BD Bioscience). For

differentiation, human ESCs were harvested using Accutase (Invitrogen), and plated on matrigel-coated 10-cm dishes at a density of  $0.5 \times 10^5$  cells/cm<sup>2</sup> in CM100 for 3 days. Hepatocyte differentiation was initiated by culture for 5 days with 100ng/ml Activin A (R&D Systems) in RPMI/B27 medium (Invitrogen) (definitive endoderm differentiation), followed by 5 days with 20 ng/ml BMP4 (Peprotech) and 10 ng/ml FGF-2 (Invitrogen) in RPMI/B27 (hepatic specification), then 5 days with 20 ng/ml HGF (Peprotech) in RPMI/B27 (immature hepatocyte differentiation) and finally for 5 days with 20 ng/ml Oncostatin-M (R&D Systems) in Hepatocyte Culture Media (Lonza) supplemented with SingleQuots (hepatocyte maturation).

10           **[00231]**       The Tet-On inducible gene expression was confirmed in human H1 ESC R/I lines (**FIGS. 5A-5C**). The EGFP driven by the Ptight promoter (an rtTET-responsive inducible promoter) was introduced into human ESC R/I lines using Fugene HD-mediated transfection of both vectors in **FIG. 5A**. Human ESCs with stable *PiggyBac* transposon integration were selected with geneticin (100 µg/ml). Images are shown in **FIG.**  
15 **5B** with human ESC R/I lines after 2 days induction with or without Doxycycline (1 µg/ml). EGFP expression was analyzed by flow cytometry in human ESC R/I lines after 4 days induction with or without Doxycycline (1 µg/ml) (**FIG. 5C**). After 4 days of Doxycycline induction, 83.3% human ESC R/I lines showed stable *PiggyBac* transposon integration by EGFP expression.

20           **[00232]**       Hepatocytes were directly induced from human ESC R/I lines through transgene expression (**FIG. 6**). Genes that are either implicated in hepatic differentiation during normal mammalian development or enriched in adult hepatocytes were cloned into the *PiggyBac* vector (**FIG. 5A**) under the control of the Ptight promoter (**Table 1**). These genes were further prioritized based on their known functional importance during normal hepatic  
25 differentiation or hepatic functions. To screen for transcription factors that are able to directly impose hepatic fate upon human ESCs, various combinations of transgene-expressing *PiggyBac* vectors along with the hPBBase-expressing vector were introduced into the human ESC R/I lines cultured in CM100 on matrigel via Fugene HD-mediated transfection or electroporation. Following Geneticin (100 µg/ml) selection for stable  
30 genomic transgene integration, Doxycycline (1 µg/ml) was added to induce transgene expression, and the CM100 was replaced with Hepatocyte Culture Media (Lonza) supplemented with SingleQuots, 20 ng/ml HGF and 50 ng/ml Oncostatin-M (HCM). Hepatic

lineage induction was monitored with mOrange marker gene expression between day 3-5 post induction. In the absence of Doxycycline induction, significant cell death was observed after 3-day culture in HCM medium in contrast to those with Doxycycline induction. The combination of transcription factors used herein are from the following: FOXA1, FOXA2-2, HHEX, HNF1A, HNF4A-2 and TBX3-1 (**Table 1**). Significant number of hepatocyte-specific promoter-driven mOrange-expressing cells were observed after five days of Doxycycline induction in HCM.

**[00233]** Another combination of genes (FOXA2, HHEX, HNF4A, GATA4, NROB2 and SCML1) were identified that are sufficient to convert human ESCs directly into hepatocyte-like cells (**FIGs. 7A-7C**). Briefly, the transgene-expressing PiggyBac vectors (2 µg each) along with the hPBase-expressing vector (4 µg) were introduced into the human ESC R/I lines cultured in mTeSR1 on matrigel via nucleofection. About 2 – 3 x 10<sup>6</sup> ESCs were used for each nucleofection (nucleofection solution: 100 µl of Ingenio® Electroporation solution from Mirus, Madison, WI; program: Amaxa B-016). Following Geneticin (100 µg/ml) selection for stable genomic transgene integration, cells were plated into 12-well matrigel plates for forward programming. The next day following plating, Doxycycline (1 µg/ml) was added to induce transgene expression, initially in mTeSR1 for 1 day followed by Hepatocyte Maintenance Medium supplemented with SingleQuot (HMM, Lonza), 20ng/ml HGF and 20ng/ml Oncostatin-M (OSM) for 5 days. After 6-day transgene induction, cells were further cultured in HMM supplemented with OSM for an additional 10 – 11 days prior to analysis. Hepatic induction was examined by immunological staining with antibodies for hepatocyte-specific markers alpha-fetoprotein (AFP) (**FIG. 7A**), albumin (ALB) (**FIG. 7B**), and asialoglycoprotein receptor 1 (ASGPR1) (**FIG. 7C**). Expression of additional genes (**Table 1**) may improve either the efficiency of hepatic lineage programming from human ESCs or hepatic functions.

**[00234]** Additional combinations that could induce hepatocyte-like cells from human ESC R/I lines via forward programming are presented in **Table 2**.

**Table 2.** Additional transgene combinations for hepatocyte forward programming

#				
C1	FOXA2	HNF1A	HNF4A	CEBPB
C2	FOXA2	HNF1A	HNF4A	FOXA1
C3	FOXA2	HNF1A	HNF4A	GATA4

C4	FOXA2	HNF1A	HNF4A	HHEX
C5	FOXA2	HNF1A	HNF4A	HLF
C6	FOXA2	HNF1A	HNF4A	HLX
C7	FOXA2	HNF1A	HNF4A	NR0B2
C8	FOXA2	HNF1A	HNF4A	NR1H3
C9	FOXA2	HNF1A	HNF4A	NR1H4
C10	FOXA2	HNF1A	HNF4A	NR1I2
C11	FOXA2	HNF1A	HNF4A	NR1I3
C12	FOXA2	HNF1A	HNF4A	NR5A2
C13	FOXA2	HNF1A	HNF4A	SCML1
C14	FOXA2	HNF1A	HNF4A	SEBOX
C15	FOXA2	HNF1A	HNF4A	ZNF391
C16	FOXA2	HNF1A	HNF4A	ZNF517

[00235] Examples of additional combinations for forward programming are shown in FIG. 8. The transgene-expressing *PiggyBac* vectors (2  $\mu\text{g}$  each) along with the hPBase-expressing vector (4  $\mu\text{g}$ ) were introduced into the human ESC R/I lines cultured in mTeSR1 on matrigel via nucleofection. About  $2 - 3 \times 10^6$  ESCs were used for each nucleofection (nucleofection solution: 100  $\mu\text{l}$  of Ingenio® Electroporation solution from Mirus, Madison, WI; program: Amaxa B-016). Following Geneticin (100  $\mu\text{g}/\text{ml}$ ) selection for stable genomic transgene integration, cells were plated into 12-well matrigel plates for forward programming. The next day following plating, Doxycycline (1  $\mu\text{g}/\text{ml}$ ) was added to induce transgene expression, in Hepatocyte Maintenance Medium supplemented with SingleQuot (HMM, Lonza), 20ng/ml HGF and 20ng/ml Oncostatin-M (OSM) for 4 days. After 4-day transgene induction, cells were further cultured in HMM supplemented with OSM for an additional 12 days prior to analysis. Hepatic induction was examined by immunostaining with antibodies for hepatocyte-specific markers alpha-fetoprotein (AFP), albumin (ALB), and asialoglycoprotein receptor 1 (ASGPR1).

\* \* \*

[00236] All of the methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the

agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

**REFERENCES**

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

U.S. Patent 4,683,202

U.S. Patent 5,030,015

U.S. Patent 5,290,684

U.S. Patent 5,302,523

U.S. Patent 5,322,783

U.S. Patent 5,384,253

U.S. Patent 5,460,964

U.S. Patent 5,464,765

U.S. Patent 5,486,359

U.S. Patent 5,538,877

U.S. Patent 5,538,880

U.S. Patent 5,550,318

U.S. Patent 5,563,055

U.S. Patent 5,580,859

U.S. Patent 5,589,466

U.S. Patent 5,591,616

U.S. Patent 5,610,042

U.S. Patent 5,624,840

U.S. Patent 5,635,387

U.S. Patent 5,656,610

U.S. Patent 5,677,136

U.S. Patent 5,681,599

U.S. Patent 5,702,932

U.S. Patent 5,716,827

U.S. Patent 5,736,396

U.S. Patent 5,736,524

U.S. Patent 5,750,397

U.S. Patent 5,759,793

U.S. Patent 5,780,448  
U.S. Patent 5,789,215  
U.S. Patent 5,811,094  
U.S. Patent 5,827,735  
U.S. Patent 5,827,740  
U.S. Patent 5,837,234  
U.S. Patent 5,837,539  
U.S. Patent 5,837,670  
U.S. Patent 5,843,780  
U.S. Patent 5,853,717  
U.S. Patent 5,925,565  
U.S. Patent 5,928,906  
U.S. Patent 5,935,819  
U.S. Patent 5,935,849  
U.S. Patent 5,945,100  
U.S. Patent 5,981,274  
U.S. Patent 5,994,136  
U.S. Patent 5,994,624  
U.S. Patent 6,013,516  
U.S. Patent 6,184,038  
U.S. Patent 6,458,589  
U.S. Patent 6,506,574  
U.S. Patent 6,833,269  
U.S. Patent 6,991,897  
U.S. Patent 7,015,037  
U.S. Patent 7,399,632  
U.S. Patent 7,410,773  
U.S. Patent 7,410,798  
U.S. Patent 7,422,736  
U.S. Patent Publn. 2002/0102265  
U.S. Patent Publn. 2003/0040038  
U.S. Patent Publn. 2003/0082561  
U.S. Patent Publn. 20030211603  
U.S. Patent Publn. 20070238170



- U.S. Appln. Serial 08/464,599
- U.S. Appln. Serial 61/058,858
- U.S. Appln. Serial 61/172,079
- U.S. Appln. Serial 61/184,546
- Alexander *et al.*, *Proc. Nat. Acad. Sci. USA*, 85:5092-5096, 1988.
- Alison *et al.*, *Hepatology*, 29:678-83, 1998.
- Amit *et al.*, *Dev. Bio.*, 227:271-278, 2000.
- Andrews *et al.*, In: *Teratocarcinomas and Embryonic Stem Cells*, Robertson (Ed.), IRL Press, 207-246, 1987.
- Asoh *et al.*, *Proc. Natl. Acad. Sci. USA*, 99(26):17107-12, 2002.
- Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publ. Assoc. Inc. & John Wiley & Sons, Inc., MA, 1996.
- Ausubel *et al.*, In: *Current Protocols in Molecular Biology*, John, Wiley & Sons, Inc, New York, 1994.
- Blomer *et al.*, *J. Virol.*, 71(9):6641-6649, 1997.
- Boyer *et al.*, *Cell*, 122(6):947-56, 2005.
- Braun *et al.*, *Nature Med.*, 6:320, 2000.
- Bublitz, *Mol. Cell Biochem.*, 108:141, 1991.
- Buss *et al.*, *Mol. Cell. Biol.*, 8:3960-3963, 1988.
- Byrne *et al.*, *Nature*, 450(7169):497-502, 2007.
- Capecchi, *Nature*, 348:109, 1990.
- Cassiede *et al.*, *J. Bone Miner. Res.*, 11(9):1264-1273, 1996.
- Castell *et al.*, In: *In vitro Methods in Pharmaceutical Research*, Academic Press, 375-410, 1997.
- Cell Encapsulation Technology and Therapeutics*, Kuhlreiber *et al.* eds., Birkhauser, Boston Mass., 1999.
- Chambers *et al.*, *Cell*, 113(5):643-55, 2003.
- Chen and Okayama, *Mol. Cell Biol.*, 7(8):2745-2752, 1987.
- Chesne *et al.*, In: *Liver Cells and Drugs*, Guillouzo (Ed.), John Libbey Eurotext, London, 343-350, 1988.
- Chevalier *et al.*, *Mol. Cell.*, 10:895-905, 2002.
- Current Protocols in Stem Cell Biology*, Bhatia *et al.* (Ed.), John Wiley and Sons, Inc., 2007.
- Derossi *et al.*, *J. Bio. Chem.*, 269:10444-10450, 1994. Derossi *et al.*, *J. Biol. Chem.*, 271:18188, 1996.

- Derossi *et al.*, *Trends in Cell Biol.*, 8:84-87, 1998.
- Durai *et al.*, *Nucleic Acids Res.*, 33:5978-5990, 2005.
- Elliott and O'Hare, *Cell*, 88:223-234, 1997.
- EP 1507865
- EP0412700
- Ercolani *et al.*, *J. Biol. Chem.*, 263:15335-15341, 1988.
- Evans, *et al.*, In: *Cancer Principles and Practice of Oncology*, Devita *et al.* (Eds.), Lippincott-Raven, NY, 1054-1087, 1997.
- Fawell *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:664-668, 1994.
- Fechheimer *et al.*, *Proc Natl. Acad. Sci. USA*, 84:8463-8467, 1987.
- Ferry *et al.*, *Hum. Gene Ther.*, 9(14):1975-81, 1998.
- Follenzi *et al.*, *Hum. Gene Ther.*, 13(2):243-60, 2002.
- Fraley *et al.*, *Proc. Natl. Acad. Sci. USA*, 76:3348-3352, 1979.
- Frankel and Pabo, *Cell*, 55(6):1189-1193, 1988.
- Gebhart and Wang, *J. Cell Sci.*, 56233-244, 1982.
- Gehrke *et al.*, *Gene*, 322:137-43, 2003.
- Ghosh and Bachhawat, In: *Liver Diseases, Targeted Diagnosis and Therapy Using Specific Receptors and Ligands*, Wu *et al.* (Eds.), Marcel Dekker, NY, 87-104, 1991.
- Ghosh *et al.*, *J. Hepatol.*, 32(1Suppl):238-52, 2000.
- Gomes-Lechon *et al.*, In: *In vitro Methods in Pharmaceutical Research*, Academic Press, 129-153, 1997
- Gomez-Lechon *et al.*, *Anal. Biochem.*, 236:296, 1996.
- Gopal, *Mol. Cell Biol.*, 5:1188-1190, 1985.
- Graham and Van Der Eb, *Virology*, 52:456-467, 1973.
- Grompe *et al.*, *Sem. Liver Dis.*, 19:7, 1999.
- Gronthos, *Blood*, 84(12):4164-4173, 1994.
- Hancock *et al.*, *EMBO J.*, 10:4033-4039, 1991.
- Harland and Weintraub, *J. Cell Biol.*, 101(3):1094-1099, 1985.
- Hayhurst *et al.*, *Mol. Cell Biol.*, 21(4):1393-403, 2001.
- Hill *et al.*, *Exp. Hematol.*, 24(8):936-943, 1996.
- Ho *et al.*, *Cancer Res.*, 61(2):474-7, 2001.
- In vitro Methods in Pharmaceutical Research*, Academic Press, 1997.
- Jaiswal *et al.*, *J. Cell Biochem.*, 64(2):295-312, 1997.
- Johnstone *et al.*, 238(1):265-272, 1998.

- Kaepler *et al.*, *Plant Cell Rep.*, 8:415–418, 1990.
- Kaneda *et al.*, *Science*, 243:375-378, 1989.
- Karin *et al.* *Cell*, 36: 371-379,1989.
- Kato *et al.*, *J. Biol. Chem.*, 266:3361-3364, 1991.
- Kilic *et al.*, *Stroke*, 34:1304-10, 2003.
- Kirchmaier and Sugden, *J. Virol.*, 72(6):4657–4666, 1998.
- Kirkeby *et al.*, *Biochem. Biophys. Meth.*, 24:225, 1992.
- Klein *et al.*, *Nature*, 327:70-73, 1987.
- Kobayashi *et al.*, *Science*, 287:1258, 2000.
- Kramer *et al.*, *Mol. Ther.*, 7(3):375-85, 2003.
- Laarse *et al.*, *Biotech Histochem.*, 67:303, 1992.
- Langle-Rouault *et al.*, *J. Virol.*, 72(7):6181–6185, 1998.
- Le *et al.*, *Blood*, 89(4):1254-9, 1997.
- Levitkaya *et al.*, *Proc. Natl. Acad. Sci. USA*, 94(23):12616–12621, 1997.
- Li *et al.*, *Genes Dev.*, 14:464-74, 2000.
- Lieber *et al.*, *Proc. Natl. Acad. Sci. USA*, 92:6210, 1995.
- Lindgren *et al.*, *Trends in Pharmacol. Sci.*, 21:99-103, 2000.
- Lindner *et al.*, *J. Virol.*, 82(12):5693-702, 2008.
- Macdonald *et al.*, pp. 252-286, *Cell Encapsulation Technology and Therapeutics*
- Macejak and Sarnow, *Nature*, 353:90-94, 1991.
- Makino *et al.*, *J. Clin. Invest.*, 103(5):697-705, 1999.
- Maniatis, *et al.*, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1988.
- Mann and Frankel, *EMBO J.*, 10:1733-1739, 1991. Mann *et al.*, *Cell*, 33:153-159, 1983.
- Manno *et al.*, *Nat. Med.*, 12(3):342-7, 2006.
- Miao *et al.*, *Mol. Ther.*, 1(6):522-32, 2000.
- Mignon *et al.*, *Nature Med.*, 4:1185, 1998.
- Miller *et al.*, *Am. J. Clin. Oncol.*, 15(3):216-221, 1992.
- Miller *et al.*, *Nat. Biotechnol.*, 29:143-148, 2011.
- Miyoshi *et al.*, *J. Biomater. Sci. Polym. Ed.*, 9:227-237, 1998.
- Nabel *et al.*, *Science*, 244(4910):1342-1344, 1989.
- Naldini *et al.*, *Science*, 272(5259):263-267, 1996.
- Ng, *Nuc. Acid Res.*, 17:601-615,1989.

- Nicolas and Rubenstein, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt, eds., Stoneham: Butterworth, pp. 494-513, 1988.
- Nicolau and Sene, *Biochim. Biophys. Acta*, 721:185-190, 1982.
- Nicolau *et al.*, *Methods Enzymol.*, 149:157-176, 1987.
- Ockerman, *Clin. Chim. Acta*, 17:201, 1968.
- Ohashi *et al.*, *Nature Med.*, 6:327, 2000.
- Overturf *et al.*, *Human Gene Ther.*, 9:295, 1998.
- Paskind *et al.*, *Virology*, 67:242-248, 1975.
- Passonneau and Lauderdale, *Anal. Biochem.*, 60:405-415, 1974.
- Pingoud and Silva, *Nat. Biotechnol.*, 25:743-744, 2007.
- PCT Appln. WO 01/81549
- PCT Appln. WO 03/059940
- PCT Appln. WO 03/059941
- PCT Appln. WO 94/09699
- PCT Appln. WO 94/09699
- PCT Appln. WO 95/06128
- PCT Appln. WO 99/20741
- PCT Appln. WO/2003/042405
- PCT Appln. WO 01/098482
- PCT Appln. WO 95/011308
- PCT Appln. WO 96/39487
- PCT/US2004/030606
- PCT/IB2010/000154
- Peeters *et al.*, *Hepatology*, 25:884, 1997.
- Pelletier and Sonenberg, *Nature*, 334:320-325, 1988.
- Potrykus *et al.*, *Mol. Gen. Genet.*, 199(2):169-177, 1985.
- Potten, *Philos. Trans. R Soc. Lond. B Biol. Sci.*, 353:821-30, 1998.
- Potter *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:7161-7165, 1984.
- Quitsche *et al.*, *J. Biol. Chem.*, 264:9539-9545, 1989.
- Reubinoff *et al.*, *Nat. Biotechnol.*, 18:399B404, 2000.
- Rhim *et al.*, *Proc. Natl. Acad. Sci. USA*, 92:4942, 1995.
- Richards *et al.*, *Cell*, 37: 263-272, 1984.
- Rippe, *et al.*, *Mol. Cell Biol.*, 10:689-695, 1990.
- Rothbard *et al.*, *Nat. Med.*, 6(11):1253-7, 2000.

- Rudolph *et al.*, *Science*, 287:1253, 2000.
- Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, 3<sup>rd</sup> Ed. Cold Spring Harbor Lab. Press, 2001.
- Sambrook *et al.*, *In: Molecular Cloning: A Laboratory Manual*, Vol. 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (7)7:19-17.29, 1989.
- Schwarze *et al.*, *Science*, 285(5433):1466-7, 1999.
- Schwarze *et al.*, *Science*, 285:1569-1572, 1999.
- Sheehan and Hrapchak, *In: Theory and Practise of Histotechnology*, 2<sup>nd</sup> Ed., Battelle Memorial Institute, Columbus, OH, 1987.
- Shen *et al.*, *DNA*, 8(2):101-8, 1989.
- Shiojiri, *J. Embryol. Exp. Morph.*, 62:139, 1981.
- Simonet *et al.*, *J. Biol. Chem.*, 268(11):8221-9, 1993.
- Smith, *In: Origins and Properties of Mouse Embryonic Stem Cells*, *Annu. Rev. Cell. Dev. Biol.*, 2000.
- Takahashi and Yamanaka, *Cell*, 126:663-676, 2006.
- Takahashi *et al.*, *Cell*, 126(4):663-76, 2007.
- Takahashi *et al.*, *Cell*, 131:861-872, 2007.
- Tanaka *et al.*, *J. Immunol.*, 170(3):1291-8, 2003.
- Temin, *In: Gene Transfer*, Kucherlapati (Ed.), NY, Plenum Press, 149-188, 1986.
- Thompson, *In: Selected Histochemical and Histopathological Methods*, Tomas (Ed.), Sprungfield, IL, 1966.
- Thomson and Marshall, *Curr. Top. Dev. Biol.*, 38:133-165, 1998.
- Thomson and Odorico, *J. Trends. Biotechnol.*, 18:53B57, 2000.
- Thomson *et al.* *Proc. Natl. Acad. Scie. USA*, 92:7844-7848, 1995.
- Thomson *et al.*, *Science*, 282:1145, 1998.
- Tur-Kaspa *et al.*, *Mol. Cell Biol.*, 6:716-718, 1986.
- VandenDriessche *et al.*, *J. Thromb. Haemost.*, 5(1):16-24, 2007.
- Vickers *In: In vitro Methods in Pharmaceutical Research*, Academic Press, 375-410, 1997
- Watt, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 353:831, 1997.
- Wender *et al.*, *Proc. Natl. Acad. Sci. USA*, 97(24):13003-8, 2000.
- Wilson *et al.*, *Science*, 244:1344-1346, 1989.
- Wong *et al.*, *Gene*, 10:87-94, 1980.
- Wu and Wu, *Biochemistry*, 27: 887-892, 1988.
- Wu and Wu, *J. Biol. Chem.*, 262:4429-4432, 1987.

- Xu *et al.*, *Nat. Biotechnol.*, 19:971-974, 2001.
- Yakubov *et al.*, *Biochemical and Biophysical Research Communications* 394: 189-193, 2010.
- Yang and Russell, *Proc. Natl. Acad. Sci. USA*, 87:4144-4148, 1990.
- Yasmineh *et al.*, *Clin. Biochem.*, 25:109, 1992.
- Ying *et al.*, *Cell*, 115:281-292, 2003.
- Yoo *et al.*, *J. Bone Joint Sure. Am.*, 80(12):1745-1757, 1998.
- Yu and Thompson, *Genes Dev.*, 22(15):1987-97, 2008.
- Yu *et al.*, *Science*, 318:1917-1920, 2007.
- Yu *et al.*, *Science*, 324(5928):797-801, 2009.
- Yull *et al.*, *Transgenic Res.*, 4(1 ):70-4, 1995.
- Zufferey *et al.*, *Nat. Biotechnol.*, 15(9):871-875, 1997.

**WHAT IS CLAIMED IS:**

1. A method of producing hepatocytes by forward programming of stem cells, comprising: culturing the stem cells under conditions to artificially increase the expression level of FOXA2, HNF4A and one or more additional hepatocyte programming factor genes selected from the group consisting of HHEX, HNF1A, FOXA1, TBX3-1, GATA4, NR0B2, SCML1, CEBPB, HLF, HLX, NR1H3, NR1H4, NR1I2, NR1I3, NR5A2, SEBOX, ZNF391, and ZNF517, thereby producing hepatocytes from forward programming of the stem cells.

2. The method of claim 1, wherein the stem cells comprise at least one exogenous expression cassette comprising one or more of the hepatocyte programming factor genes.

3. The method of any of preceding claims, comprising contacting the stem cells with hepatocyte programming factors comprising gene products of the hepatocyte programming factor genes in an amount sufficient to cause forward programming of the stem cells to hepatocytes.

4. The method of any of preceding claims, wherein the gene products are polypeptides or RNA transcripts of the hepatocyte programming factor genes.

5. The method of any of preceding claims, wherein the stem cells are mesenchymal stem cells, hematopoietic stem cells, or induced pluripotent stem cells.

6. The method of any of preceding claims, wherein the number of the hepatocyte programming factor genes are two, three, four, five or six.

7. The method of any of preceding claims, wherein the hepatocyte programming factor genes comprise Forkhead box protein A1 (FOXA1), forkhead box A2 (FOXA2), hematopoietically-expressed homeobox protein (HHEX), hepatocyte nuclear factor 1 homeobox A (HNF1A), hepatocyte nuclear factor 4 alpha (HNF4A), GATA binding protein 4 (GATA4), NR0B2 (nuclear receptor subfamily 0, group B, member 2), sex comb on midleg-like 1 (SCML1) or T-box transcription factor (TBX3).

8. The method of any of preceding claims, wherein the hepatocyte programming factor genes comprise FOXA1, FOXA2, HHEX, HNF1A, HNF4A and TBX3.

9. The method of any of preceding claims, wherein the hepatocyte programming factor genes comprise FOXA2, HHEX, HNF4A, GATA4, NR0B2 and SCML1.

10. The method of any of preceding claims, wherein the stem cells or progeny cells thereof further comprise a reporter expression cassette comprising a hepatocyte-specific transcriptional regulatory element operably linked to a reporter gene.

11. The method of any of preceding claims, wherein the hepatocyte-specific transcriptional regulatory element is a promoter of albumin,  $\alpha$ -1-antitrypsin (AAT), cytochrome p450 3A4 (CYP3A4), apolipoprotein A-I, or APOE.

12. The method of any of preceding claims, wherein the hepatocytes comprise one or more of the hepatocyte characteristics comprising:

(i) expression of one or more hepatocyte markers including glucose-6-phosphatase, albumin,  $\alpha$ -1-antitrypsin (AAT), cytokeratin 8 (CK8), cytokeratin 18 (CK18), asialoglycoprotein receptor (ASGR), alcohol dehydrogenase 1, arginase Type I, cytochrome p450 3A4 (CYP3A4), liver-specific organic anion transporter (LST-1), or a combination thereof;

(ii) activity of glucose-6-phosphatase, CYP3A4, bile production or secretion, urea production, or xenobiotic detoxification;

(iii) hepatocyte morphological features; or

(iv) *in vivo* liver engraftment in an immunodeficient subject.

13. The method of any of preceding claims, further comprising selecting or enriching for hepatocytes.

14. The method of any of preceding claims, wherein the stem cells or progeny cells thereof are cultured in a medium comprising one or more growth factors including Oncostain M (OSM).

15. The method of any of preceding claims, wherein the medium further comprises hepatocyte growth factor (HGF).



16. The method of any of preceding claims, wherein the stem cells and progeny cells thereof are cultured in a medium essentially free of fibroblast growth factor (FGF).

17. The method of any of preceding claims, wherein the stem cells and progeny cells thereof are cultured in a medium essentially free of epidermal growth factor (EGF).

18. The method of any of preceding claims, wherein the stem cells and progeny cells thereof are cultured in a medium essentially free of nicotinamide.

19. The method of any of preceding claims, comprising obtaining the hepatocytes less than or about 10 days after culturing in said conditions.

20. The method of any of preceding claims, comprising obtaining the hepatocytes less than or about 5 days after culturing in said conditions.

21. A method of assessing a compound for a pharmacological or toxicological effect on a hepatocyte, comprising:

a) contacting the hepatocyte provided by the method in accordance with any of claims 1 to 20 with the compound; and

b) assaying a pharmacological or toxicological effect of the compound on the hepatocyte.

22. A hepatocyte or stem cell comprising:

a) one or more exogenous expression cassettes comprising FOXA2, HNF4A and one or more additional hepatocyte programming factor genes selected from the group consisting of HHEX, HNF1A, FOXA1, TBX3-1, GATA4, NR0B2, SCML1, CEBPB, HLF, HLX, NR1H3, NR1H4, NR1I2, NR1I3, NR5A2, SEBOX, ZNF391; and

b) a reporter expression cassette comprising a hepatocyte-specific promoter operably linked to a reporter gene.

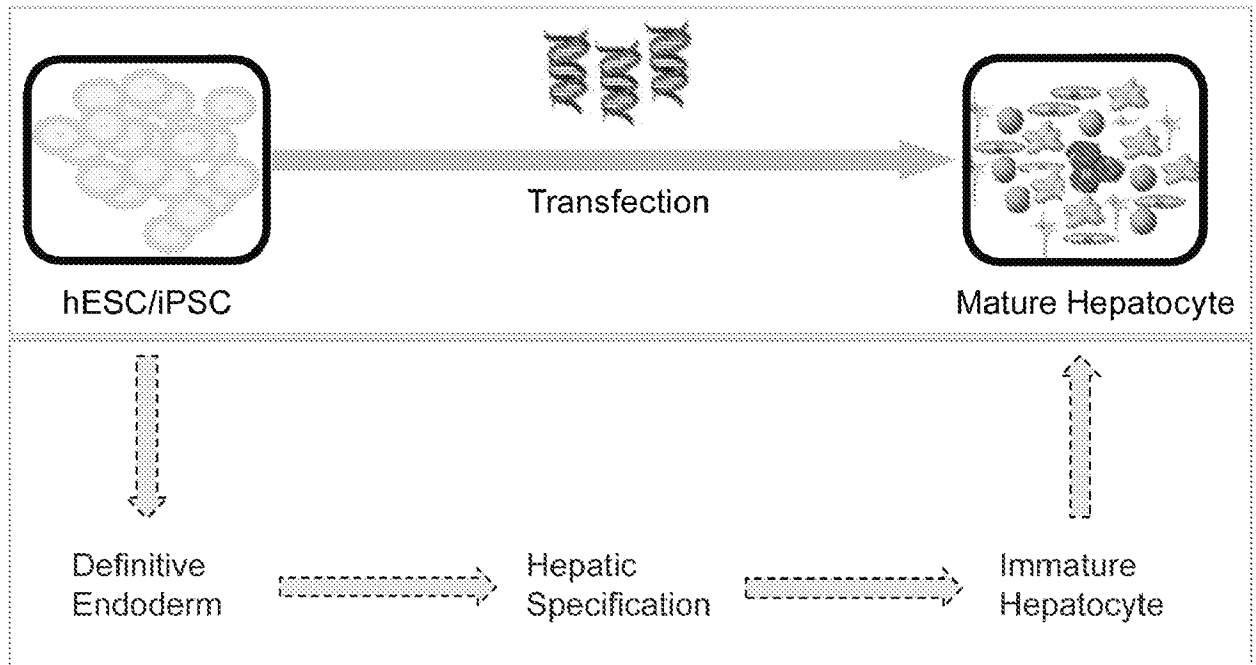
23. A hepatocyte or stem cell comprising one or more exogenous expression cassettes, wherein the one or more exogenous expression cassettes comprise FOXA2, HNF4A and one or more additional hepatocyte programming factor genes selected from the group consisting of HHEX, HNF1A, FOXA1, TBX3-1, GATA4, NR0B2, SCML1, CEBPB,

HLF, HLX, NR1H3, NR1H4, NR1I2, NR1I3, NR5A2, SEBOX, ZNF391, and at least one of the exogenous expression cassettes is operably linked to an externally inducible transcriptional regulatory element.

24. A cell population comprising hepatocytes, wherein at least 80% of the hematopoietic precursor cells comprise one or more exogenous expression cassettes that comprise FOXA2, HNF4A and one or more additional hepatocyte programming factor genes selected from the group consisting of HHEX, HNF1A, FOXA1, TBX3-1, GATA4, NR0B2, SCML1, CEBPB, HLF, HLX, NR1H3, NR1H4, NR1I2, NR1I3, NR5A2, SEBOX, ZNF391.

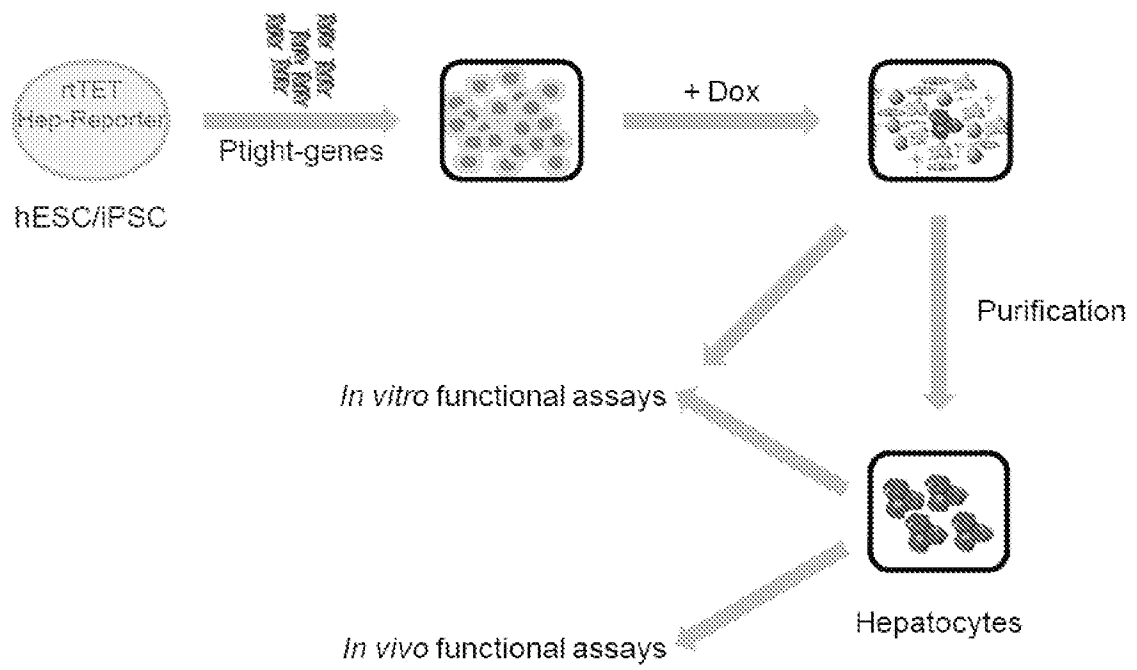
25. Hepatocytes, obtained by the method in accordance with any of claims 1 to 20, hepatocytes of claim 22 or 23, or a cell population of claim 24 for treating a subject having or at risk of having a liver dysfunction.

**Forward Programming**



**Normal differentiation**

**FIG. 1**



**FIG. 2**

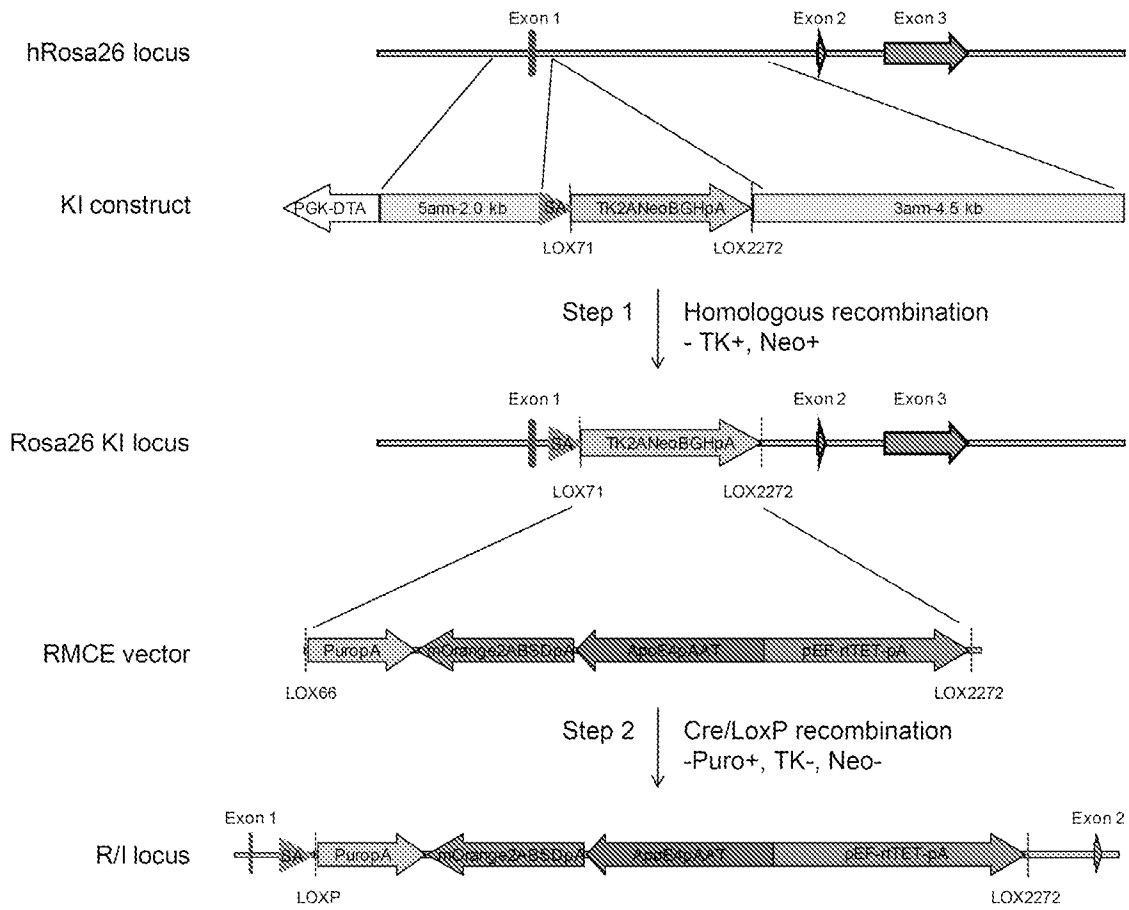
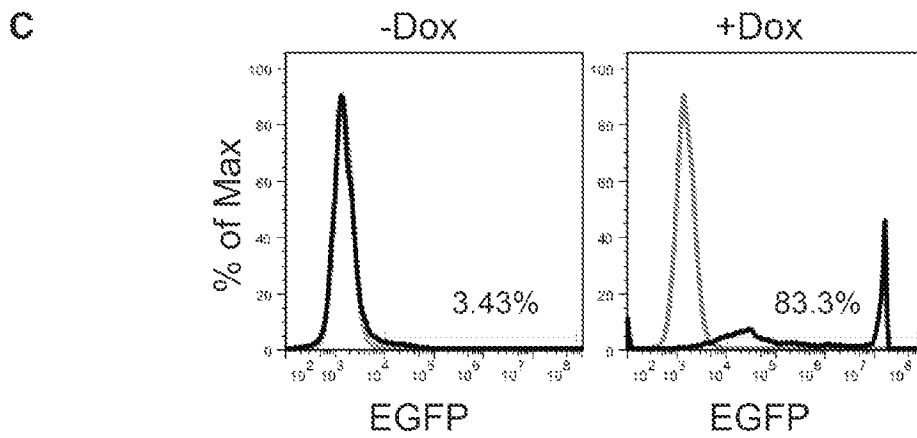
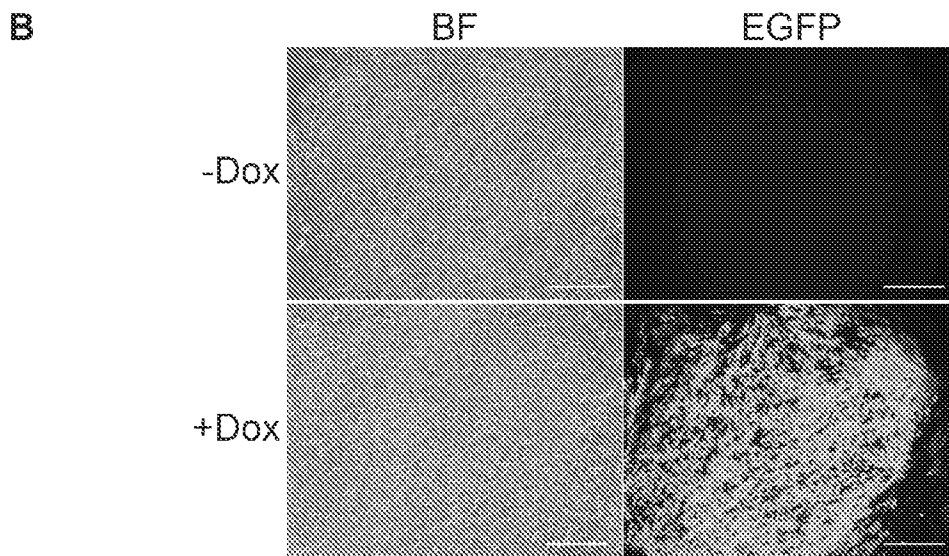
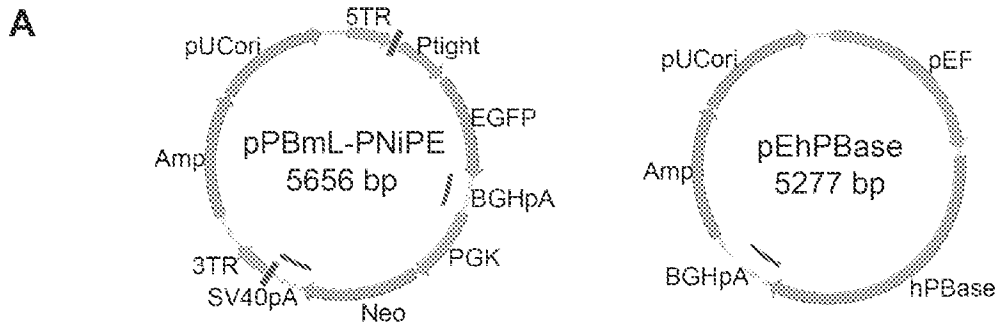


FIG. 3



5/8



FIGS. 5A-5C

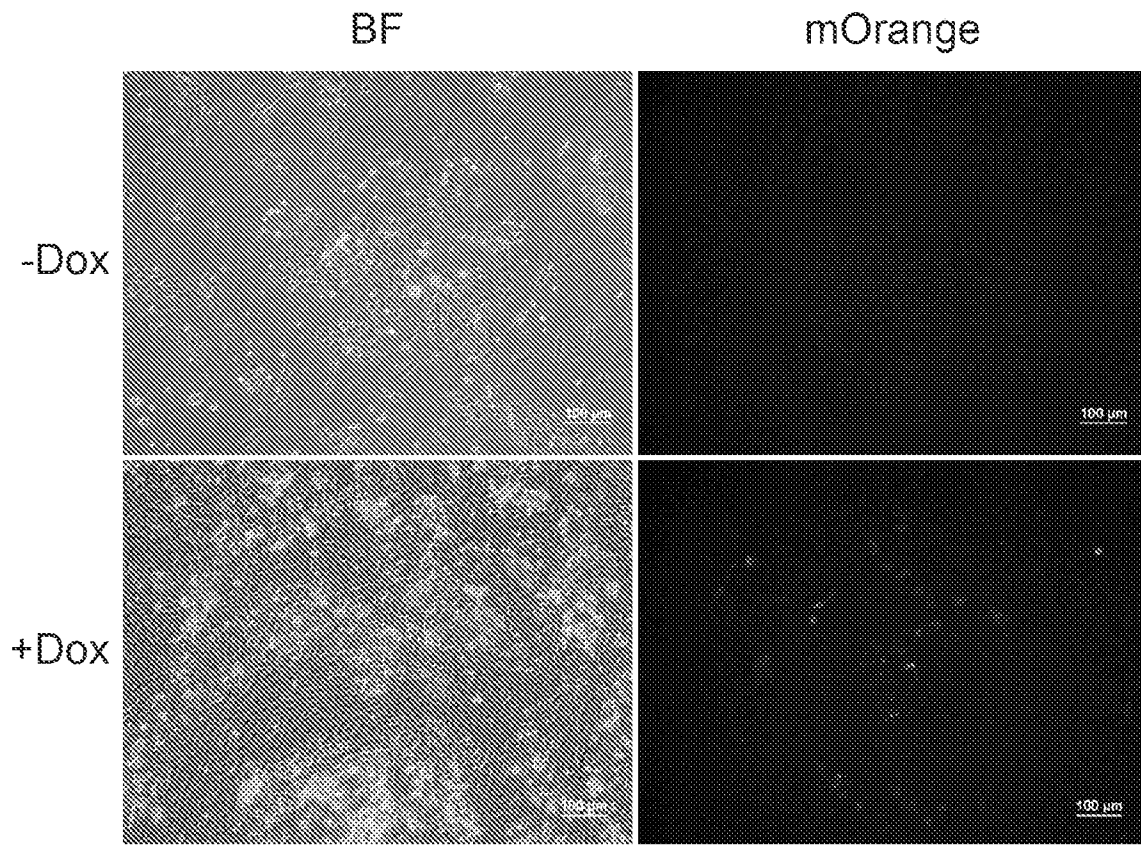
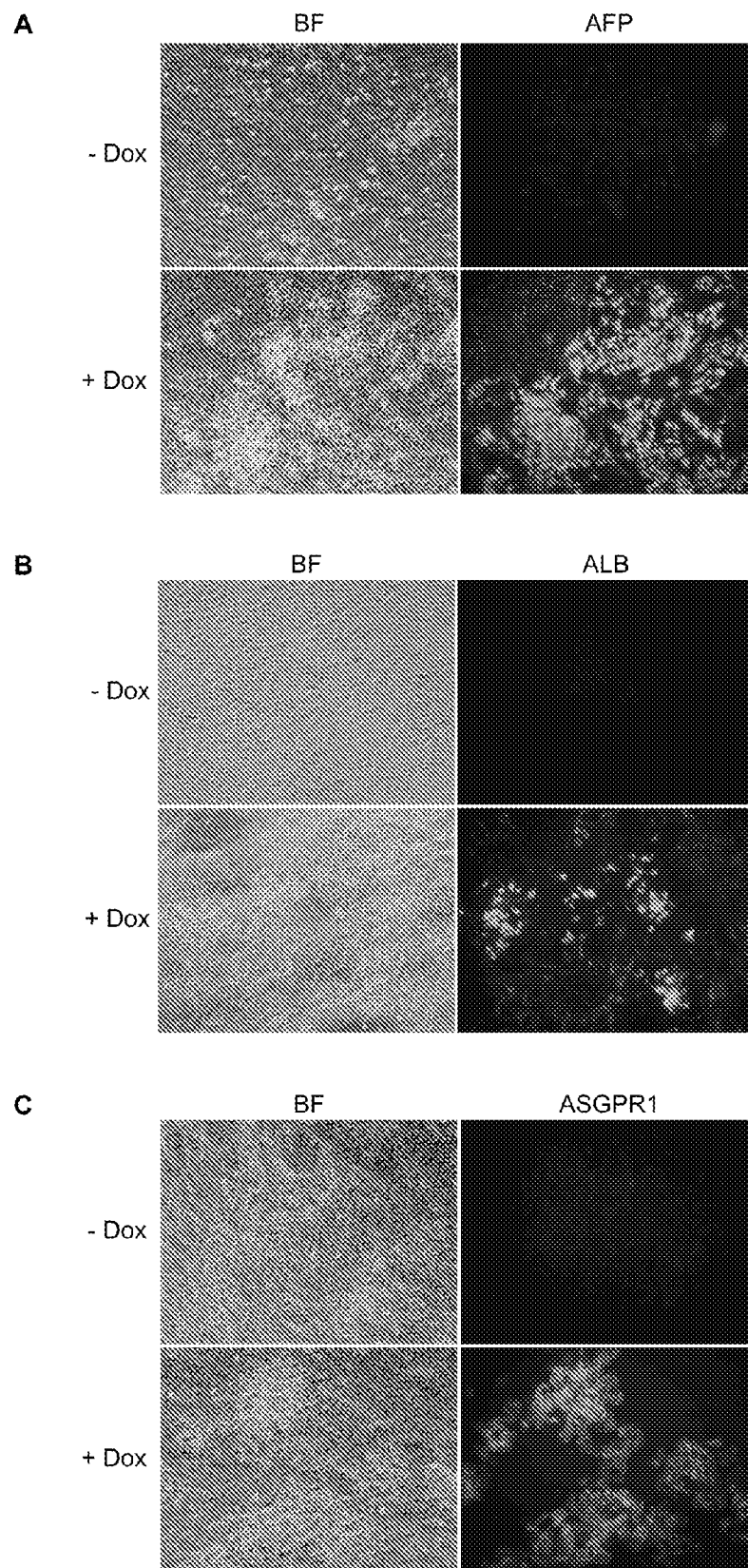


FIG. 6





FIGS. 7A-7C

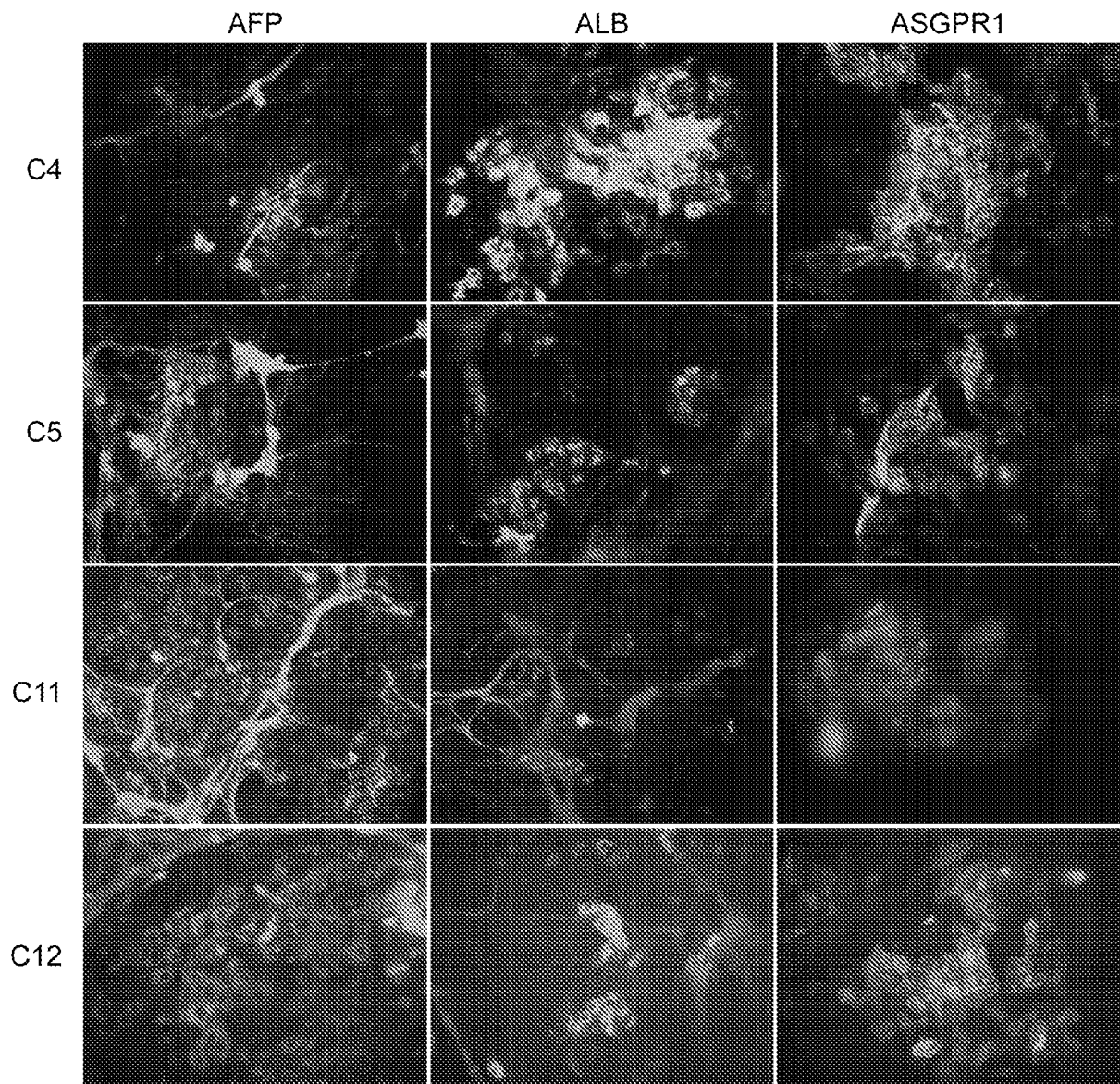


FIG. 8