



US 20240084257A1

(19) **United States**

(12) **Patent Application Publication**  
Slukvin et al.

(10) **Pub. No.: US 2024/0084257 A1**

(43) **Pub. Date: Mar. 14, 2024**

(54) **REAGENTS AND METHODS FOR PRODUCING ERYTHROMYELOID PROGENITOR CELLS, NK CELLS, AND MEGAKARYOCYTES**

(71) Applicant: **Wisconsin Alumni Research Foundation**, Madison, WI (US)

(72) Inventors: **Igor Slukvin**, Madison, WI (US); **Ho Sun Jung**, Madison, WI (US)

(21) Appl. No.: **18/463,211**

(22) Filed: **Sep. 7, 2023**

**Related U.S. Application Data**

(60) Provisional application No. 63/404,508, filed on Sep. 7, 2022.

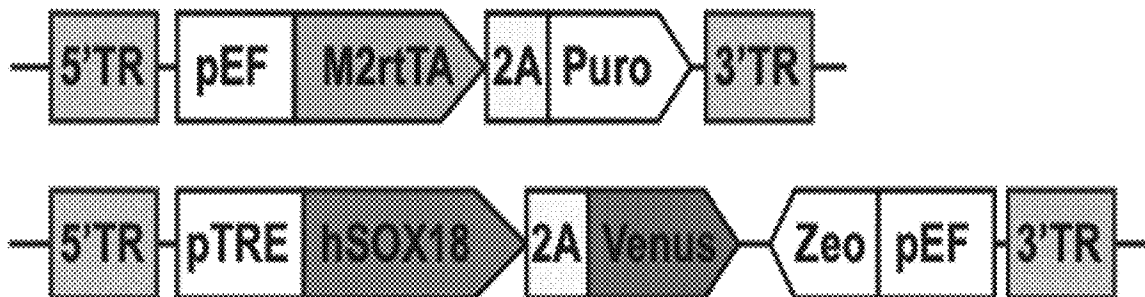
**Publication Classification**

(51) **Int. Cl.**  
*C12N 5/0783* (2006.01)  
*A61K 35/17* (2006.01)  
*A61K 35/19* (2006.01)  
*C12N 5/078* (2006.01)  
(52) **U.S. Cl.**  
CPC ..... *C12N 5/0646* (2013.01); *A61K 35/17* (2013.01); *A61K 35/19* (2013.01); *C12N 5/0644* (2013.01); *C12N 2501/60* (2013.01); *C12N 2506/45* (2013.01)

(57) **ABSTRACT**  
This disclosure provides reagents and methods for producing NK progenitor cells, NK cells, megakaryocytes and CAR derivatives thereof from hematopoietic stem cells. Pharmaceutical compositions comprising NK progenitor cells produced by the reagents and methods of the invention and immunotherapeutic methods using these pharmaceutical compositions are also provided.

**Specification includes a Sequence Listing.**

**PB-M2rtTA/PB-SOX18**



**PB-M2rtTA/PB-SOX18**

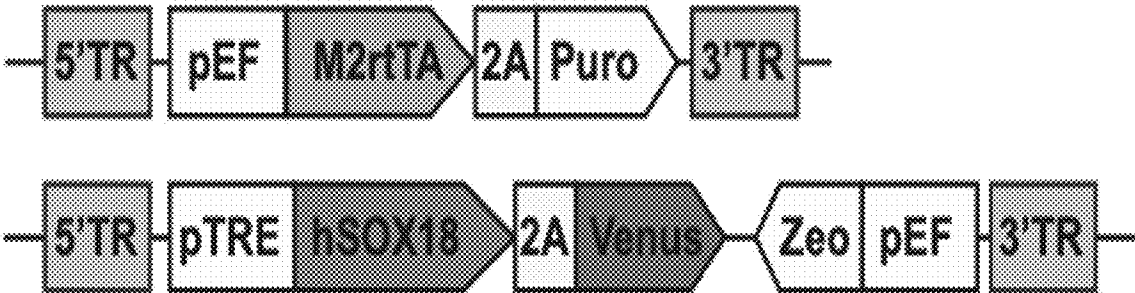


Figure 1

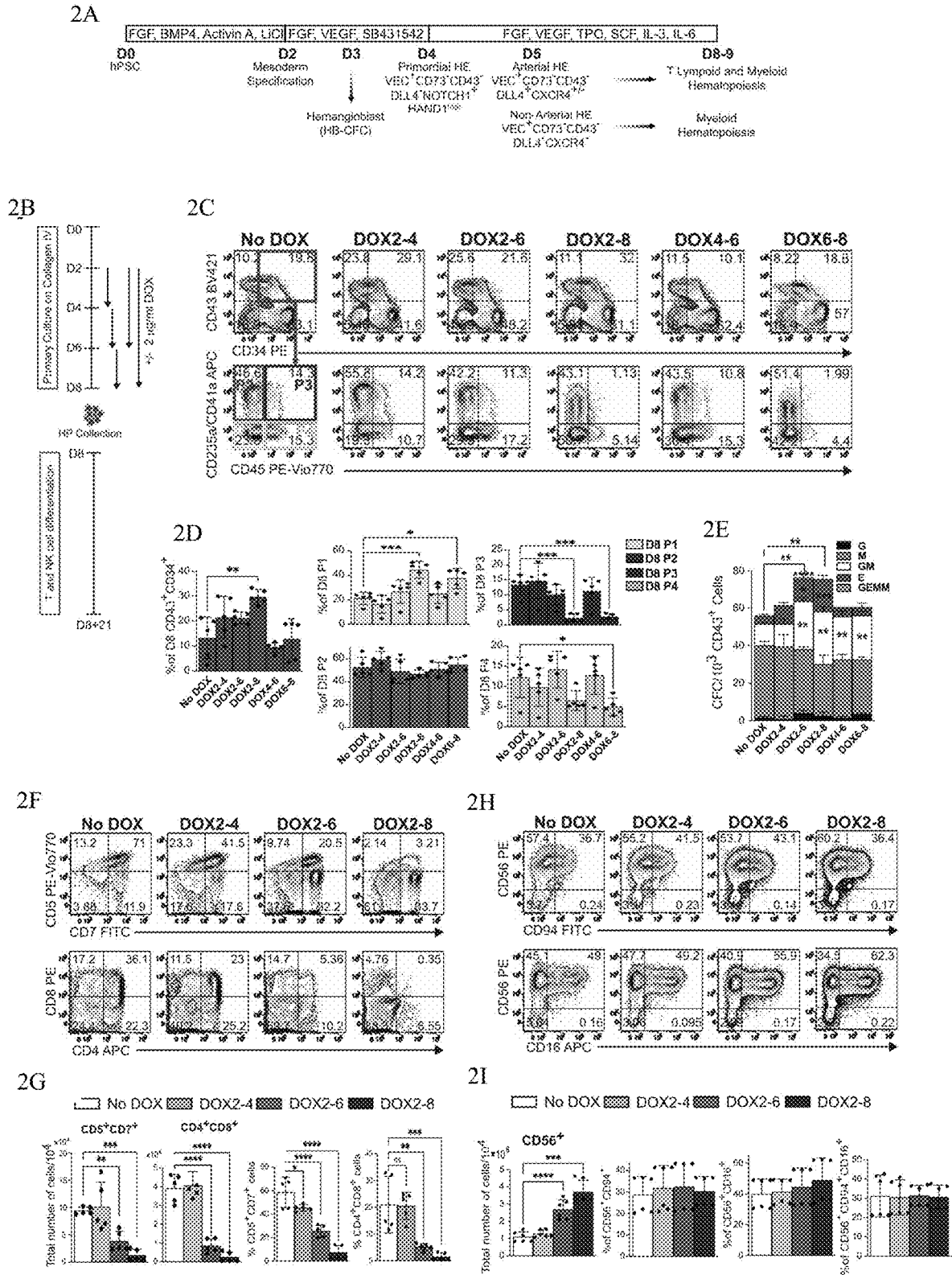


Figure 2

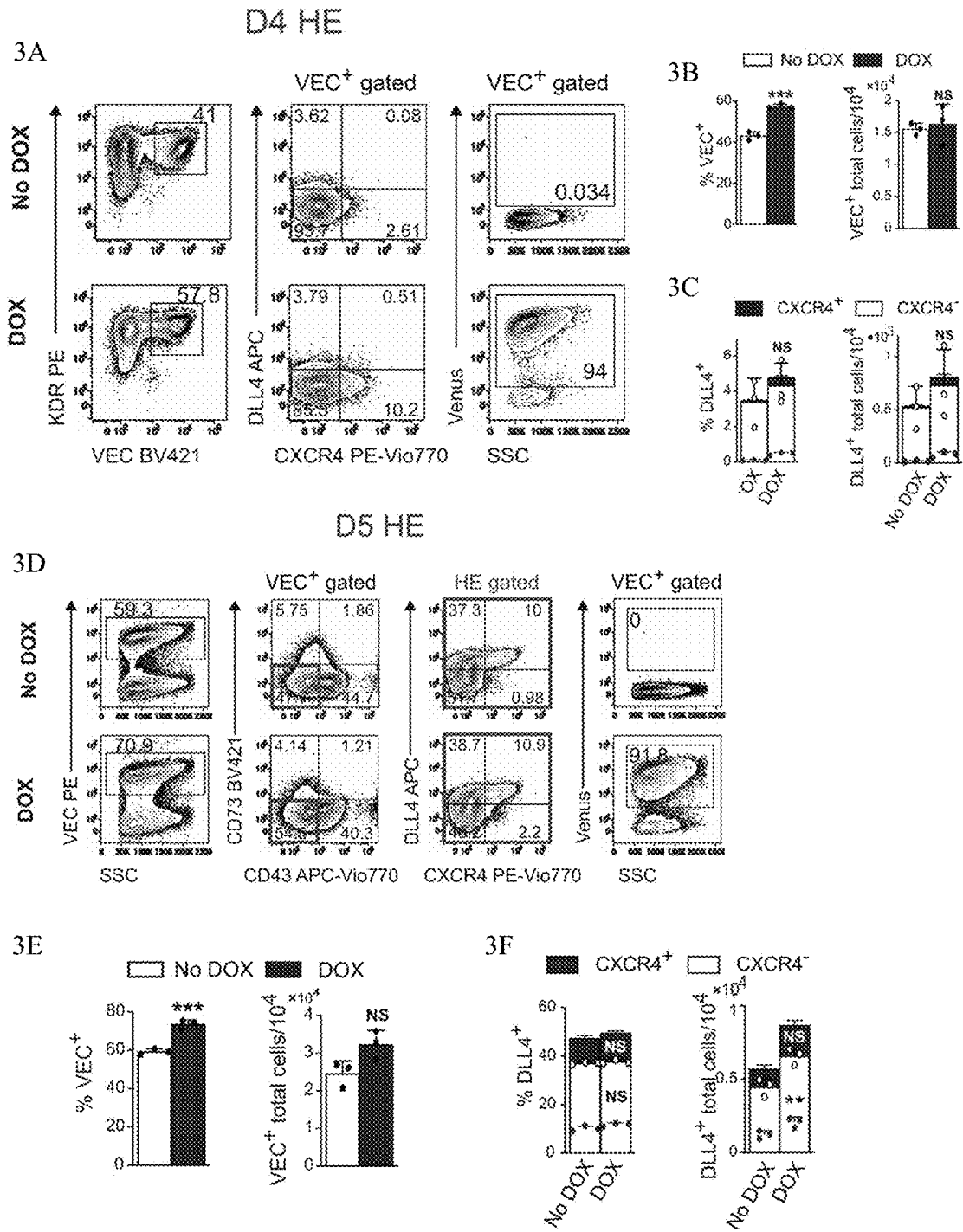


Figure 3

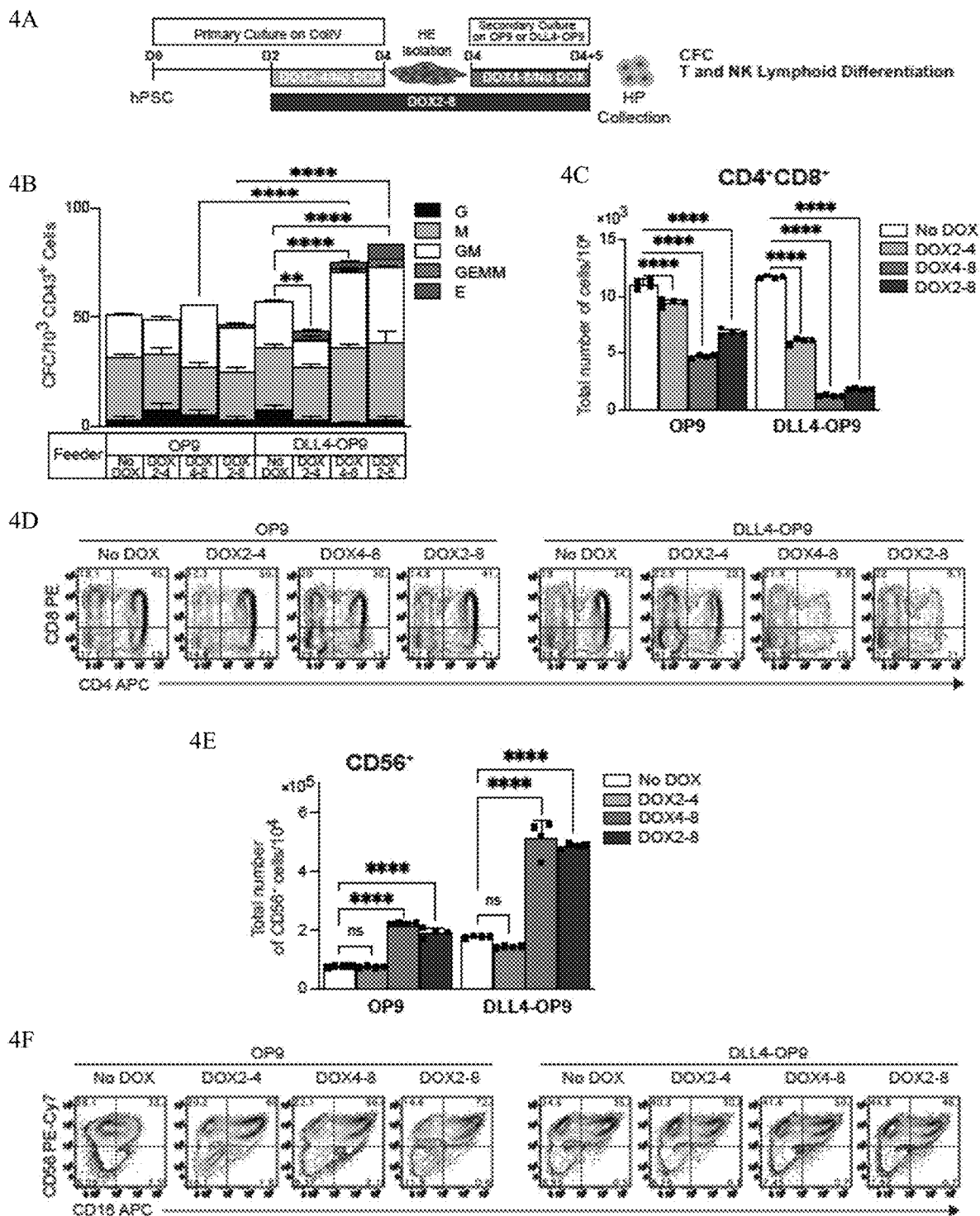
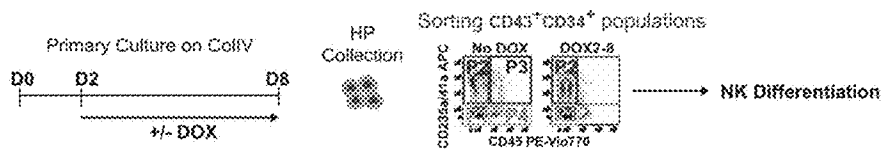
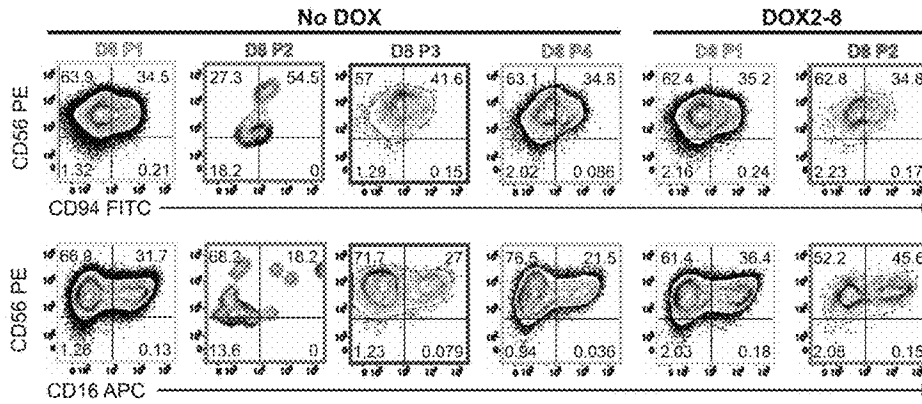


Figure 4

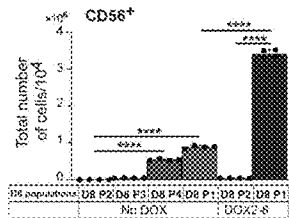
5A



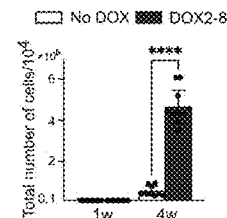
5B



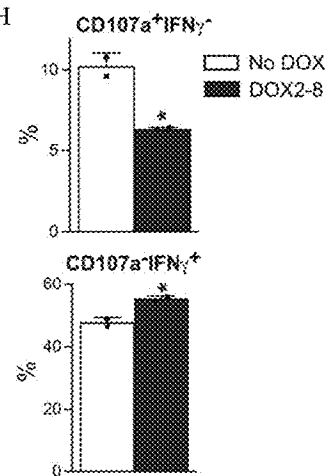
5C



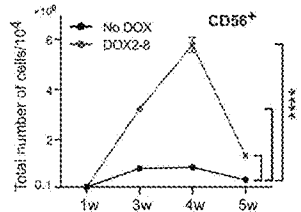
5D



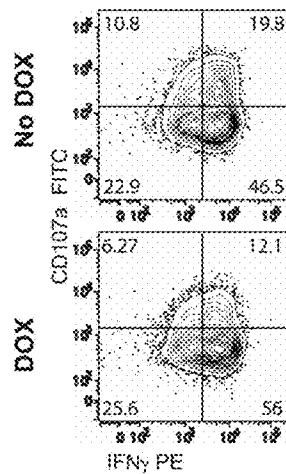
5H



5E



5G



5F

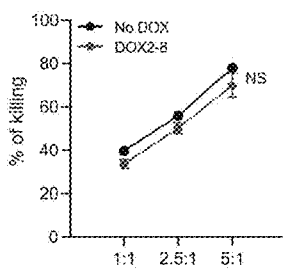


Figure 5

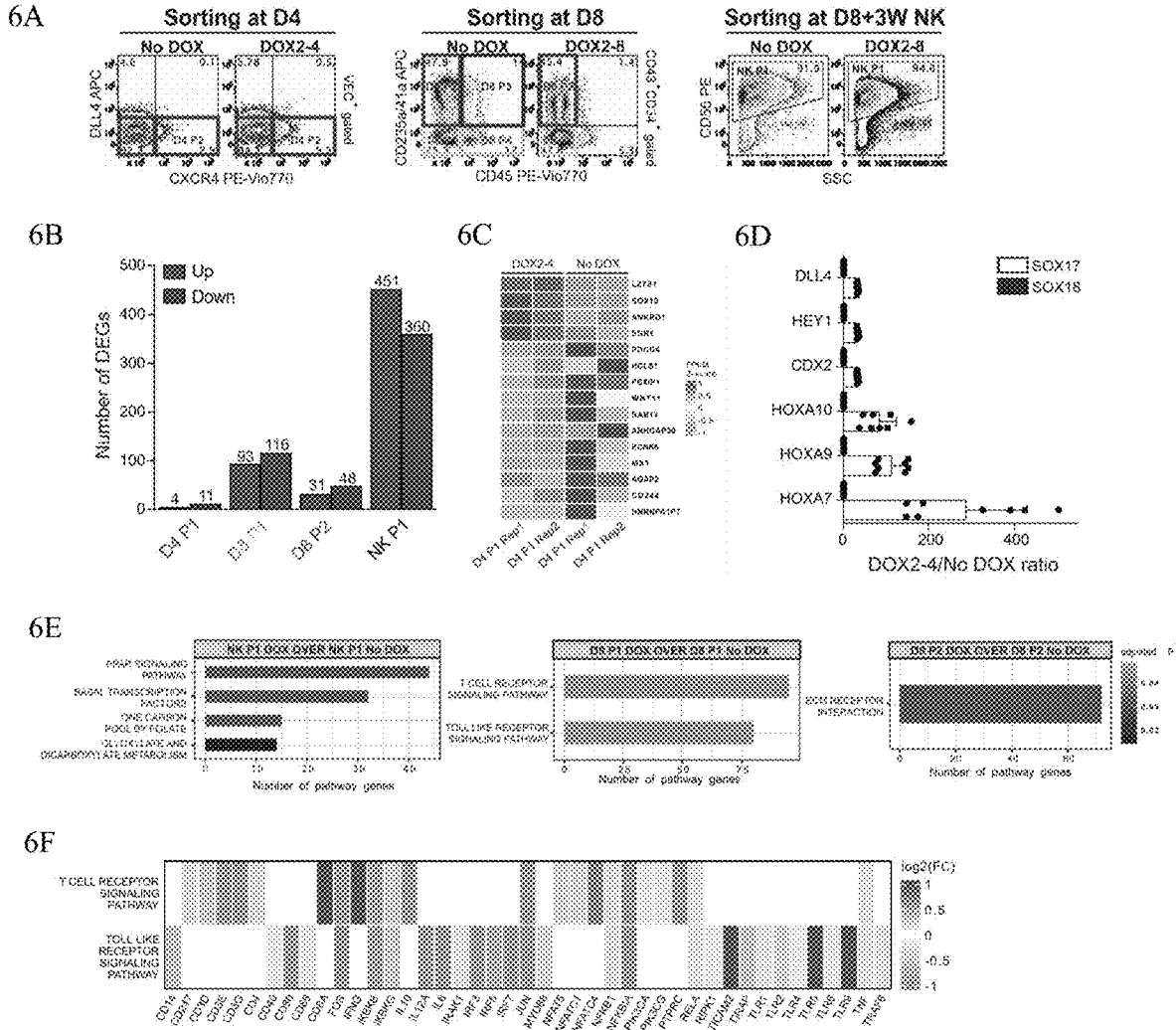


Figure 6

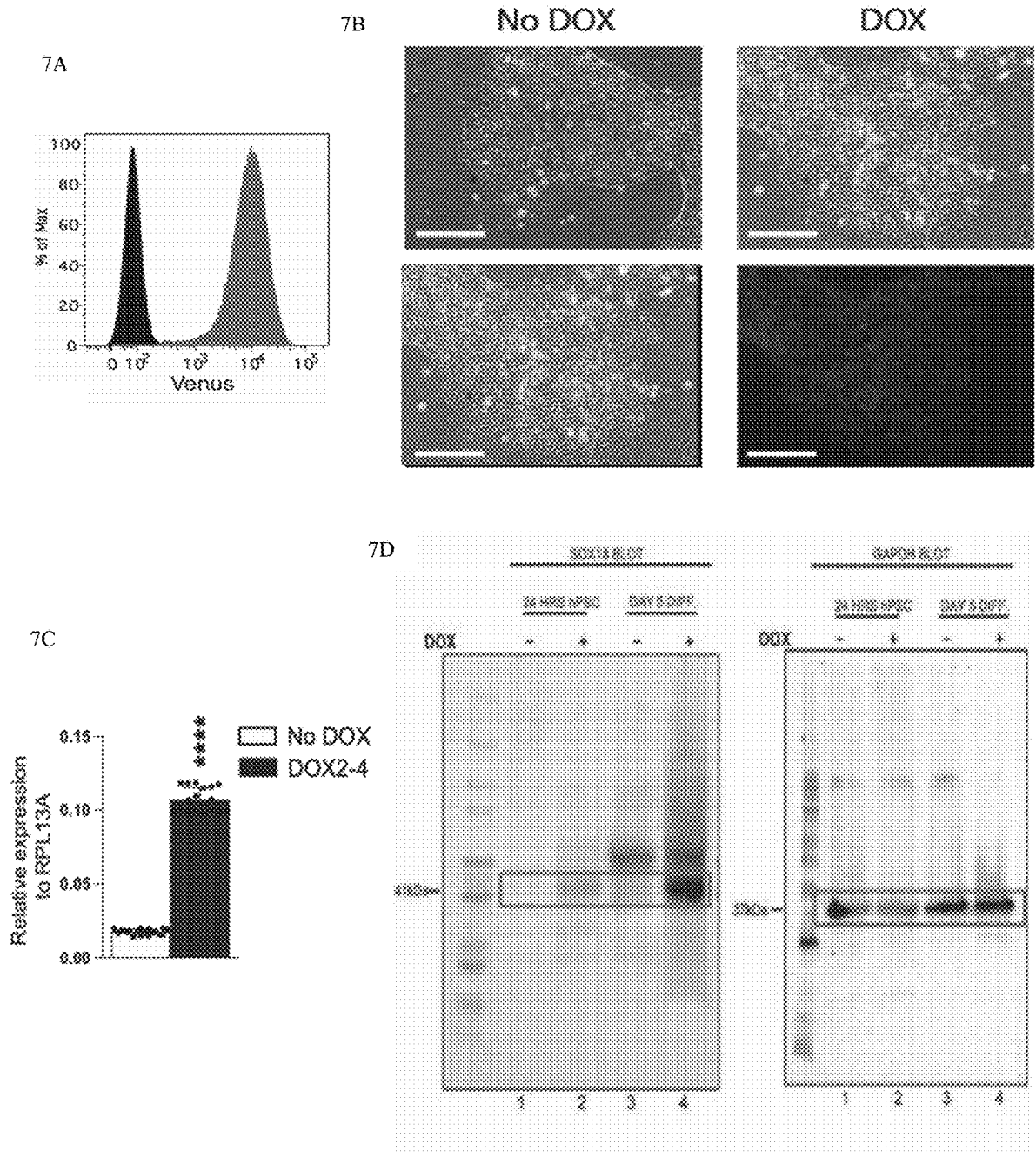


Figure 7



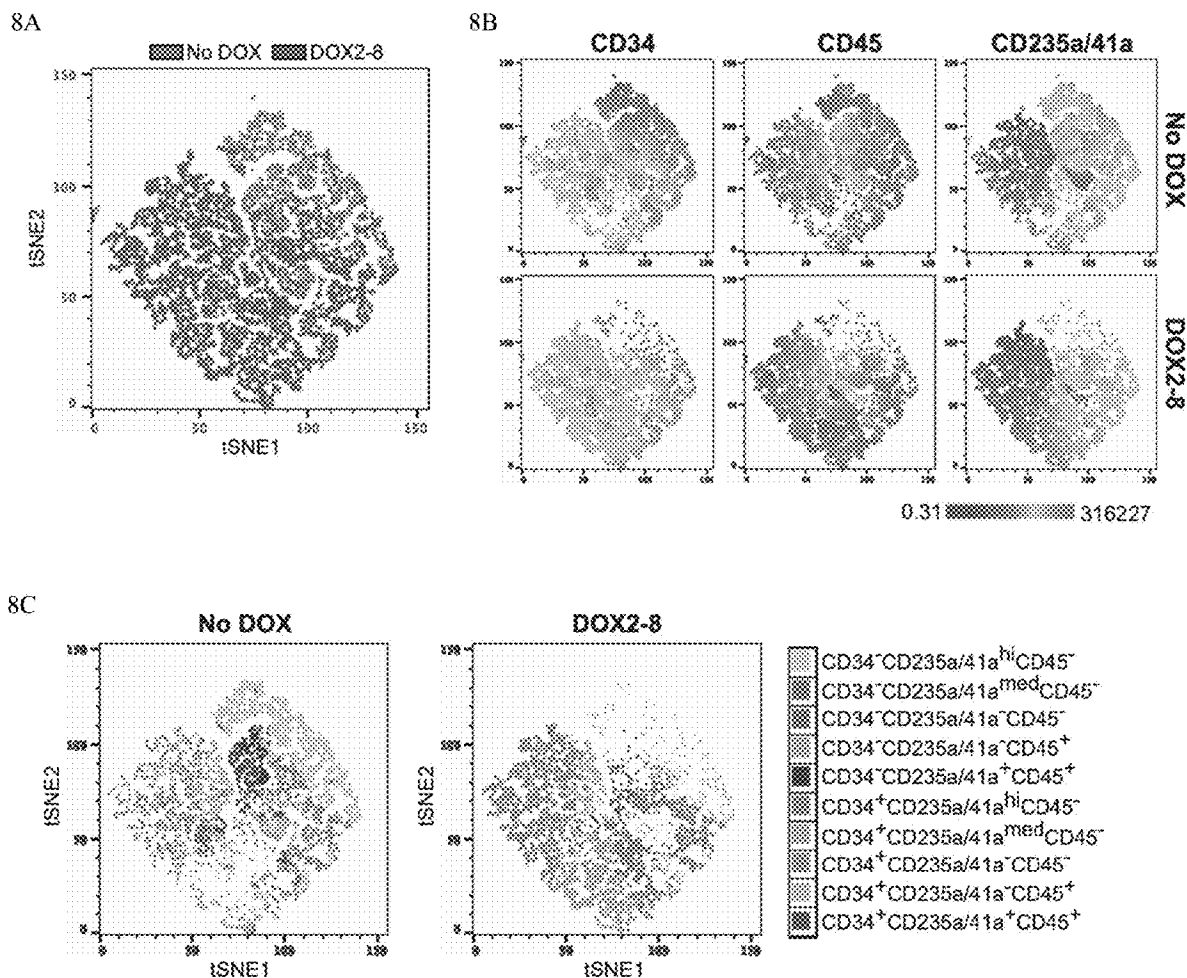


Figure 8

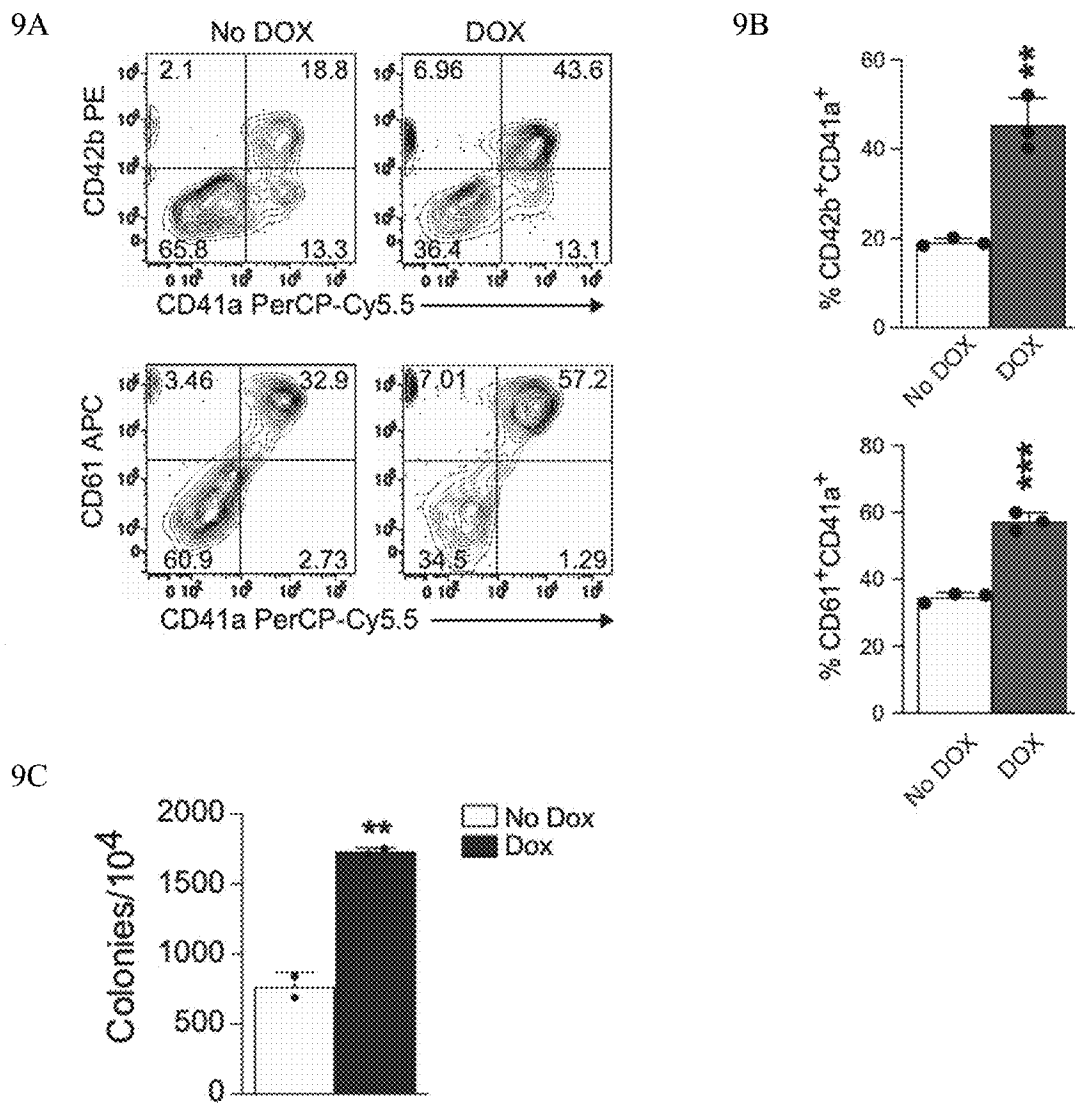
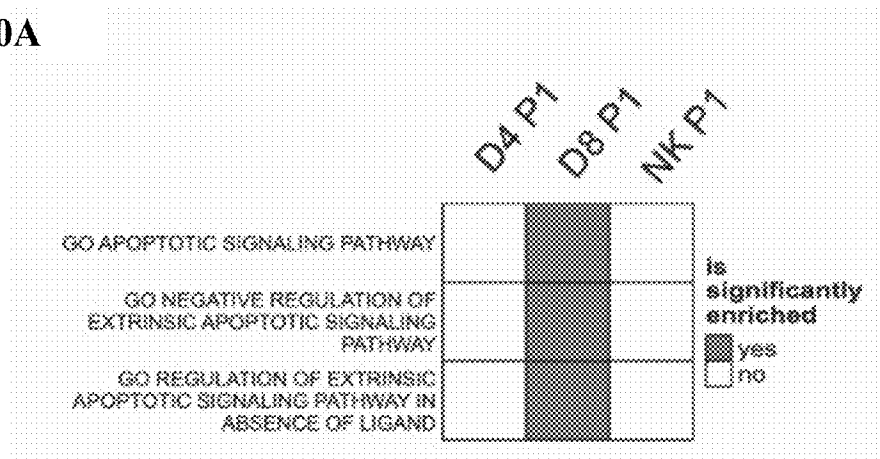


Figure 9

10A



10B

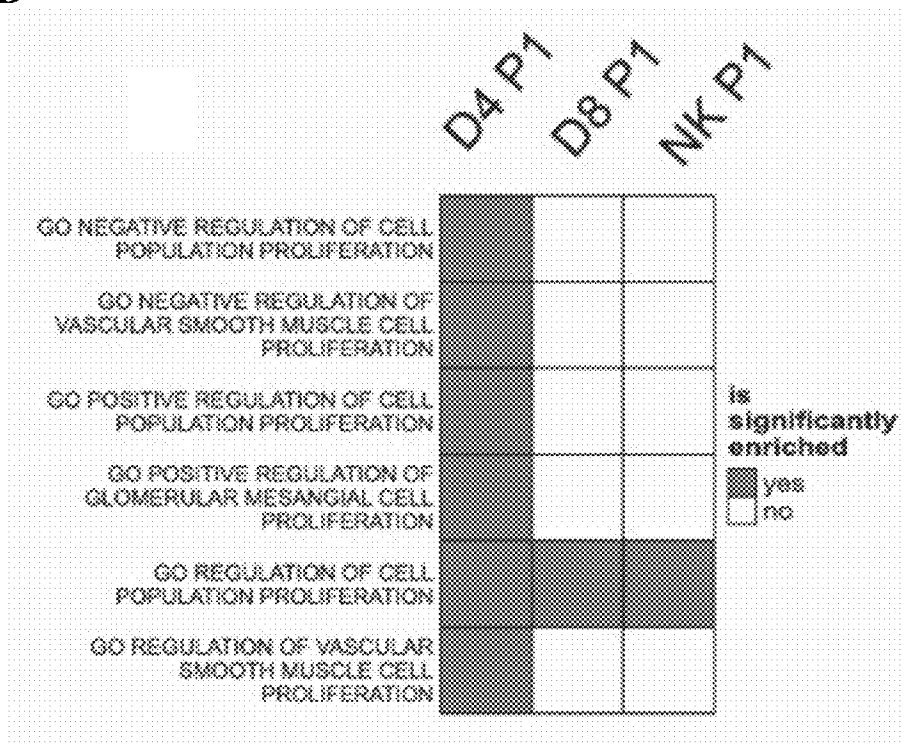
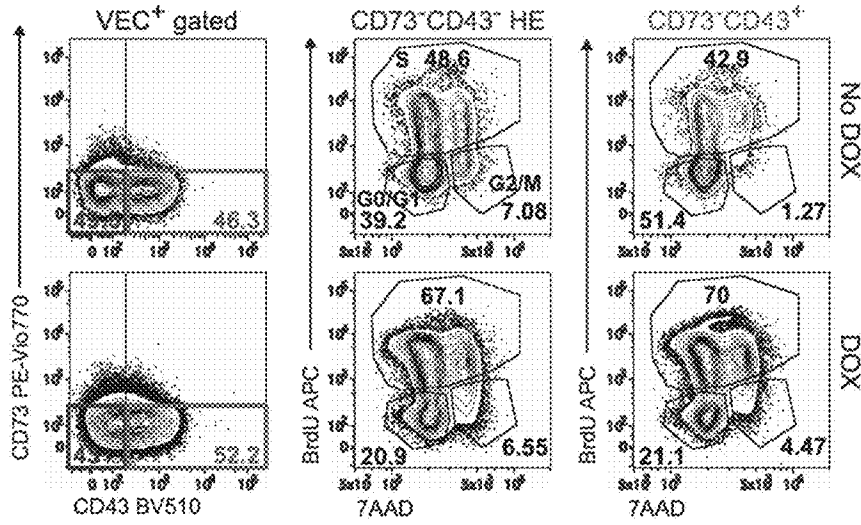


Figure 10

10C



10D

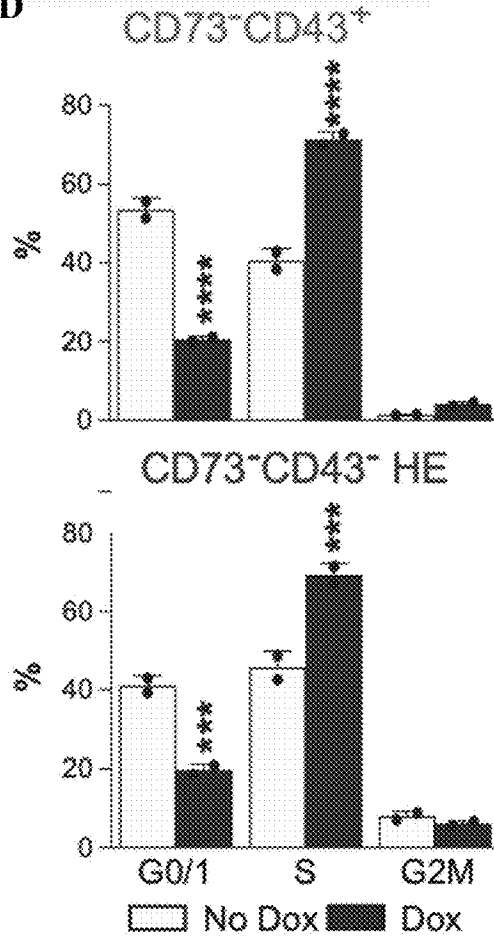


Figure 10 Cont.

**REAGENTS AND METHODS FOR  
PRODUCING ERYTHROMYELOID  
PROGENITOR CELLS, NK CELLS, AND  
MEGAKARYOCYTES**

FEDERALLY SPONSORED RESEARCH

**[0001]** This invention was made with government support under HL142665, OD011106 and HL134655 awarded by the National Institutes of Health. The government has certain rights in the invention.

STATEMENT REGARDING SEQUENCE  
LISTING

**[0002]** A computer readable form of the Sequence Listing is filed with this application by electronic submission and is incorporated into this application by reference in its entirety. The Sequence Listing is contained in the file created on Sep. 5, 2023, having the file name "22-0629-US.xml" and is 37 kb in size.

BACKGROUND OF THE DISCLOSURE

**[0003]** Elucidation of developmental pathways using human embryonic stem cells and induced pluripotent stem cells over the past 25 years have enabled production of therapeutically useful amounts of specific cell types that can, inter alia, provide allogenic sources of cells specific for individual patients. A specific example of such cell types is produced during hematopoiesis to provide a number of important components of blood function and homeostasis, including for example cells involved in immunological surveillance and function. One such cell is the natural killer or NK cell, a species of white blood cell in the lymphoid lineage which is characterized by having granules containing enzymes that can kill tumor cells or cells infected with a virus. Yet another are megakaryocytes, cells found in bone marrow that are responsible for producing thrombocytes (platelets).

**[0004]** Another consequence of the ability to study and manipulate stem cells and particularly hematopoietic stem cells has been an understanding of endogenous proteins produced during differentiation that effect progression to final effector cells like NK cells. These proteins, including transcription factors are responsible for activating specific genes in developmental pathways leading to particular differentiation outcomes. One example of such transcription factors is the SOXF family of transcription factors, comprising SOX7, SOX17 and SOX18, that have been recognized as critical regulators of angiogenesis, cardiovascular and hematopoietic development. See, for example, Lilly et al., 2017 SOXF transcription factors in cardiovascular development. *Seminars in cell & developmental biology* 63: 50-57.

**[0005]** Murine embryonic studies have shown that Sox7 is required for formation of the earliest multipotent hematopoietic progenitor cells (HPs) with erythro-myeloid potential. Gandillet et al., 2009, Sox7-sustained expression alters the balance between proliferation and differentiation of hematopoietic progenitors at the onset of blood specification. *Blood*.114: 4813-4822. Forced expression of Sox7 in cells from E7.5 mouse embryo or from in vitro differentiated mouse embryonic stem cells (ESCs) promotes self-renewal of early CD41<sup>+</sup> hemogenic progenitors with erythro-myeloid potential and blocks their differentiation. Lilly et al.,

2016, Interplay between SOX7 and RUNX1 regulates hemogenic endothelial fate in the yolk sac. *Development* 143: 4341-4351. A similar phenotype was observed following overexpression of Sox18 in in vitro-differentiated mouse ESCs. Serrano et al., 2010, Contrasting effects of Sox17- and Sox 18-sustained expression at the onset of blood specification. *Blood* 115: 3895-3898. SOX17 has been shown using human embryonic stem cells to be a master regulator of HOXA and arterial programs in hemogenic endothelium (HE) and is required for the specification of HE with robust lympho-myeloid potential and DLL4<sup>+</sup>CXCR4<sup>+</sup> phenotype resembling arterial HE at sites of HSC emergence. Jung et al., 2021, SOX17 integrates HOXA and arterial programs in hemogenic endothelium to drive definitive lympho-myeloid hematopoiesis. *Cell Rep.* 34(7):108758.

**[0006]** There is a need in this art to further elucidate functions of transcription factors in the SOXF family to develop reagents and methods for manipulating hematopoiesis in human stem cells to produce cells for therapeutic intervention and other purposes.

SUMMARY OF THE DISCLOSURE

**[0007]** Provided herein are reagents and methods for manipulating pluripotent stem cells, and particularly human stem cells, for producing specific cell types resulting from hematopoietic differentiation. In specific embodiments, such cells are natural killer (NK) cells. In other specific embodiments are megakaryocytes. Specific embodiments of the methods provided herein result in enforced SOX18 expression in stem cells and hemogenic endothelial cells during the endothelial-to-hematopoietic transition. In particular, such enforced SOX18 expression produces preferential commitment to NK progenitor cells. In specific embodiments such NK progenitor cells are derived from multipotent hematopoietic progenitors having a phenotype of CD34<sup>+</sup>CD43<sup>+</sup>CD235a/CD41a<sup>+</sup>CD45<sup>-</sup>. In particular embodiments the invention provides such cells, inter alia, for use in immunotherapy.

**[0008]** Yet another aspect of the disclosure provides a pharmaceutical composition comprising NK cells produced by forced expression of SOX18 in hemogenic endothelium. In some embodiments the NK cell is a genetically modified cell such as a CAR-NK cell. In certain embodiments, CAR-encoding expression constructs are introduced into pluripotent stem cells (PSCs) and those cells differentiated to CAR-NK cells by forced expression of SOX18. In alternative embodiments NK cells are produced from PSCs and CAR-encoding expression constructs introduced into the NK cells thereby produced.

**[0009]** In still further aspects of the disclosure are provided megakaryocytes through forced expression of SOX18 in mesoderm and hemogenic endothelium.

**[0010]** These and other features, objects, and advantages of the present invention will become better understood from the description that follows. In the description, reference is made to the accompanying drawings, which form a part hereof and in which there is shown by way of illustration, not limitation, embodiments of the invention. The description of preferred embodiments is not intended to limit the invention to cover all modifications, equivalents, and alternatives. Reference should therefore be made to the claims recited herein for interpreting the scope of the invention.

## DESCRIPTION OF THE DRAWINGS

[0011] The disclosure will be better understood and features, aspects, and advantages other than those set forth above will become apparent when consideration is given to the following detailed description thereof. Such detailed description refers to the following drawings.

[0012] FIG. 1 is a diagram of vectors used in the PiggyBac system to generate DOX-inducible SOX18 in an H1 hESC line.

[0013] FIGS. 2A-2I show the effects of SOX-18-enforced expression on hematopoietic differentiation of human pluripotent stem cells (hPSCs). FIG. 2A is a diagram of a timeline of hematopoietic differentiation, wherein D represents day of differentiation. Cell types arising after treatment of developing cell cultures set forth horizontal are shown at each developmental day underneath the timeline. FIG. 2B is a diagram of the experimental design. Cells were differentiated following DOX treatment as indicated. Floating HPs were collected on D8 of differentiation and evaluated for phenotype, CFC and lymphoid potential. FIG. 2C shows flow cytometric counter plots illustrating a phenotype of hematopoietic progenitors isolated from D8 differentiation cultures of DOX-induced SOX18 hPSCs. FIG. 2D illustrates percentages of CD34+CD43+ cells and composition of CD34+CD43+ subsets on D8 of differentiation (results are mean  $\pm$ SD, n=5 for two independent experiments performed in triplicate and duplicate). \*p<0.05, \*\*p<0.01 and \*\*\*p<0.001, one-way ANOVA, Tukey's multiple comparisons test. FIG. 2E is a graph showing CFC potential of D8 HPs, wherein results shown are mean  $\pm$ SD, n=2. \*\*p<0.01, two-way ANOVA. FIG. 2F shows Flow cytometric analysis of T cell differentiation induced by DOX on differentiating human pluripotent stem cells (hPSCs). FIG. 2G shows the total number of T cells generated from 104 of D8 CD43+ cells and percentages of CD5+CD7+ and CD4+CD8+ cells (results are means  $\pm$ SD, n=5 experiments). \*p<0.05, \*\*p<0.01 \*\*\*p<0.001, and \*\*\*\*p<0.0001, one-way ANOVA, Dunnett's multiple comparisons test. FIG. 2H shows flow cytometric analysis of NK cell differentiation induced by DOX on differentiating hPSCs. FIG. 2I is a graphical representation showing the total number of CD56+ cells generated from 104 CD43+ cells and the expression of CD94 and CD16 by NK cells (results are means  $\pm$ SD, n=6 for two independent experiments performed in triplicate). \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, and \*\*\*\*p<0.0001, one-way ANOVA, Dunnett's multiple comparisons test.

[0014] FIGS. 3A-3E show experimental results showing that SOX18 has a limited effect on specification of hemogenic endothelium. FIG. 3A shows flow cytometric results demonstrating expression of arterial markers and Venus in VEC+ SOX18 cells on day 4 of differentiation with or without DOX. FIG. 3B and FIG. 3C are graphical representations showing percentages and total number of VEC+ cells, DLL4+CXCR4+ and VEC+DLL4+CXCR4- cells generated from 10,000 hPSCs on day 4 of differentiation, wherein the results are shown as mean  $\pm$ SD, n=3. \*\*\*p<0.001, t-test. FIG. 3D shows flow cytometric results demonstrating expression of arterial markers and Venus in VEC+iSOX18 cells on D4 of differentiation with or without DOX. FIG. 3E and FIG. 3F are graphical representations showing percentages and total number of VEC+ cells, VEC+DLL4+CXCR4+ and VEC+DLL4+CXCR4- cells generated from 10<sup>4</sup>hPSCs on D5 of differentiation with or without DOX

(wherein results are shown as the mean  $\pm$ SD, n=3). \*\*\*p<0.001, t-test, and \*\*p<0.01, two-way ANOVA, Sidak's multiple comparisons test.

[0015] FIGS. 4A-4F show results of experiments illustrating that SOX18 overexpression affects lymphoid specification from HE cells by shifting the balance of NK and T lymphocyte differentiation. FIG. 4A is a schematic diagram illustrating the experimental design as set forth in Example 3. FIG. 4B is a graphical representation of CFC potential of HP generated from D4 HE after coculture on OP9 or DLL4-OP9 for 5 days with or without DOX, the results expressed as mean  $\pm$ SD, n=2. \*\*p<0.01 and \*\*\*\*p<0.0001, two-way ANOVA, Tukey's multiple comparisons test. FIG. 4C is a graphical representation of total number of CD4+ CD8+ T cells generated from 10,000 CD43+ cells, the results expressed as mean  $\pm$ SD, n=4. \*\*\*\*p<0.0001, two-way ANOVA, Sidak's multiple comparisons test. FIG. 4D shows flow cytometric results demonstrating the effect of SOX18 overexpression on T cell differentiation from hemogenic endothelium. FIG. 4E is a graphical representation of the total number of CD56+ NK cells generated from 10,000 CD43+ cells, the results expressed as mean  $\pm$ SD, n=4. \*\*\*\*p<0.0001, two-way ANOVA, Sidak's multiple comparisons test. FIG. 4F shows flow cytometric results demonstrating the effect of SOX18 overexpression on NK cell differentiation from hemogenic endothelium.

[0016] FIGS. 5A-5G illustrate characterization of NK cell potential in iSOX 18 hPSCs. FIG. 5A is a diagram of the experimental design showing DOX treatment schedule and cell subsets analyses as set forth in Example C. FIG. 5B shows flow cytometric results demonstrating NK cell differentiation potential of cell subsets isolated from cultures in the presence and absence of DOX. FIG. 5C is a graphical representation of total number of CD56+ NK cells generated from 10,000 cells from the indicated cell subset, the results expressed as mean  $\pm$ SDs, n=4. \*\*\*\*p<0.0001, one-way ANOVA, Tukey's multiple comparisons test. FIG. 5D is a graphical representation of the total numbers of CD56+ cells in cultures from No DOX and DOX2-8 HP after 1- and 4-weeks differentiation in NK cultures (results are means  $\pm$ SDs, n=10-11). \*\*\*\*p<0.0001, two-way ANOVA, Sidak's multiple comparisons test. FIG. 5E is a graphical representation of a growth curve of CD56+ cells in NK differentiation cultures initiated from D8 CD43+ cells generated in No DOX and DOX2-8-treated cultures. Cell cultures were started with 10<sup>5</sup>CD43+ cells and continued for 5 weeks (results are means  $\pm$ SDs, n=2). FIG. 5F is a graphical representation of a cytotoxicity assay against K562 targets. Different amount of target cells plated with CD56+ cells (effector: target ratio from 1:1 to 5:1) for 4 h (Results are means  $\pm$ SDs, n=2 experiments). FIG. 5G is a flow cytometric counter plot shows CD107a and interferon gamma (IFN $\gamma$ ) expression by CD56+ cells after K562 stimulation. FIG. 5H is a graphical representation of percentages of CD107a+IFN $\gamma$ - or CD107a-IFN $\gamma$ + cells after PMA/ionomycin stimulation. The results are expressed as mean  $\pm$ SD, \*p<0.05, t-test.

[0017] FIGS. 6A-6E illustrate molecular profiling of HE, HPs and NK cells generated from iSOX18 hPSCs. FIG. 6A shows flow cytometric results illustrating cell populations isolated for RNAseq analysis. FIG. 6B is a graphical representation showing the number of DEGs in indicated cell populations. FIG. 6C is a heatmap showing differentially expressed genes in D4 HE P1 population from cultures with

and without DOX. FIG. 6D shows bar plots illustrating the presence of significantly enriched KEGG pathways in selected differential expression comparisons. The y axis shows the enriched pathway categories and x axis shows the number of significantly genes in each pathway. FIG. 6E is a bar graph that shows significantly enriched KEGG pathways in selected differential expression comparisons. The y axis shows the enriched pathway categories and x axis shows the number of significantly enriched genes in each pathway. FIG. 6F shows the fold changes of selected genes in the two KEGG pathways that are significantly enriched in D8 P1 DOX+ cells over D8 P1 No DOX cells.

**[0018]** FIGS. 7A-7D illustrate generation of DOX-inducible SOX18 in H1 hESC line. FIG. 7A shows a flow Cytometric analysis in undifferentiated iSOX18 cells cultured with or without DOX. FIG. 7B are fluorescent images show the expression of Venus reporter in undifferentiated iSOX18 cells cultured with or without DOX. Scale bars are 200uM. FIG. 7C shows a qRT-PCR analysis shows SOX18 expression in D4 HE generated in iSOX18 hPSCs differentiation cultures with or without DOX. Results are mean  $\pm$  SD (N=9). \*\*\*\*p<0.0001, t-test. FIG. 7D is a western blot showing upregulation of SOX18 expression in undifferentiated iSOX18 cells 24 hours after DOX treatment and D5 differentiated iSOX18 hPSCs treated with DOX at D2-D5.

**[0019]** FIGS. 8A-8C provide results of an in-depth analysis of D8 CD43+ subsets collected from cultures without and with D2-8 DOX using t-distributed stochastic neighbor embedding algorithm (t-SNE). FIG. 8A is an overlaid t-SNE plot of No DOX and DOX2-8 CD43+ subsets at D8. Each dot represents one cell. FIG. 8B shows that No DOX or DOX2-8 samples created single tSNE with single surface marker in each plot. Expression levels of marker defines with a continuous grayscale scale. FIG. 8C shows tSNE maps generated for the No DOX or DOX2-8 subsets. Each shade indicates No DOX or DOX2-8 cells at D8 CD43+. Each dot represents one cell and cells are shaded according to their assigned subsets. Cell populations were defined by the manual gating strategy.

**[0020]** FIGS. 9A-9C show the effect of SOX18 overexpression on development of megakaryocytes and hemangioblasts (HBs). FIG. 9A show counter plots demonstrating expression megakaryocytic markers in megakaryocyte differentiation cultures initiated using CD43+ cells collected from No DOX and DOX2-8 cultures at D8. FIG. 9B is a graphical representation of the percentage of megakaryocytic cells generated in these cultures. The results are presented as mean  $\pm$ SD, n=3. \*\* p<0.01 and \*\*\*p<0.001, t-test. FIG. 9C is a graphical representation of the HB-CFC potential of iSOX18 cells. Cells were collected from D3 differentiation without DOX or with D2-3 DOX treatment. The graph shows number of HB colonies per 10,000 cells collected on D3 of differentiation. The results are presented as the mean  $\pm$ SD, n=2. \*\*p<0.01, t-test.

**[0021]** FIGS. 10A-10D illustrate the impact of SOX18 overexpression on cell cycle and apoptosis. FIG. 10A shows an analysis of gene enrichment in GO "Apoptosis". FIG. 10B shows the analysis of gene enrichment in GO "Regulation of Cell Population Proliferation" gene sets in indicated cell subsets in DOX-treated and untreated cultures. Shaded areas indicates significant enrichment in DOX-treated over untreated cultures. FIG. 10C show representative dot plots of flow cytometric analysis of cell cycle in the D5 HE in DOX and No DOX iSOX18 hPSC cultures. FIG.

10D is a bar graph showing the means  $\pm$  SD (D) of duplicated experiments of the flow cytometric analysis of cell cycle in the D5 HE in DOX and No DOX iSOX18 hPSC cultures. \*\*\*p<0.001, and \*\*\*\*p<0.0001, one-way ANOVA Dunnett's multiple comparisons test.

#### DETAILED DESCRIPTION OF THE DISCLOSURE

**[0022]** The present disclosure is based, at least in part, on experimental demonstration that forced SOX18 expression in cells in the hematopoietic lineage preferentially promotes formation of progenitors of natural killer cells and megakaryocytes.

**[0023]** For the purposes of promoting an understanding of the principles of the disclosure, reference will now be made to embodiments and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of the disclosure is thereby intended, such alteration and further modifications of the disclosure as illustrated herein, being contemplated as would normally occur to one skilled in the art to which the disclosure relates.

#### Definitions

**[0024]** As used in the specification, articles "a" and "an" are used herein to refer to one or to more than one (i.e., at least one) of the grammatical object of the article. By way of example, "an element" means at least one element and can include more than one element.

**[0025]** "About" is used to provide flexibility to a numerical range endpoint by providing that a given value can be "slightly above" or "slightly below" the endpoint without affecting the desired result. The term "about" in association with a numerical value means that the numerical value can vary by plus or minus 5% or less of the numerical value.

**[0026]** Throughout this specification, unless the context requires otherwise, the word "comprise" and "include" and variations (e.g., "comprises," "comprising," "includes," "including") will be understood to imply the inclusion of a stated component, feature, element, or step or group of components, features, elements, or steps but not the exclusion of any other integer or step or group of integers or steps.

**[0027]** As used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations where interpreted in the alternative ("or").

**[0028]** Recitation of ranges of values herein are merely intended to serve as a succinct method of referring individually to each separate value falling within the range, unless otherwise indicated herein. Furthermore, each separate value is incorporated into the specification as if it were individually recited herein. For example, if a range is stated as 1 to 50, it is intended that values such as 2 to 4, 10 to 30, or 1 to 3, for example, are expressly enumerated in this disclosure. These are only examples of what is specifically intended, and all possible combinations of numerical values between and including the lowest value and the highest value enumerated are to be considered to be expressly stated in this disclosure.

**[0029]** The term "contacting" includes the physical contact of at least one substance to another substance.

**[0030]** As used herein, "treatment" refers to the clinical intervention made in response to a disease, disorder, or physiological condition of the subject or to which a subject

can be susceptible. The aim of treatment includes the alleviation or prevention of symptoms, slowing or stopping the progression or worsening of a disease, disorder, or condition and/or the remission of the disease, disorder, or condition.

**[0031]** The terms “effective amount” or “therapeutically effective amount” refer to an amount sufficient to effect beneficial or desirable biological and/or clinical results. In other words, a “therapeutically effective” amount is an amount that will provide some alleviation, mitigation, or decrease in at least one clinical symptom in the subject.

**[0032]** The terms “express” or “expression” refer to transcription and translation of a nucleic acid coding sequence resulting in production of the encoded polypeptide. “Express” or “expression” also refers to antigens that are expressed on cell surfaces.

**[0033]** Unless otherwise defined, all technical terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art to which this disclosure belongs.

**[0034]** As used herein, the term “subject” refers to both human and nonhuman animals. The term “nonhuman animals” of the disclosure includes all vertebrates, e.g., mammals and non-mammals, such as nonhuman primates, sheep, dog, cat, horse, cow, chickens, amphibians, reptiles, and the like. The human subject can be of any age (e.g., an infant, child, or adult).

**[0035]** The term “construct” refers to an artificially-designed segment of DNA that can be used to incorporate genetic material into a target cell (e.g., an hPSC).

**[0036]** The term “sequence identity” refers to the number of identical or similar nucleotide bases on a comparison between a test and reference oligonucleotide or nucleotide sequence. Sequence identity can be determined by sequence alignment of a first nucleic acid sequence to identify regions of similarity or identity to second nucleic acid sequence. As described herein, sequence identity is generally determined by alignment to identify identical residues. Matches, mismatches, and gaps can be identified between compared sequences by techniques known in the art. Alternatively, sequence identity can be determined without taking into account gaps as the number of identical positions/length of the total aligned sequence  $\times 100$ . In one embodiment, the term “at least 90% sequence identity to” refers to percent identities from 90 to 100%, relative to the reference nucleotide sequence. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplary purposes a test and reference polynucleotide sequence length of 100 nucleotides are compared, no more than 10% (i.e., 10 out of 100) of the nucleotides in the test oligonucleotide differ from those of the reference oligonucleotide. Differences are defined as nucleic acid substitutions, insertions, or deletions.

**[0037]** Any appropriate method can be used to detect expression of biological markers characteristic of cell types described herein. For example, the presence or absence of one or more biological markers can be detected using, for example, RNA sequencing (e.g., RNA-seq), immunohistochemistry, polymerase chain reaction, quantitative real time PCR (qRT-PCR), or other technique that detects or measures gene expression. RNA-seq is a high-throughput sequencing technology that provides a genome-wide assessment of the RNA content of an organism, tissue, or cell. Alternatively, or additionally, one can detect the presence or absence of, or measure the level of, one or more biological markers of

HPCs using, for example, Fluorescence in situ Hybridization (FISH; see W098/45479 published October 1998), Southern blotting, Northern blotting, or polymerase chain reaction (PCR) techniques, such as qRT-PCR. In exemplary embodiments, a cell population obtained according to a method provided herein is evaluated for expression (or the absence thereof) of biological markers of HPCs such as CD34, CD45, CD43, and CD90. Quantitative methods for evaluating expression of markers at the protein level in cell populations are also known in the art.

**[0038]** The term “genetically engineered” as used herein refers to cells that have been manipulated using biotechnology to change the genetic makeup of the cells, including the transfer of genes within and across species boundaries to produce improved or non-naturally occurring cells. A human pluripotent stem cell, hemogenic endothelium, megakaryocytes or NK cells that contains an exogenous, recombinant, synthetic, and/or otherwise modified polynucleotide is considered to be a genetically engineered cell and, thus, non-naturally occurring relative to any naturally occurring counterpart. In some cases, genetically engineered cells contain one or more recombinant nucleic acids. In other cases, genetically engineered cells contain one or more synthetic or genetically engineered nucleic acids (e.g., a nucleic acid containing at least one artificially created insertion, deletion, inversion, or substitution relative to the sequence found in its naturally occurring counterpart). Procedures for producing genetically engineered cells are generally known in the art, for example, as described in Sambrook et al, *Molecular Cloning, A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (2012) and Doudna et al., *CRISPR-Cas, A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (2016).

**[0039]** A genetically engineered cell can be a cell that has been modified using a gene editing technique. Gene editing refers to a type of genetic engineering in which DNA is inserted, deleted, modified, or replaced in the genome of a living cell. In contrast to other genetic engineering techniques that can randomly insert genetic material into a host genome, gene editing can target the insertions to site specific locations (e.g., AAVS1 alleles). Examples of gene editing techniques including, but are not limited to, restriction enzymes, zinc finger nucleases, TALENs, and CRISPR-Cas9.

**[0040]** A genetically engineered cell can be a stem cell (e.g., a human pluripotent stem cell) or any of their differentiated progeny cells (e.g., mesoderm cells, hemangioblast cells, hemogenic endothelium cells, hematopoietic progenitor cells, megakaryocytes and NK cells) that have been modified to express. Any of the cells described herein can be genetically engineered. In some embodiments, a genetically engineered cell refers to a cell that is differentiated from a cell that has been genetically engineered.

**[0041]** The term “tumor cell” as used herein refers to abnormal cells that divide continuously. In some embodiments, the tumor cell is a solid tumor cell. A solid tumor is an abnormal mass of cells that typically does not contain cysts or a liquid area. Examples of solid tumors include, but are not limited to, sarcomas and carcinomas. Cancers of the blood (e.g., leukemias) typically do not form solid tumors. In some embodiments, the “tumor cell” is not a blood cancer cell. “Tumor cells” as used herein refers to a group of tumor cells and/or a single tumor cell.



[0042] Human pluripotent stem cells (hPSCs), either embryonic or induced, provide access to the earliest stages of human development and offer a platform on which to derive a large number of hematopoietic progenitor cells or blood cells for cellular therapy and tissue engineering. Accordingly, the methods provided herein can comprise differentiating human pluripotent stem cells under conditions that promote differentiation of mesodermal cells (e.g., arterial endothelial cells) into hematopoietic progenitor cells into megakaryocytes and NK cells.

[0043] Methods of differentiating hPSCs into progenitor mesoderm, hemogenic endothelium and hematopoietic progenitors are known in the art. In exemplary embodiments the culture medium to be used in any of the above-described differentiation methods comprises an IF9S medium, as described herein. In one embodiment, the IF9S medium to be used is the IF9S medium having the formulation set forth in Table 2. In some embodiments, any of the above-referenced cells (e.g., human pluripotent stem cells) are cultured on Tenascin C. In some embodiments, any of the referenced cells are seeded on a substrate treated with an amount of Tenascin-C sufficient to adhere 10,000 cells/cm<sup>2</sup> to the substrate. In some embodiments, the Tenascin-C to be used is human Tenascin C. In some embodiments, the substrate is treated with Tenascin C at a concentration of at least about 0.25 µg/cm<sup>2</sup> to 1 µg/cm<sup>2</sup>, e.g., 0.4 µg/cm<sup>2</sup>, 0.5 µg/cm<sup>2</sup>, 0.7 µg/cm<sup>2</sup>, 0.8 µg/cm<sup>2</sup>, or another concentration from at least about 0.25 µg/cm<sup>2</sup> to 1 µg/cm<sup>2</sup>. In a preferred embodiment, the cells are cultured on plates coated with Collagen IV, as described in Uenishi et al., 2014, *Stem Cell Reports* 3: 1073-1084 and U.S. Pat. No. 9,938,499.

[0044] In some embodiments, in the cell culture medium to be used in the differentiation methods set forth herein, the concentration of: BMP4 is about 50 ng/ml to about 250 mg/ml; Activin A is about 10 ng/ml to about 15 ng/ml; FGF2 is about 10 ng/ml to about 50 ng/ml; LiCl is about 1 mM to about 2 mM; VEGF is about 20 ng/ml to about 50 ng/ml; SCF is about 50 ng/ml to about 100 ng/ml; TPO is about 50 ng/ml to about 100 ng/ml; IL-6 is about 50 ng/ml to about 100 ng/ml, and IL-3 is about 5 ng/ml to about 15 ng/ml.

[0045] In some embodiments, any of the above-referenced cells are cultured in a xeno-free cell culture medium. Of central importance for clinical therapies is the absence of xenogenic materials in the derived cell populations, i.e., no non-human cells, cell fragments, sera, proteins, and the like. Preferably, the present invention arrives at xenogen-free differentiated cells by use of Tenascin C or Collagen IV as a platform, which essentially replaces contact with OP9 cells used in earlier differentiation systems. In addition, the media disclosed herein are chemically defined and, in some embodiments, are made xeno-free, and incorporate human proteins, which can be produced using recombinant technology or derived from placenta or other human tissues in lieu of animal-derived proteins. In some embodiments, all proteins added to the medium are recombinant proteins.

[0046] As used herein, the term “mesoderm cell” refers to a cell having mesoderm-specific gene expression, capable of differentiating into a mesodermal lineage such as bone, muscle such as cardiac muscle, skeletal muscle, and smooth muscle (e.g., of the gut), connective tissue such as the dermis and cartilage, kidneys, the urogenital system, blood or hematopoietic cells, heart, and vasculature. Mesoderm-spe-

cific biomarkers include Brachyury (7). Culturing can take place on any appropriate surface (e.g., in two-dimensional or three-dimensional culture).

[0047] Medium and substrate conditions for culturing pluripotent stem cells, as used in the methods described herein, are well known in the art. In some cases, pluripotent stem cells to be differentiated according to the methods disclosed herein are cultured in mTESR-1 medium (Stem-Cell Technologies, Inc., Vancouver, British Columbia.), E8 medium, or Essential 8 medium (Life Technologies, Inc.) on a MATRIGEL™ substrate (BD Biosciences, NJ) or Vitronectin (Life Technologies) according to the manufacturer's protocol.

[0048] As used herein, the term “albumin-free conditions” indicates that the culture medium used contains no added albumin in any form including, without limitation, Bovine Serum Albumin (BSA), any form of recombinant albumin, or any other animal albumin.

[0049] As used herein, the terms “chemically-defined medium” and “chemically-defined culture medium” also refer to a culture medium containing formulations of fully disclosed or identifiable ingredients, the precise quantities of which are known or identifiable and can be controlled individually. As such, a culture medium is not chemically-defined if (1) the chemical and structural identity of all medium ingredients is not known, (2) the medium contains unknown quantities of any ingredients, or (3) both. Standardizing culture conditions by using a chemically-defined culture medium minimizes the potential for lot-to-lot or batch-to-batch variations in materials to which the cells are exposed during cell culture. Accordingly, the effects of various differentiation factors are more predictable when added to cells and tissues cultured under chemically-defined conditions.

[0050] As used herein, the term “serum-free” refers to cell culture materials that do not contain serum or serum replacement, or that contains essentially no serum or serum replacement. For example, an essentially serum-free medium can contain less than about 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% serum. “Serum free” also refers to culture components free of serum obtained from animal (e.g., fetal bovine) blood or animal-derived materials, which is important to reduce or eliminate the potential for cross-species viral or prion transmission. For avoidance of doubt, serum-containing medium is not chemically-defined.

[0051] Human pluripotent stem cells (e.g., human ESCs or iPS cells) can be cultured in the absence of a feeder layer (e.g., a fibroblast feeder layer), a conditioned medium, or a culture medium comprising poorly defined or undefined components. As used herein, “feeder-free” refers to culture conditions that are substantially free of a cell feeder layer. Cells grown under feeder-free conditions can be grown on a substrate, such as a chemically-defined substrate, and/or grown as an adherent culture. Suitable chemically-defined substrates include vitronectin.

[0052] A method of producing a hematopoietic progenitor cell can comprise culturing human pluripotent stem cells in a serum-free, albumin-free, chemically-defined culture medium that promotes differentiation to mesoderm. In this manner, pluripotent stem cell-derived mesodermal cells are differentiated according to the HPC differentiation methods provided herein, thus producing pluripotent stem cell-derived HPCs.

**[0053]** As used herein, “pluripotent stem cells” appropriate for use according to a method of the invention are cells having the capacity to differentiate into cells of all three germ layers. Suitable pluripotent cells for use herein include human embryonic stem cells (hESCs) and human induced pluripotent stem (iPS) cells. As used herein, “embryonic stem cells” or “ESCs” mean a pluripotent cell or population of pluripotent cells derived from an inner cell mass of a blastocyst. See Thomson et al., *Science* 282:1145-1147 (1998). These cells can express Oct-4, SSEA-3, SSEA-4, TRA-1-60, and TRA-1-81. Pluripotent stem cells appear as compact colonies comprising cells having a high nucleus to cytoplasm ratio and prominent nucleolus. ESCs are commercially available from sources such as WiCell Research Institute (Madison, WI).

**[0054]** As used herein, “induced pluripotent stem cells” or “iPS cells” refers to pluripotent cell or population of pluripotent cells that can vary with respect to their differentiated somatic cell of origin, that can vary with respect to a specific set of potency-determining factors and that can vary with respect to culture conditions used to isolate them, but nonetheless are substantially genetically identical to their respective differentiated somatic cell of origin and display characteristics similar to higher potency cells, such as ESCs, as described herein. See, e.g., Yu et al., *Science* 318:1917-1920 (2007).

**[0055]** Induced pluripotent stem cells exhibit morphological properties (e.g., round shape, large nucleoli, and scant cytoplasm) and growth properties (e.g., doubling time of about seventeen to eighteen hours) akin to ESCs. In addition, iPS cells express pluripotent cell-specific markers (e.g., Oct-4, SSEA-3, SSEA-4, Tra-1-60, or Tra-1-81, but not SSEA-1). Induced pluripotent stem cells, however, are not immediately derived from embryos. As used herein, “not immediately derived from embryos” means that the starting cell type for producing iPS cells is a non-pluripotent cell, such as a multipotent cell or terminally differentiated cell, such as somatic cells obtained from a post-natal individual.

**[0056]** As set forth herein, NK cells and megakaryocytes can be preferentially produced, compared with T cells, by inducing expression in SOX18 in progenitor cells. Generally, preferential differentiation can be achieved by overexpressing SOX18 at a specific time (i.e., starting at day 2) and for a defined period (i.e., 2-8 days) during differentiation. SOX18 is a member of the Sry-related high mobility group domain (SOX) family of transcription factors and is key developmental regulator of endothelial and hematopoietic lineages. In the methods of the present invention, differentiating hPSCs are forced to overexpress SOX18 during the mesoderm differentiation by introducing an inducible SOX18 transgene into the population of progenitor cells.

**[0057]** The SOX18 transgene used with the present invention can comprise any nucleic acid sequence encoding the SOX18 protein. For example, the SOX18 transgene can be obtained by amplifying the SOX18 gene sequence from the genomic locus in human cells or by amplifying SOX18 mRNA from hPSCs differentiated into endothelial and blood cells and converting it into cDNA. Alternatively, genomic DNA or cDNA clones can be obtained commercially (e.g., from Sino Biological, Origene, or IDT). In some embodiments, the transgene comprises SEQ ID NO:16, a cDNA sequence encoding the human SOX18 protein (having an amino acid sequence set forth in SEQ ID NO: 17).

**[0058]** In some embodiments, the SOX18 transgene further comprises a vector sequence that can be used to drive the expression of the SOX18 transgene within the cells. In these embodiments, the transgene is introduced by into the population of hPSCs by transducing the cells with said vector. As used herein, the term “vector” refers to a nucleic acid molecule capable of propagating another nucleic acid to which it is linked. The term includes the vector as a self-replicating nucleic acid structure as well as the vector incorporated into the genome of a host cell into which it has been introduced. Certain vectors are capable of directing the expression of nucleic acids to which they are operatively linked. Such vectors are referred to herein as “expression vectors”.

**[0059]** Vectors suitable for use with the present invention comprise a nucleotide sequence encoding a SOX18 transgene and a heterogeneous sequence necessary for proper propagation of the vector and expression of the encoded polypeptide. The heterogeneous sequence (i.e., sequence from a different species than the transgene) can comprise a heterologous promoter or heterologous transcriptional regulatory region that allows for expression of the polypeptide. Suitable vectors for the expression of the SOX18 transgene include plasmids and viral vectors. In a preferred embodiment, the vector comprises heterologous sequence that allows the transient and/or Inducible expression of the encoded SOX18 protein.

**[0060]** In some embodiments, the vector includes a transposase system, such as the PiggyBac transposon system (see Examples). The PiggyBac transposon is a TTAA-specific mobile genetic element that efficiently transposes between vectors and chromosomes via a “cut and paste” mechanism. PiggyBac transposase recognizes transposon-specific inverted terminal repeat sequences (ITRs) and moves the intervening contents to a TTAA insertion site in a chromosome or another vector. Thus, inserting a gene of interest between two ITRs in a transposon vector allows one to efficiently insert the gene into a target genome. Other suitable transposase systems for use with the present invention include, for example, Sleeping Beauty.

**[0061]** In other embodiments, the vector is a plasmid, a viral vector, a cosmids, or an artificial chromosome. Suitable plasmids include, for example, *E. coli* cloning vectors. Many suitable viral vectors are known in the art and include, but are not limited to, an adenovirus vector; an adeno-associated virus vector; a pox virus vector, such as a fowlpox virus vector; an alpha virus vector; a baculoviral vector; a herpes virus vector; a retrovirus vector, such as a lentivirus vector; a Modified Vaccinia virus Ankara vector; a Ross River virus vector; a Sindbis virus vector; a Semliki Forest virus vector; and a Venezuelan Equine Encephalitis virus vector. In one particular embodiment, the vector comprises SEQ ID NO:15.

**[0062]** In some embodiments, the vector is an expression vector that comprises a promoter that drives the expression of the SOX18 transgene, preferably transient or inducible expression of the SOX18 transgene. As used herein, the term “promoter” refers to a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a coding sequence. Although promoters are most commonly found immediately upstream of a coding sequence, they can also be found downstream of or within the coding sequence. Promoters can be derived in their entirety from a native gene or can be composed of multiple

elements, including elements derived from promoters found in nature or elements comprising synthetic DNA sequences. It is understood by those skilled in the art that different promoters can direct the expression of a gene in different tissues or cell types, at different stages of development, or in response to different environmental conditions. Preferably, the promoters used with the present invention are inducible promoters. An “inducible promoter” is a promoter that is activated (i.e., initiates transcription) only in the presence of a particular molecule. Inducible promoters allow tight control the expression of a transgene within cells. Many suitable inducible expression systems are known in the art and include, for example, Tet-On gene expression systems that allow one to induce the expression of a gene by administering tetracycline (Tc) or tetracycline-derivatives like doxycycline (DOX). Suitable Tet-On systems for use with the present invention include, without limitation, Tet-On Advanced and Tet-On 3G. Tet-On systems utilize several promoters, including both minimal promoters (e.g., CMV) flanked by a tetracycline response element (TRE) and engineered Tet-inducible promoters (e.g., TRE2 and TREtight). For instance, in the Examples, the SOX18 transgene is inserted (i.e., via conventional cloning methods) downstream of the doxycycline-inducible TREtight promoter within a vector. This vector was introduced into the hPSCs, allowing the inventors to induce expression of the SOX18 transgene at the desired stage of differentiation by adding doxycycline to the cell culture to activate expression from the TREtight promoter. Those of skill in the art are aware of many additional inducible gene expression systems, including both chemical-inducible and temperature-inducible systems. Other suitable inducible gene expression systems for use with the present invention include, without limitation, the glucocorticoid-responsive mouse mammary tumor virus promoter (MMTVprom), the tamoxifen-responsive hormone-binding domain of the estrogen receptor (ERTAM), the ecdysone-inducible promoter (EcP), heat shock inducible promoters (e.g., Hsp70 or Hsp90- derived promoters), and the T7 promoter/T7 RNA polymerase system (T7P). The SOX18 transgene can be introduced into the hPSCs using any suitable method, for example by transfection or transduction. In one embodiment, the transgene is introduced by transducing the hPSCs with a vector comprising the SOX18 transgene. In another embodiment, the hPSCs are transduced with an exogenous SOX18 modified mRNA (mmRNA). In yet another embodiment, the hPSCs are transduced with the SOX18 protein. Typically, mmRNAs comprise (i) a 5' synthetic cap for enhanced translation; (ii) modified nucleotides that confer RNase resistance and an attenuated cellular interferon response, which would otherwise greatly reduce translational efficiency; and (iii) a 3' poly-A tail.

**[0063]** As provided herein, SOX18 can be enforced in progenitor cells by introduction of an inducible recombinant genetic construct encoding human SOX18 having a nucleotide sequence set forth herein. As used herein, the term “enforced” when used with regard to SOX18 expression in a progenitor cell for NK cells or megakaryocytes produced according to the methods disclosed herein can be understood by the skilled worker to mean that SOX18 expression in the cell is increased over endogenous SOX18 expression in the cell by introduction therein of an exogenous SOX18-encoding construct wherein SOX18 expression is controlled by an inducible promoter. Endogenous SOX18 can be enforced, inter alia, using reagents and methods known in the art,

including but not limited to disclosure in U.S. Patent Application Nos. U.S. 2018/0010124 and 2018/0142207, the disclosures of which are expressly incorporated by reference herein. Such techniques and methodologies can comprise enforced expression of SOX18 in said cells.

**[0064]** As used herein, the term “forced expression” refers to inducing an increase in the level of a protein of interest (e.g., a transcription factor) in a population of host cells, e.g., hPSCs. Forced expression can include one or more of the following in any combination: introducing exogenous nucleic acids encoding the protein of interest (e.g., by viral transduction, plasmid expression vector transfection, or modified mRNA transfection); protein transduction; genomic modification of a host cell, e.g., replacing a promoter to increase the expression of an endogenous (native) gene; and contacting host cells with a small molecule that induces increased expression of an endogenous protein.

#### Pharmaceutical Formulations and Methods of Treatment

**[0065]** Administration of a therapeutically effective amount of NK cells or megakaryocytes provided herein into the recipient subject is generally effected using methods well known in the art, and usually involves directly injecting or otherwise introducing a therapeutically effective dose of NK progenitor cells into the subject using clinical tools known to those skilled in the art (e.g., U.S. Pat. Nos. 6,447,765; 6,383,481; 6,143,292; and 6,326,198). For example, introduction of NK cells or megakaryocytes of the present invention can be injected locally or systemically via intravascular administration, such as intravenous, intramuscular, or intra-arterial administration, intraperitoneal administration, and the like. Cells can be injected into an infusion bag (e.g., Fenwal infusion bag (Fenwal, Inc.)) using sterile syringes or other sterile transfer mechanisms. The cells can then be immediately infused via IV administration over a period of time, such as 15 minutes, into a free flow IV line into the patient. In some embodiments, additional reagents such as buffers or salts are provided to the recipient subject concurrently with the cells.

**[0066]** Various exemplary embodiments of compositions and methods according to this invention are now described in the following non-limiting Examples. The Examples are offered for illustrative purposes only and are not intended to limit the scope of the present invention in any way. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and the following examples and fall within the scope of the appended claims.

#### EXAMPLES

**[0067]** The Examples set forth herein incorporate and rely on certain experimental and preparatory methods and techniques performed as exemplified herein.

#### Materials and Methods

**[0068]** Cell Culture. Wild type H1 and human pluripotent stem cells (hPSCs) were obtained from WiCell (Madison, WI).

**[0069]** Construction of vectors, generation, and validation of iSOX18 hPSC line. A doxycycline (DOX)-inducible SOX18 H1 hESC line was generated using the PiggyBac

system disclosed in Park et al., 2018, *Curr. Protoc. Stem Cell Biol.* 47: e63 and illustrated in FIG. 1. Human SOX18 CDS was cloned into PiggyBac transposon vector (Transposagen) downstream of TREtight promoter of pTRE-P2A-Venus-rpEFla-Zeo plasmid, and co transfected with pEF1 $\alpha$ -M2rtTA-T2A-Puro and transposase plasmid into H1 cells using human stem cell nucleofector kit 2 (Lonza). Cells were selected in Zeocin (0.5  $\mu$ g/ml, ThermoFisher) and Puromycin (0.5  $\mu$ g/ml, Sigma) for 10 days and resistant clones screened by Venus expression with DOX (Sigma) treatment.

**[0070]** hPSC maintenance and hematopoietic differentiation. hPSCs were maintained and passaged on Matrigel in mTeSR1 media (WiCell). Hematopoietic differentiation was performed on collagen IV (ColIV)-coated plate in chemically defined serum-free medium as previously described as disclosed in Uenishi et al., 2014, *Stem Cell Reports* 3: 1073-1084. The iSOX18 H1 line from hPSCs (H1 hESC line from WiCell) was maintained and passaged on Matrigel in mTeSR1 media (WiCell). The cell lines were differentiated on a collagen IV (ColIV)-coated plate. Briefly, to initiate differentiation, cells were plated at 5,000 cells/cm<sup>2</sup> onto 6 well plates with E8 media and 10  $\mu$ M Rock inhibitor (Y-27632, Cayman Chemicals). This media was changed the following day to IF9S media with 50 ng/ml FGF2 (PeproTech), 50 ng/ml BMP4 (PeproTech), 15 ng/ml Activin A (PeproTech), and 2 mM LiCl (Sigma), and cells were cultured in hypoxia (5% CO<sub>2</sub>, 5% O<sub>2</sub>). On day 2, the media was changed to IF9S media with 50 ng/ml FGF2, 50 ng/ml VEGF (PeproTech), and 2.5  $\mu$ M TGF- $\beta$ 3 inhibitor (SB-431542, Cayman), and cells were cultured in hypoxia (5% CO<sub>2</sub>, 5% O<sub>2</sub>). On days 4 and 6, the media was changed to IF9S media with 50 ng/ml FGF2, 50 ng/ml VEGF, 50 ng/ml TPO (PeproTech), 50 ng/ml IL-6 (PeproTech), 20 ng/ml SCF (PeproTech), and 10 ng/ml IL-3 (PeproTech), and cells were cultured in normoxia (20% CO<sub>2</sub>, 5% O<sub>2</sub>).

**[0071]** Isolation and culture of hemogenic endothelium and hematopoietic progenitors. CD31<sup>+</sup> cells from day 4 (D4) cultures of iSOX18 cells were isolated by magnetic activated cell sorting (MACS) using CD31 antibodies. In D4 differentiation cultures almost all CD31<sup>+</sup> cells co-express VE-cadherin. Selected cells were plated on OP9 or DLL4-OP9 in 10%  $\alpha$ -MEM with 10% FBS (Hyclone) with TPO, SCF (50 ng/ml), IL-6 (20 ng/ml), IL-3 and FLT3L (10 ng/ml). Medium was changed on the following day and extra medium was added on the 3rd day of OP9 co-culture. After 5 days in secondary culture, cells were collected and assessed for colony-forming cells (CFCs), T, and NK potential. Hematopoietic precursor cells (HPs) generated from iSOX18 cells were collected at day 8 (D8) of differentiation. Four different subsets of CD34<sup>+</sup>CD43<sup>+</sup> population CD235a/CD41a-CD45<sup>-</sup> (P1 subset), CD235a/CD41a+CD45<sup>-</sup> (P2), CD235a/CD41a+CD45<sup>+</sup> (P3) and CD235a/CD41a-CD45<sup>+</sup> (P4) were isolated using a MA900 cell sorter (Sony Biotechnology) and cultured on OP9-DLL4 in NK cell differentiation conditions.

**[0072]** Hemangioblast (HB)-CFC and hematopoietic CFC assay. HB-CFC were detected as described in Vodyanik et al., 2010, A mesoderm-derived precursor for mesenchymal stem and endothelial cells, *Cell Stem Cell* 7: 718-729. Hematopoietic CFCs were detected using serum containing H4435 MethoCult (Stem Cell Technologies). HB-CFCs were detected using a semisolid colony-forming serum-free medium (CF-SFM) containing 40% ES-Cult M3120 meth-

ylcellulose (2.5% solution in IMDM, Stem Cell Technologies), 25% StemSpan serum-free expansion medium (SFEM, Stem Cell Technologies), 25% human endothelial serum-free medium (ESFM, ThermoFisher), 10% BIT 9500 supplement (Stem Cell Technologies), GlutaMAX (1/100 dilution, ThermoFisher), Ex-Cyte (1/1000 dilution, Millipore), 100 mM MTG, 50 mg/mL ascorbic acid and 20 ng/mL FGF (PeproTech)

**[0073]** Megakaryocyte differentiation. Floating hematopoietic cells from iSOX18 hPSC cultures were collected at D8 of differentiation and cultured in StemSpan serum-free expansion medium (SFEM, Stem Cell Technologies) with 20 ng/ml SCF, TPO, and IL-11 on an ultra-low attachment 6-well plate for 5 days. Fresh media (2 mL) was added every 2 days. All cytokines were purchased from PeproTech.

**[0074]** NK-cell differentiation. For NK cell differentiation, floating hematopoietic cells from D8 primary cultures or day 4+5 secondary OP9/OP9-DLL4-cocultures were cultured in  $\alpha$ -MEM (Invitrogen) with 20% FBS (Hyclone), 25 ng/ml SCF, 5 ng/ml FLT3L, IL-3 and IL-7 and 10 ng/ml IL-15 (PeproTech) on OP9-DLL4 for 5 days. Cells were then cultured in the same media without IL-3 for 3-4 weeks. Cells were passaged weekly onto fresh DLL4-OP9 cells and analyzed by flow cytometry for NK cell surface markers after 3-4 weeks.

**[0075]** Functional analysis of NK cells. To assess cytotoxicity, CD56<sup>+</sup> cells were isolated using a MA900 cell sorter and incubated with K562-GFP target cells for 4 hours at 37° C., at effector:target (E:T) ratios of 1:1, 2.5:1, and 5:1, in 96-well plates. Cells were collected in FACS buffer and stained using 7-Aminoactinomycin D (7-AAD) and Annexin V (BD). Specific killing was calculated by subtracting spontaneous K562 death (7-AAD<sup>+</sup> cells in no effector control). Production of IFN $\gamma$  and CD107a expression in isolated CD56<sup>+</sup> cells was assessed after incubation with Phorbol 12-myristate 13-acetate (PMA) and ionomycin (1:500) (BioLegend) for 4 hours. Brefeldin A (1:1000; ThermoFisher) was added at the beginning of the stimulation. Cells were washed with FACS buffer and stained with Live/dead violet 540 (TONBO Bioscience) with CD107a antibody (BD). Cells were treated with fixation/permeabilization buffer (eBioscience) and stained for intracellular IFN $\gamma$  (BD).

**[0076]** T-cell differentiation. For T cell differentiation, floating hematopoietic cells from D8 of primary differentiation cultures or day 4+5 secondary OP9/OP9-DLL4-cocultures were cultured in  $\alpha$ -MEM (Invitrogen) with 20% FBS (Hyclone), 10 ng/ml SCF, 5 ng/ml FLT3L and IL-7 (PeproTech) on OP9-DLL4 for 3 weeks. Cells were passaged weekly onto fresh OP9-DLL4 cells. Cells were analyzed by flow cytometry for T cell surface markers after 21 days. All cytokines were purchased from PeproTech.

**[0077]** Flow cytometry and t-distributed stochastic neighbor embedding algorithm (tSNE) analysis. Flow cytometric analysis was performed using antibodies listed in Table 1 with MACSQuant Analyzer 10 (Miltenyi Biotech) and FlowJo software (FlowJo LLC). For tSNE analysis, individual DOX2-8 and No DOX fcs files were imported into FlowJo to exclude doublets, debris, and dead cells. A subset or 11,000 CD43<sup>+</sup> cells were selected for each sample and concatenated. To generate a tSNE map, the concatenated data were analyzed with the parameters 30 perplexity, 550 iteration number and 1540 learning rate (Eta). Concatenated cells were divided manually into No DOX or DOX2-8 and

cell subsets were defined by the manual gating. First, cells CD34<sup>+</sup> or CD34<sup>-</sup> were gated and then gated into CD235a/CD41a<sup>hi</sup>CD45<sup>-</sup>, CD235a/CD41a<sup>med</sup>CD45<sup>-</sup>, CD235a/CD41a<sup>-</sup>CD45<sup>-</sup>, CD235a/CD41a<sup>-</sup>CD45<sup>+</sup>, or CD235a/CD41a<sup>+</sup>CD45<sup>+</sup> cells.

**[0078]** Apoptosis and cell cycle analysis. Apoptosis was detected by flow cytometry using Annexin V (BD). For cell-cycle analysis, D5 cells were incubated in culture medium with bromodeoxyuridine (BrdU) (10  $\mu$ M, BD Pharmingen) for 2 hours and stained with antibodies. For BrdU detection, the BrdU flow kit with 7 AAD was used and performed per the manufacturer's instructions. Fluorescent reagents used for analysis, cell viability, apoptosis, and proliferation are listed in Table 2.

**[0079]** Real Time Quantitative Polymerase Chain Reaction (qPCR). RNA was extracted from D4 CD31<sup>+</sup> cells isolated from control and DOX-treated cultures of iSOX18 of iSOX173 hPSCs using the RNeasy Plus Micro Kit (QIAGEN). RNA was reverse transcribed into cDNA using random hexamer primers (QIAGEN) with SMART MMLV reverse transcriptase (TaKaRa). qPCR was conducted using TB Green Advantage qPCR Premix (TaKaRa). RPL13A was used as the reference gene to normalize the data. Primer sequences are listed in Table 2.

**[0080]** RNA-Seq data processing and analysis. One hundred nanograms of total RNA was used to prepare sequencing libraries following the Ligation Mediated Sequencing (LM-Seq) protocol (see, Gandillet et al., 2009, *Blood* 114: 4813-422) and quantified with a Qubit fluorometer (Life Technologies). Final cDNA libraries were quantitated with the Quant-iT PicoGreen Assay Kit (ThermoFisher Scientific), multiplexed, loaded at a final concentration of 1 nM or 2.5 nM, and sequenced as single reads on the NextSeq 2000 (Illumina), respectively. RNA-seq reads were aligned by STAR (version 2.5.2b) to the human genome (version hg38) with GENCODE basic gene annotations (version 38). Gene expression levels were quantified by RSEM (version 1.3.0), and differential expression was analyzed by edgeR (version 3.34.1). A differentially expressed gene was required to have at least two-fold changes and an adjusted p-value <0.05. Gene set enrichment analysis was performed by fgsea (version 1.18.0) with KEGG gene sets from the Molecular Signatures Database (version 7.1). RNA-seq data has been deposited to GEO with the accession code GSE195670 and a review token as ylfefuwwtvcfnx.

**[0081]** Western Blot. For Western Blot experiment, iSOX18 hPSCs were cultured with (2  $\mu$ g/ml) and without DOX for 24 hours and harvested. In a similar manner, total cells from iSOX18 hPSC differentiation cultures with (D2-D5) and without DOX were collected at day 5 of differentiation. The cells were lysed using Pierce IP lysis buffer with Pierce protease inhibitors. For cell lysate analysis, protein levels were quantified using the Pierce BCA Assay kit (Thermo Fisher, Waltham, MA) and normalized to 8  $\mu$ g of total protein (depending on the individual blot) prior to running on pre-cast 4-12% gradient SDS-PAGE gels and subsequent transfer to PVDF membranes using Bio-Rad Trans-Blot Turbo Transfer System. The membrane was blocked with 5% BSA (Fisher Scientific, BP1600-100) and 5% Difco Skim Milk (BD, 232100) in TBST (1%) for human SOX18 antibody (R&D Systems, 1:1000) and anti-GAPDH (Santa Cruz Biotechnology, 1:5000) probing respectively. The membranes were incubated with primary antibodies overnight at 4° C. after blocking with mild

agitation. The membranes were blotted with their corresponding HRP-linked secondary antibodies at room temperature for one hour. The antibodies were diluted in 1% BSA and 1% milk in TBST for SOX18 and GAPDH detection respectively. 1% TBST was used to wash the membranes for three times at 5 minutes intervals. Sheep and rabbit HRP-linked secondary antibodies were purchased from R&D Systems and Santa Cruz Biotechnology. Images were collected using Bio-Rad ChemiDoc XRS+.

**[0082]** Statistical analysis. Data were analyzed using GraphPad Prism version 9 (GraphPad Software Inc.) and Microsoft Excel (Microsoft Corporation). Tests for statistical significance are listed with each experiment; these included two-sided Student's t-tests for paired analyses and one-way ANOVAs, and two-way ANOVAs for experiments with multiple comparisons of variables or grouped variables, accompanied by the Tukey and Sidak post-hoc test, as inferred to be most appropriate by the software.

#### Example 1

##### SOX18-Enforced Expression Enhances Production of Erythromyeloid Progenitors

**[0083]** To determine the impact of SOX18 overexpression on hematopoietic development in humans, H1 hESC carrying doxycycline (DOX)-inducible SOX18-P2A-Venus were generated (FIGS. 1 and 7) and these cells differentiated under a chemically defined culture system in which all stages of hematopoietic development are temporally, phenotypically, and functionally defined (FIG. 2A), as explained with more specificity in Example 2.

**[0084]** To determine the stages of hematopoietic development sensitive to SOX18 modulation and the optimal duration of SOX18 overexpression to achieve a maximal effect on hematopoietic output, hESC differentiation cultures were treated with DOX at different times and the phenotype and colony-forming cell (CFC) potential of hematopoietic cells collected on day 8 (D8) of differentiation were analyzed (FIG. 2B). As shown in FIGS. 2C-2D, treatment of cultures with DOX from day 2 (D2) through D8 (DOX2-8) resulted in the highest percentages of CD34<sup>+</sup>CD43<sup>+</sup> HPs and CFCs. This treatment also increased the proportion of CD235a/CD41<sup>-</sup>CD45<sup>-</sup> progenitors (D8 P1 population) within the CD33<sup>+</sup>CD43<sup>+</sup> cell population and decreased a proportion of CD235a/CD41a<sup>+</sup>CD45<sup>+</sup> progenitors (D8 P3 population). Although a similar effect on subset composition within CD34<sup>+</sup>CD43<sup>+</sup> progenitors was observed in DOX6-8 and to a lesser degree in DOX4-6 cultures, DOX treatments at these stages of differentiation had a minimal effect on the percentages of CD34<sup>+</sup>CD43<sup>+</sup> cells therein. In contrast, earlier and shorter treatments (DOX2-4) mildly increased CD34<sup>+</sup>CD43<sup>+</sup> cells but had a minimal effect on their composition. These findings suggested that SOX18 has the most profound effect on hematopoietic differentiation when continuously upregulated from D2 through D8 of differentiation. Analysis of colony forming unit (CFC) potential revealed that SOX18 had the most significant effect on GM<sup>-</sup> and E-CFCs. This was the most pronounced in DOX2-6 and DOX2-8 cultures (FIG. 2E). DOX-treated cultures also had a higher megakaryocytic differentiation potential than non-treated controls (FIG. 9A). These observations indicated that SOX18 overexpression predominantly promoted erythromyeloid progenitors.

**[0085]** To visualize the in-depth phenotype of D8 hematopoietic cells from No DOX and DOX2-8 cultures, CD43+ gated cells were analyzed using the t-distributed stochastic neighbor embedding algorithm (tSNE) which showed distinct single cell deposition between No DOX and DOX2-8 (FIG. 8A). tSNE analysis also highlighted a suppression of CD45<sup>+</sup> populations and enrichment of CD45<sup>-</sup> and CD235a/41a<sup>-</sup> populations following DOX treatment. In addition, DOX treatment reduced populations with CD235a/41a<sup>hi/med</sup>CD34<sup>-</sup> phenotype (FIGS. 8B-8C) corresponding to more mature erythromyeloid progenitors. This observation was consistent with prior findings in murine systems, which demonstrates blocking effect of Sox18 on differentiation of ESC<sup>-</sup> and yolk sac-derived HPs (see, Serrano et al., 2010, *Blood* 115: 3895-3898).

#### Example 2

##### SOX18-Enforced Expression Promotes Megakaryocyte Production

**[0086]** To determine the impact of SOX18 overexpression on hematopoietic development in humans, H1 hESCs carrying doxycycline (DOX)-inducible SOX18-P2A-Venus (FIGS. 1 and 7) were generated and these cells differentiated under a chemically defined culture system in which all stages of hematopoietic development are temporally, phenotypically, and functionally defined. In this differentiation system, the most primitive hemogenic cells with FGF2-dependent hemangioblast colony-forming cell (HB-CFCs) potential are detected on day 3 (D3) of differentiation. The first immature/primordial VEC<sup>+</sup>CD43<sup>-</sup>CD73<sup>-</sup>NOTCH1<sup>+</sup> HE cells expressing high levels of HAND1 mesodermal gene and lacking arterial and venous gene expression arise on D4 (D4 HE). Subsequently on D5, HE specifies into DLL4<sup>+</sup>CXCR4<sup>+/+</sup> arterial-type HE with definitive lymphomyeloid potential and DLL4<sup>-</sup> non-arterial-type HE with myeloid-restricted potential (FIG. 2A). D5 DLL4<sup>+</sup>CXCR4<sup>+</sup> HE is highly enriched in lymphoid-myeloid progenitors and expresses the highest levels of HOXA and arterial genes, including SOX17 and NOTCH4, as compared to other DLL4<sup>+</sup> and DLL4<sup>-</sup> HE populations. All hematopoietic progenitors derived from hPSC cultures on D8 of differentiation can be identified by CD43 expression (Vodyanik et al., *Blood* 108:2095-2105 (2006); Choi et al., *Stem Cell* 27:559-567 (2009)) D8 CD34<sup>+</sup>CD43<sup>+</sup> HPs are composed of at least three major subpopulations: (1) CD235a<sup>+</sup>CD41a<sup>+</sup>CD45<sup>-</sup> progenitors enriched in erythro-megakaryocytic cells, (2) CD41a<sup>lo</sup>CD235a<sup>+/+</sup>CD45<sup>+</sup> progenitors with erythro-myeloid potential, and (3) CD235a<sup>-</sup>CD41a<sup>-</sup>CD43<sup>+</sup>CD45<sup>+</sup> multilineage progenitors that are lacking lineage markers, and display CD90<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>-</sup> phenotype (Vodyanik et al., *Blood* 108:2095-2105 (2006); Choi et al., *Stem Cell* 27:559-567 (2009); Suknuntha et al. *Stem Cell Res.* 15:678-693 (2015); Mesquita et al. *Sci. Rep.* 9:6622 (2019)) typical for human hematopoietic stem/progenitor cells (HPSCs) Ivanovs et al., *Stem Cell Rep.* 2:449-456 (2014). In addition, it has been shown that SOX18 expression is initiated on D4 differentiation in HE and remains present in D5 HE and D8 HPs.

**[0087]** To determine the stages of hematopoietic development sensitive to SOX18 modulation and the optimal duration of SOX18 overexpression to achieve a maximal effect on hematopoietic output, hESC differentiation cultures were treated with DOX at different time points and the phenotype and CFC potential of hematopoietic cells collected on D8 of differentiation analyzed (FIG. 2B). As shown in FIGS.

2C-2D, treatment of cultures with DOX from D2 through D8 (DOX2-8) resulted in the highest percentages of CD34<sup>+</sup>CD43<sup>+</sup> HPs and CFCs. This treatment increased proportion of CD235a/CD41a<sup>-</sup>CD45<sup>-</sup> progenitors (D8 P1 population) within CD34<sup>+</sup>CD43<sup>+</sup> population and decreased a proportion of CD235a/CD41a<sup>+</sup>CD45<sup>+</sup> progenitors (D8 P3 population). Although a similar effect on subset composition within CD34<sup>+</sup>CD43<sup>+</sup> progenitors was observed in DOX6-8 and to a lesser degree in DOX4-6 cultures, DOX treatments at these stages of differentiation had a minimal effect on the percentages of CD34<sup>+</sup>CD43<sup>+</sup> cells. In contrast, earlier and shorter treatments (DOX2-4) mildly increased CD34<sup>+</sup>CD43<sup>+</sup> cells but had a minimal effect on their composition. These findings suggested that SOX18 had the most profound effect on CD34<sup>+</sup>CD43<sup>+</sup> HPs and CFCs when continuously upregulated from D2 through D8 of differentiation.

**[0088]** Analysis of CFC potential revealed that SOX18 has the most significant effect on GM- and E-CFCs. This was the most pronounced in DOX2-6 and DOX2-8 cultures (FIG. 2E). Furthermore, DOX treated cultures were found to have a higher megakaryocytic differentiation potential than non-treated controls (FIGS. 9A-9B). These observations indicated that SOX18 overexpression starting from D2 of differentiation predominantly promoted erythromyeloid progenitors.

#### Example 3

##### SOX18-Enforced Expression Suppresses T Cell Differentiation and proMotes NK Cell Production

**[0089]** To evaluate the effect of SOX18 on lymphoid differentiation, cultures were treated with DOX as shown in FIG. 2B and T cell and NK cell potential of HPs generated on day 8 of 9 differentiation were evaluated. Treatment of differentiation cultures with DOX suppressed T cell potential of day 8 HPs. This suppression of T cell potential was more pronounced in cultures treated with DOX from D2 through D8 (FIG. 2F-2G). However, HP from DOX-treated cultures were able to produce CD56<sup>+</sup> NK cells, of which the output increased most dramatically following prolonged SOX18 upregulation during hematopoietic differentiation (DOX2-6 and DOX2-8 cultures; FIGS. 2H-2I). These DOX treatments also increased expression of CD16 by CD56<sup>+</sup> cells, while CD94 expression increased during D2-4 or D2-6 but decreased during D2-8 DOX treatment. Overall these findings suggested that SOX18-enforced expression during day 2-8 (D2-8) of differentiation promotes NK cell development and slightly affects the expression of CD16 and CD94.

#### Example 4

##### SOX18-Enforced Expression Promotes HB-CFCs but Have a Limited Effect on Specification of Hemogenic Endothelium Formation

**[0090]** To determine stages of hematopoiesis mostly affected by SOX18 overexpression, the effect of DOX treatment on the formation of the hemangioblast (HB) colonies and HE, including arterial HE specification was analyzed. Enforced expression of SOX18 on D2 of differentiation resulted in almost 3-fold increase in the numbers of HB colonies (FIG. 9B). However, SOX18 overexpression had a limited effect on HE formation and arterial HE specification on D4 and D5 of differentiation (FIGS. 3A-3F). Although a slight increase in the proportion of VE-cadherin<sup>+</sup> (VEC<sup>+</sup>) endothelial cells in DOX-treated cultures was observed, no significant differences were observed in the

proportion of DLL4<sup>+</sup>CXCR4<sup>+/-</sup> arterial HE in DOX and No DOX cultures. Formation of a DLL4<sup>-</sup>CXCR4<sup>+</sup> population in DOX-treated D4 cultures was also observed. However, transcriptional profiling of DLL4<sup>-</sup>CXCR4<sup>+</sup> and DLL4<sup>-</sup>CXCR4<sup>-</sup>D4 VEC<sup>+</sup> populations revealed no differentially expressed genes (as set forth below).

**[0091]** To assess the effect of SOX18 on HE, D4 HE generated in No DOX and DOX conditions were isolated and cultured on OP9 or OP9-DLL4 in presence or absence of DOX (FIG. 4A). Analysis of CFC potential DOX treatment had the most profound effect on HE cultured on OP9-DLL4. In these cocultures, continued treatment with DOX significantly increased CFC formation, while HE from primary differentiation cultures pretreated with DOX (DOX2-4) without additional DOX treatment during cocultures generated fewer CFCs as compared to No DOX controls (FIG. 4B).

**[0092]** Analysis of T cell potential revealed that HE from DOX2-4 treated cultures demonstrated decreased total T cell output when cultured on OP9 and OP9-DLL4. This effect was more pronounced when DOX treatment was initiated on D4 in coculture of HE with OP9 or OP9<sup>-</sup> DLL4, or in DOX2-8 treatment cultures (FIGS. 4C-4D). In contrast, treatment of HE in OP9 or OP9-DLL4 cultures markedly increased NK cell potential of D8 HPs, while D2-4 DOX treatment of primary differentiation cultures had no effect on NK cells (FIGS. 4E-4F). These findings suggested that SOX18 overexpression predominantly affected lymphoid specification from HE by shifting the balance of NK versus T lymphocyte differentiation potential.

#### Example 5

##### SOX18-Enforced Expression Promotes Expansion of CD34<sup>+</sup>CD43<sup>+</sup>CD235a<sup>+</sup>CD41a<sup>-</sup>CD45<sup>-</sup> (D8 P1) Population Enriched in NK cell Potential

**[0093]** As shown in FIGS. 2C and 2D, D2-D8 SOX18 overexpression significantly promoted development of CD235a<sup>+</sup>CD41a<sup>-</sup>CD45<sup>-</sup> (D8 P1) populations and inhibited development of CD45<sup>+</sup>populations, including CD235a<sup>+</sup>CD41a<sup>+</sup> (D8 P3) and CD235a<sup>+</sup>CD41a<sup>-</sup> (D8 P4) populations within CD34<sup>+</sup>CD43<sup>+</sup> HPs. To define cell populations enriched in NK cells following DOX treatment, major populations of HPs formed on D8 of culture were isolated and their NK cell potential assessed (FIG. 6A). As shown in FIG. 6B and 6C, in control cultures NK cell potential was mostly detected in D8 P1 and P4 populations, while P2 population failed to produce NK cells. Although CD235a<sup>+</sup>CD41a<sup>+</sup>CD45<sup>+</sup> (D8 P3) cells were capable of producing NK cells, the total NK cell output from this subset was negligible, as compared to the P1 and P4 populations. Due to dramatic inhibition of CD45<sup>+</sup>cell development in DOX treated cultures, the NK cell potential of only two major populations were evaluated: i) the D8 P1 population which markedly expanded following DOX treatment and ii) the D8 P2 population. These studies revealed that, in DOX treated cultures, NK cell potential was restricted to the D8 P1 population and that this population possessed much stronger NK cell differentiation potential as compared to the NK-producing populations in the control No DOX cultures (FIG. 5B and 5C). Overall, HPs collected from DOX2-8 treated cultures expanded more rapidly and longer in NK cell differentiation cultures and were able to generate up to 5-fold more NK cells as compared to the control (FIG. 5D). Although differences in cytotoxic potential against K562 cells of NK cells generated from DOX<sup>+</sup> and DOX<sup>-</sup> conditions were not observed (FIG. 5E), in response to PMA, NK

cells from DOX-treated cultures exhibited robust IFN $\gamma$  production, but slightly decreased degranulation response as compared to NK cells from non-treated cultures (FIG. 5F and 5G). Overall, these studies indicated that enforced SOX18 expression predominantly expanded an CD34<sup>+</sup>CD43<sup>+</sup>CD235a<sup>+</sup>CD41a<sup>-</sup>CD45<sup>-</sup> (D8 P1) HP population with superior NK cell potential.

#### Example 6

##### Molecular Characterization of SOX18-Induced Changes in HE and Blood Cells

**[0094]** To define changes in transcriptional program induced by enforced SOX18 expression, RNAseq analysis was performed on D4 HE, D8 CD34<sup>+</sup>CD43<sup>+</sup> subsets and NK cells generated from No DOX and DOX-treated cultures (FIG. 6A). Only 15 differentially expressed genes (DEGs) were detected in the D4 DLL4<sup>-</sup>CXCR4<sup>-</sup>VEC<sup>+</sup> population (D4 P1 population) following D2-DOX treatment (FIG. 6B), suggesting minimal effect of SOX18 on D4 HE. In contrast to prior findings of D2-4 SOX17 overexpression (see, Jung et al., 2021, Cell Rep. 34: 108758), RNAseq analysis of SOX18 D4 HE from No DOX and DOX cultures showed no significant increase in expression of HOXA genes, CDX2, or genes involved in NOTCH signaling pathways (FIG. 6C). These findings were confirmed using qPCR analysis of selected genes in D4 HE cells isolated from control and D2-4 DOX-treated differentiation cultures of iSOX17 and iSOX18 hPSCs (FIG. 6D). Since flow cytometric analysis showed that SOX18 overexpression induced DLL4<sup>-</sup>CXCR4<sup>+</sup> subset within D4 HE (FIG. 3A and 6A), DEGs were analyzed in DLL4<sup>-</sup>CXCR4<sup>+</sup> (D4 P1) and DLL4<sup>-</sup>CXCR4<sup>-</sup> (D4 P2) populations from DOX-treated cultures. These studies revealed no DEGs in these subsets including expression of arterial genes, thus suggesting that CXCR4 expression in D4 HE without co-expression of DLL4 does not reflect activation of arterial program. The most significant changes in transcriptional program were observed in the D8 P1 multipotent HP population and in CD56<sup>+</sup>NK cells which showed 209 and 811 DEGs in their corresponding DOX treated and untreated cultures (FIG. 6B). Gene set enrichment analysis (GSEA) in D8 P1 cells revealed enrichment in Kyoto Encyclopedia of Genes and Genomes (KEGG) categories related to T cell receptor (TCR) and Toll like receptor (TLR) signaling pathways (FIG. 6E), suggesting that SOX18 overexpression has a major effect on establishing T lymphoid transcriptional program in the D8 P1 multipotent HPs arising from HE. As shown in FIG. 6F, downregulated genes in these categories included CD3 complex genes, CD4, API complex genes JUN and FOS, PI3 signaling genes and the majority of TLR genes, while upregulated genes included IFNG, TICAM2, TLR8 and TLR9. D8 P2 population in DOX<sup>+</sup> and DOX<sup>-</sup> conditions showed only 79 DEGs enriched in extracellular matrix (ECM) receptor interactions KEGG categories (FIG. 6E). GSEA in NK cells generated from DOX treated and untreated cultures demonstrated enrichment in KEGG categories related to glyoxylate and dicarboxylate metabolism, one carbon pool by folate, and PPAR signaling pathway (FIG. 6E). However, no differential expression of NK cell activating receptors KLRK1, NCR2, NCR1 or key NK cell transcription factors EOMES and TBX21 was observed between NK cells from DOX<sup>-</sup> treated and untreated cultures (Data S1). These findings indicated that overexpression of SOX18 during endothelial-to-hematopoietic transition (EHT) has a major impact on the transcriptional program regulating metabolism of NK cells.

[0095] To determine whether SOX18 overexpression affects proliferation and apoptosis in major cell subsets, gene sets enrichment analysis was performed in GO categories related to “Regulation of Cell Population Proliferation” and “Apoptosis”. The first GO gene set was significantly altered in D4 and D8 P1 subsets and NK cells in DOX+ versus No DOX comparisons, whereas apoptotic genes were mostly affected in D8 P1 subset (FIG. 10A and 10B). To confirm the effect of SOX18 on cell proliferation, we performed cell-cycle analysis of D5 differentiated cells using bromodeoxyuridine (BrdU). Consistent with gene expression analysis, SOX18 overexpression led to a significant cell-cycle shift from GO/G1 to S phases in HE cells and emerging CD43+ HPs (FIG. 10C and 10D).

TABLE 1

List of antibodies used in this disclosure				
Antigen	Conjugate	Source	Clone	Cat No.
CD4	APC	BD Biosciences	RPA-T4	555349
CD5	PE-Vio770	Miltenyi Biotec	REA782	130-111-109
CD7	FITC	BD Biosciences	M-T701	555360
CD8	PE	BD Biosciences	HITBa	555635
CD18	APC	BD Biosciences	3G8	561248
CD31	MicroBeads	Miltenyi Biotec	N/A	130-091-935
CD34	PE	BD Biosciences	581	555822
CD41a	APC	BD Biosciences	HIP8	559777
CD41a	PerCP-Cy5.5	BD Biosciences	HIP8	340930
CD42b	PE	BD Biosciences	HIP1	555473
CD43	APC-Vio770	Miltenyi Biotec	DF-T1	130-101-174
CD43	BV421	BD Biosciences	1G10	562916
CD45	PE-Vio770	Miltenyi Biotec	581	130-113-119
CD56	PE	BD Biosciences	B159	555516
CD56	PE-cy7	BD Biosciences	B159	560916
CD61	APC	BD Biosciences	VI-PL2	564174
CD73	BV421	BD Biosciences	AD2	562430
CD94	FITC	BD Biosciences	HP-3D9	555888
CD144	BV421	BD Biosciences	55-7H1	565670
CD144	PE	Miltenyi Biotec	REA199	130-118-495
CD184	PE-Vio770	Miltenyi Biotec	REA649	130-120-797
CD235a	APC	BD Biosciences	GA-R2 (HIR2)	551336
CD309	PE	BD Biosciences	89106	560494
DLL4	APC	Miltenyi Biotec	MHD4-46	130-096-560
Ghost dye	Violet 540	TONBO Biosciences	TONBO	13-0879-T100

TABLE 2

List of Primer Pairs used for RT-qPCR					
SEQ ID NO.	Gene	Forward Sequence (5' - 3')	SEQ ID NO.	Reverse Sequence (5' - 3')	Purpose
1	CDX2	CCGAACAGGG ACTTGTTTA  GAG	2	AGGTTGGCTC TGGCATTAT A G	qPCR
3	DLL4	CAGTGGGCAG CGAAGCTAC A	4	ACAGGCAGTG GTAGCCATCC TC	qPCR
5	HEY1	GGACTATCGG AGTTTGGGA TTT	6	TGGGAAGCGT AGTTGTTGAG	qPCR
7	HOXA7	AGGTCCAGGA TCAGGGTAT T	8	CCAGAGAAGG AGGGATTGAT TC	qPCR

TABLE 2-continued

List of Primer Pairs used for RT-qPCR					
SEQ ID NO.	Gene	Forward Sequence (5' - 3')	SEQ ID NO.	Reverse Sequence (5' - 3')	Purpose
9	HOXA9	GCGCCTTCTC TGAAAACAAT	10	CAGTTCAGG GTCTGGTGT	qPCR
11	HOXA10	GCAAAGAGTG GTCGGAAGA A	12	CGCTCTCGAG TAAGGTACAT ATTG	qPCR
13	RPL13A	CCTGGAGGAG AAGAGGAAA  GAGA	14	TTGAGGACCT CTGTGTATTT G TCAA	qPCR
19	SOX18	GGCAAAGCGT GGAAGGA	20	CGGCCCGTAC TTGTAGTTG	

22-0629-US-PRO\_Sequence-Listing\_ST25  
 1. PiggyBack-SOX18 construct sequences (SEQ ID NO.: 15)  
 ACTCTTCCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTG  
 TCTCATGAGCGGATACATATTGAAATGATTTAGAAAAATAAACA  
 AATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTAAATT  
 GTAAGCGTTAATATTTTGTAAATTCGCGTTAAATTTTGTGTAA  
 ATCAGCTCATTTTTTAAACCAATAGCCGAAATCGGCAAAATCCCT  
 TATAAATCAAAAAGATAGACCAGATAGGGTTGAGTGTGTTCCCA  
 GTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTC  
 AAAGGGCGAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAA  
 CCATCACCCCTAATCAAGTTTTTTGGGGTCGAGGTGCGCTAAAGCA  
 CTAAATCGGAACCCCTAAAGGGAGCCCCGATTTAGAGCTTGACGG  
 GGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAAGCGAAA  
 GGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGC  
 GTAACCACCACACCCCGCGCTTAATGCGCCGTACAGGGCGCG  
 TCCCATTGCGCATTGAGGCTGCGCAACTGTTGGGAAGGGCGATCG  
 GTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGT  
 GCTGCAAGGCGATTAAGTTGGGTAACGCCAGGTTTTCCCAAGTCA  
 CGACGTTGTAACACGACGGCCAGTGAGCGCGCCTCGTTCAATCAC  
 GTTTTTGAACCCGTGGAGGACGGGCAGACTCGCGGTGCAAAATGTG  
 TTTTACAGCGTGATGGAGCAGATGAAGATGCTCGACACGCTGCAG  
 AACACGCGAGCTAGATTAACCCCTAGAAGATAATCATATTGTGACG  
 TACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGTTTTATC  
 GGTCTGTATATCGAGGTTTTATTATTAATTTGAATAGATATTAAG  
 TTTTATTATATTTACACTTACATACTAATAATAAATTCACAAAC  
 AATTTATTTATGTTTATTTATTTATTAATAAATAAATAAATAAATAA  
 AAATTTCTTCTATAAAGTAACAAAATTTTATCGAATTCCTGCAG



- continued

CCCGGGGGATCCACTAGTTCTAGAGGGACAGCCCCCCCCAAAGC  
 CCCCAGGGATGTAATTACGTCCCTCCCCCGCTAGGGGGCAGCAGC  
 GAGCCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCCGCATCC  
 CCGAGCCCGCAGCGTCCGGGGACAGCCCGGGCACGGGGAAGGTGG  
 CACGGGATCGCTTTCCTCTGAACGCTTCTCGCTGCTCTTTGAGCC  
 TGCAGACACCTGGGGGATACGGGGAAAAGGCTCCAAGGCCAGC  
 TTCCACAATAAGTTGGGTGAATTTTGGCTCAATTCCTCTTTCTA  
 TAGGATTGAGGTGAGAGCTTTGTGATGGGAATTCGTGGAATGTG  
 TGTCAAGTTAGGGTGTGAAAGTCCCGCATCGATCCTAGCTTCGT  
 GAGGCTCCGGTCCCGTCACTGGGCGAGAGCGCACATCGCCACAG  
 TCCCCGAGAAGTTGGGGGGAGGGGTCCGCAATTGAACCGGTGCC  
 AGAGAAGGTGGCCGCGGGTAACTGGGAAAGTATGATGCTGCTACT  
 GGCTCCGCTTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTG  
 CAGTAGTCGCGTGAACGTTCTTTTTCGCAACGGGTTTGCGCCA  
 GAACACAGGTAAGTGCCTGTGTGGTTCCCGCGGGCTGGCCCTCT  
 TTACGGGTTATGGCCCTTGCCTGCTTGAATTAATTCACCTGGC  
 TCCAGTACGTGATTTCTGATCCCGAGCTGGAGCCAGGGGGGGCC  
 TTGCGCTTTAGGAGCCCTTCGCTCGTGTGAGTTGAGGCCCTG  
 GCCTGGGCGCTGGGGCCGCGCGTGCGAATCTGGTGGCACCTTCG  
 CGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTT  
 TTGATGACCTGCTGCGACGCTTTTTTCTGGCAAGATAGTCTTGT  
 AAATGCGGGCCAGGATCTGCACACTGGTATTTTCGGTTTTTGGGCC  
 CGCGGCCGGCGACGGGGCCGTGCGTCCAGCGCACATGTTCCGGC  
 GAGGCGGGGCTCGCAGCGCGCCACCGAGAATCGGACGGGGGTA  
 GTCTCAAGCTGGCCGCTGCTCTGGTGCCTGGCCCTCGCGCCGCC  
 GTGTATCGCCCCCGCTGGGGCGCAAGGCTGGCCCGGTCCGCCACC  
 AGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTCCAGG  
 GGGCTCAAAATGGAGGACGCGCGCTCGGGAGAGCGGGCGGGTGA  
 GTCACCCACAAAGGAAAGGGCTTTCCGCTCTCAGCCGTCGC  
 TTCATGTGACTCCACGGAGTACCGGGCGCGCTCCAGGCACCTCGA  
 TTAGTTCTGGAGCTTTTGGAGTACGCTGCTTTAGGTTGGGGGA  
 GGGGTTTTATCGATGGAGTTTCCCCACACTGAGTGGGTGGAGAC  
 TGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCCTGTTGGAATT  
 TGCCCTTTTGTAGTTTGGATCTTGGTTCAATTCGAAGCCTCAGAC  
 AGTGGTTCAAAGTTTTTTCTTCCATTTAGGTGTCGTGAACAG  
 TGGTCCGCGCCATCTAGAGCAACTAGCATGGCCAAGTTGACCAGT  
 GCCGTTCCGGTGTCAACCGCGCGCAGCTCGCCGGAGCGGTCCGAG  
 TTCTGGACCACCGGCTCGGGTTTCCCGGGACTTCGTGGAGGAC  
 GACTTCGCGGGTGTGGTCCGGGACGACGTGACCTGTTATCAGC

- continued

GCGGTCCAGGACCAGGTGGTGGCCGGACAACACCTGGCCTGGGTG  
 TGGGTGCGCGGCTGGACGAGCTGTACGCCGAGTGGTCCGGAGTGC  
 GTGTCCACGAACCTCCGGGACGCTCCGGGCCGGCCATGACCGAG  
 ATCGGGCAGCAGCCGTGGGGGGGAGTTCCGCCCTGCGCAGCCCG  
 CCGGCAACTGCGTGCCTTCCGTGGCCGAGGAGCAGGACTGAATCT  
 AAGATAATTGATGAGTTTGGACAAACCACTAGAAATGCGAGTG  
 AAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATT  
 TGTAAACATTATAAGCTGCAATAAACAGTTAAACAACAACAATTG  
 CATTCAATTTATGTTTCAGGTTCCAGGGGAGGTGTGGGAGTTTT  
 TTAAGCAAGTAAACCTCTACAATGTGGTATGGCTGATTATGA  
 TCCGGCTGCCCTCGCGCATGCATCATAGAGCCACCAGCATCCCAG  
 CATGCCCTGCTATGTCTTCCCAATCTCCCCCTTGTGTCTCTGCC  
 CCACCCACCCCCGAGAATAGAATGACACCTACTCAGACAATGCG  
 ATGCAATTTCTCATTTTATTAGGAAAGGACAGTGGGAGTGGCAC  
 CTTCCAGGGTCAAGGAGGACGGGGGAGGGGCAACAACAGATG  
 GCTGGCAACTAGAAGGCACAGTCCAGGCTGATCAGCCATGATTA  
 CTTGTACAGCTCGTCCATGCCGAGAGTATCCCGCGCGGCTCAC  
 GAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGCTTTT  
 GCTCAGGGCGGACTGGTAGCTCAGGTAGTGGTTGTGGGCAGCAG  
 CACGGGGCCGTCGCCGATGGGGTGTCTGCTGGTAGTGGTCGGC  
 GAGCTGCACGCCCGCTCCTCGATGTTGTGGCGGATCTTGAAGTT  
 GGCCTTGATGCCGTTCTTCTGCTTGTGGCGGTGATATAGACGTT  
 GTGGCTGTTGATGTTGACTCCAGCTTGTGCCCCAGGATGTTGCC  
 GTCCTCCTTGAAGTTCGATGCCCTTCCAGCTCGATGCGGTTCCACCAG  
 GGTGTGCGCCTCGAACTTCACTCGCGCGGGTCTTGTAGTTGCC  
 GTCGTCCTTGAAGAAGATGGTGGCTCCTGGACGTAGCCTTCGGG  
 CATGGCGGACTTGAAGAAGTCTGCTGCTTCAATGTTGGTGGGGTA  
 GCGGGCGAAGCACTGCAGGCGTAGCCAGGGTGGTCCAGGAGGT  
 GGGCCAGGGCACGGGACGCTTCCCGTGGTGCAGATCAGCTTCCAG  
 GGTGAGCTTCCCGTAGGTGGCATCGCCCTCGCCCTCGCCGACAC  
 GCTGAACCTTGTGGCCGTTTACGTGCGCGTCCAGCTCGACCAGGAT  
 GGGCACCCCCGGTGAACAGCTCCTCGCCCTGCTCACCATGGG  
 TCCAGGGTTCTCTCCACGCTCTCCAGCCTGCTTCAGCAGGCTGAA  
 GTTAGTAGCTCCGCTTCCGTATACGCCGGAGATGCACGCGCTGTA  
 ATAGACCGCGCTGCTGGCGTCCGACAGCGCGGAGATCAGGCTGCT  
 CTCCTCTGGGCAGGACATGGCGCGCGGGCCAGTTTGGCCAGTGC  
 CACGTGGTACGGGAGCCCGGGGCGTCCGGCCGAGTCCGGCTGCA  
 GTTGGGTTACTGGTCCAACTCGGTGAGGTTCCAGTCCGCCACAG  
 ATCGGGCGGGCCCCAGCGGCTCGGCGCTCTCCAGCGCGGGCC

- continued

CTCGGGCGGCGGCAGCGGGCCGGGTACGGGCCGGGCGTGCC  
CAGGGTGCCGTAGTACAGGCCAGCGAGCGGCGCCGGGGGGCGC  
GGTCTGAGCCCTCCGCCAGGGGAGCCCGTAGCAACCGCCGG  
GTCCCGCACAACCTCGGTGGGCGGTAGGGCGCGCGGAAGGGCCG  
CAGCGCGCAGTCTCGGGCGCCGGGGCGGTGGGAAGAAGGCAGC  
CTCGCCGGGCTCAGGCCCTCAGAGCGAGCGCTCGGGCGTGGG  
CAGCCCCAGGCCCTCGAATCGGCGCCAGCGGGGCGAGCTCGCG  
GAAGGCGCGAGCCGAGCCAGACGCCGCGGGGAAAGGCTCGGGCG  
TGCTGCGGGGGCGCTAATCCCGGAGCAGGAGGCCGGGCTCCAG  
CCGCCGGGCTTGCGCGCTGCTTCTTGCGGCGCGCCGGTACTT  
GTAGTTGGGGTGGTCCGCAAGTGTGACCGCGCAGCCGTTCCGG  
TTCTCCACGAAGGGCCGCTTCTCCGCGGTTGAGCTCCTTCCA  
CGCTTTGCCAGCATCTTGTGAGCACCGCTTGTGAGGTCGGG  
GTTCTGCTGAGCCAGCCGCTTGCGCTCGTCTTGGCCACACCAT  
GAAGGCGTTCATGGGCGCCGGATGCGCGACTGCTGCGCCCTG  
GCGTTCCCGCGCGCGCCGGGCTGAGGCATAGCCCGCCGGCTC  
GGGGCTGCGCGGGGACTGCGCTGCGGGCTGGGCGGCGAGCGGG  
CGCGGCGGGCGCGGCGAGGGCGGGCGCGGCGGAGGCCGCG  
CGTGTGAGCGCGCCCGCTGTCGCGGGCCATGCACAGTCGCG  
GCGGGCGGGCGGTGCTCTGTGCGCCGTAGCCGGGCGGCGATCT  
CTGCATGGATCCTCCAGGCGATCTGACGGTTCACTAAACGAGCTC  
TGCTTATATAGGCTCCACCGTACACGCTACCTCGACATACGT  
TCTCTATCACTGATAGGGAGTAAACTCGACATACGTCTCTATCA  
CTGATAGGGATAAACTCGACATACGTTCTCTATCACTGATAGGGA  
GTAAACTCGACATACGTTCTCTATCACTGATAGGGAGTAAACTCG  
ACATACGTTCTCTATCACTGATAGGGAGTAAACTCGACATCGTTC  
TCTATCACTGATAGGGAGTAAACTCGACATACGTTCTCTATCACT  
GATAGGGAGTAAACTCGCTCGAGACGCTATACCTAGGATCCGGC  
CGGCTGCGAGGTCTCTCACAGGAACGAAGTCCCTAAAGAAACAG  
TGGCAGCCAGGTTTAGCCCGGAATTGACTGGATTCTTTTTTAG  
GGCCATTGGTATGGCTTTTTCCCGTATCCCCCAGGTGCTGCG  
AGGCTCAAAGAGCAGCGAGAAGCGTTCAGAGGAAAGCGATCCCGT  
GCCACCTTCCCGTGCCCGGGCTGTCCCGCACGCTGCCGGCTCG  
GGGATGCGGGGGAGCGCCGACCGAGCGGAGCCCGGGCGGCT  
CGCTGCTGCCCTAGCGGGGAGGACGTAATTACATCCCTGGG  
GGCTTTGGGGGGGGCTGTCCCTCTAGAGCGGCGCCACCGCGGT  
GGAGCTCCAGCTTTTGTCCCTTTAGTGAGGGTTAATTAGATCTT  
AATACGACTCACTATAGGGCAATTGGGTACCGGGCCCCCTCG  
AGGTCGACGGTATCGATAAGCTTGATATCTATAACAAGAAAATAT

- continued

ATATATAATAAGTTATCACGTAAGTAGAACATGAAATAACAATAT  
AATTATCGTATGAGTTAAATCTTAAAAGTACGTAAAAGATAATC  
ATGCGTCATTTTACTCACGCGGTCGTTATAGTTCAAATCAGTG  
ACACTTACCGCATTGACAAGCACGCCTCACGGGAGCTCCAAGCGG  
CGACTGAGATGTCTAAATGCACAGCGACGGATTGCGCTATTTA  
GAAAGAGAGAGCAATATTTCAAGAAATGCATGCGTCAATTTTACGC  
AGACTATCTTCTAGGGTTAATCTAGCTGCATCAGGATCATATCG  
TCGGGTCTTTTTTCCGGCTCAGTCATCGCCCAAGCTGGCGCTATC  
TGGGCATCGGGGAGGAAGAGCCCGTGCCCTTTCCCGCGAGGTTG  
AAGCGGCATGGAAGAGTTTGCCGAGGATGACTGCTGCTGCATTG  
ACGTTGAGCGAAAACGCACGTTTACCATGATGATTCGGGAAGGTG  
TGGCCATGCACGCTTTAACGGTGAAGTTCGTTTACGGCCACCT  
GGGATACCAGTTCGTGCGGCTTTTCCGACACAGTTCGGATGG  
TCAGCCCGAAGCGCATCAGCAACCCGAACAATACCGGCGACAGCC  
GGAAGTGCCTGCGGTTGTGAGATTAATGACAGCGGTGCGGCGC  
TGGGATATTACGTGAGCGAGGACGGGTATCCTGGCTGGATGCCCG  
AGAAATGGACATGGATACCCCGTGAAGTACCCGGCGGGCGGCTT  
GGCGTAATCATGGTATAGCTGTTTCTGTGTGAAATTTTATCC  
GCTCACAATCCACACAACATACGAGCGGAAGCATAAAGTGTA  
AGCCTGGGGTGCCTAATGAGTGAAGTAACTCACATTAATTGCGTT  
GCGCTCACTGCCGCTTTCAGTCGGGAAACCTGCTGTCAGCT  
GCATTAATGAATCGGCCAACCGCGGGGAGAGCGGTTTTCGCTAT  
TGGGCGCTCTCCGCTTCTCGCTCACTGACTGCTGCGCTCGGT  
CGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGCGGTAAT  
ACGGTTATCCACAGAATCAGGGGATAACCGAGGAAAGAACATGTG  
AGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAGGCCCGGCTT  
GCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAA  
AAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATA  
AAGATACCAGGCGTTTTCCCTGGAAGCTCCCTCGTGGCTCTCC  
TGTTCCGACCTGCGGCTTACCGGATACCTGTCCGCTTTTCTCCC  
TTCCGGGAGCGTGGCGCTTCTCATAGCTCACGCTGATAGGATCT  
CAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGACGA  
ACCCCCGTTACGCCCGACCGCTGCGCTTATCCGGTAACTATCG  
TCTTGAGTCCAACCCGGTAAAGACAGACTTATCGCCACTGGCAGC  
AGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGC  
TACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG  
GACAGTATTTGGTATCTGCGCTCTGTGAAGCCAGTTACCTTCGG  
AAAAAGAGTTGGTAGCTTTGATCCGGCAAACAACACCGCTGG  
TAGCGGTGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAA

-continued

AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGCTCTGA  
CGCTCAGTGGAAACGAAACTCACGTTAAGGGATTTTGGTCATGAG  
ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATG  
AAGTTTTAAATCAATCTAAAGTATATATAGTAAACTTGGTCTGA  
CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTG  
TCTATTTCTGTTTCCATAGTAGTTCCTGACTCCCCGTCGTGTAGAT  
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAAT  
GATACCGCAGACCCACGCTCACCGCTCCAGATTTATCAGCAAT  
AAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAAC  
TTATCCGCTCCATCCAGTCTATTAATTGTGCGGGAAGCTAG  
AGTAAGTAGTTCGCCAGTAAATAGTTTGGCCAACGTTGTTGCCAT  
TGCTACAGGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTC  
ATTACAGTCCGGTCCCAACGATCAAGGCAGTTACATGATCCCC  
CATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGT  
TGTCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGC  
AGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTT  
TTCTGTGACTGGTGGTACTCAACCAAGTATTCTGAGAATAGTG  
TATGCGCGCACCAGTGTCTTTCGCCGCGCTCAATACGGGATAA  
TACCGCGCCACATAGCAGAACTTTAAAAGTGTCTCATCTGGAAA  
ACGTTCTTCGGGCGAAAACCTCAAGGATCTTACCCTGTTGAG  
ATCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTTCAGC  
ATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG  
GCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGAAAATGTTG  
AATACTCAT

1A.  
Human SOX18 CDS (ORF) (SEQ ID NO. : 16)  
\*\*\*  
ATGCAGAGATCGCCGCCGCTACGGCGCACAGGACGACCCGCC  
GCCCGCCGCGACTGTGCATGGGCCCGGGACACGGGGCCGCCCT  
GACACGCGCGCCTCGCCGCCGCCCGCCGCCCTCGCCGCCGCC  
GCCGCGCCGCCCTCGCCGCCCGCCCGCAGCGCAGTCCCCGCGC  
AGCCCCGAGCCGGGCGCTATGGCCTCAGCCCGCCGGCCGCGGG  
GAACGCCAGGCGGACAGAGTCCGCGCATCCGGCGCCCATGAAC  
GCCTTTCATGGTGTGGGCAAGGACGAGCGCAAGCGGCTGGCTCAG  
CAGAACC CGACCTGCACAACCGGTGCTCAGCAAGATGCTGGGC  
AAAGCGTGAAGGAGCTGAACGCGCGGAGAGCGGCCCTTCGTG  
GAGGAAGCCGAACGGCTGCGCTGCAGCACTTGCAGCACCACCC  
AACTACAAGTACCGCCGCCGCCAAGAAGCAGGCGCGCAAGGCC  
CGCGGCTGGAGCCCGGCTCTGCTCCCGGATTAGGCCCCCG  
CAGCCACCGCCGAGCCTTCCCGCGGCGTCTGGCTCGGCTCGC

-continued

GCCTTCCGCGAGCTGCCCCGCTGGGCGCCGAGTTCGACGGCCTG  
GGGCTGCCACGCCCCGAGCGCTCGCTCTGGACGGCCTGGAGCCC  
GGCGAGGCTGCCTTCTCCACCGCCCGCGGCGCCGAGGACTGC  
GCGCTGCGGCCCTTCCGCGGCCCTACGCGCCACCAGTGTGTCG  
CGGGACCCCGCGGTTGCTACGGGGCTCCCCTGGCGGAGGCGCTC  
AGGACCGCGCCCCCGCGCGCCGCTCGCTGGCCTGTACTACGGC  
ACCCTGGGCACGCCCCGGCCGTAACCCGCCCCGCTGTGCGCCCG  
CCCAGGCCCGCCGCTGGAGAGCGCCGAGCCGCTGGGGCCCGCC  
GCCGATCTGTGGCCGACGTGGACCTCACCGAGTTCGACCACTAC  
CTCAACTGCAGCCGACTCGGCCGACGCCCCGGGCTCCCGTAC  
CACGTGGCACTGGCCAACTGGGCCCGCGCGCCATGCTCTGCCA  
GAGGAGAGCAGCCTGATCTCCGCGCTGTGCGACCCAGCAGCGCG  
GTCTATTACAGCGCGTGCATCTCCGGC  
\*\*\* = start (ATG) codon

SOX18 amino acid sequence (SEQ ID NO. : 17)  
MQRSPPGYGAQDDPPARRDCAWAPGHAAADTRGLAAGPAALAAP  
AAPASPPSPQRSPPRSPPEPGRYGLSPAGRGERQAADSRIRRPMN  
AFMVWAKDERKRLAQNPDLHNAVLSKMLGKAWKELNAAEKRPV  
EEAERLRVQHLRDHPNYKYRPRRKKQARKARRLEPGLLLPLGLAPP  
QPPPEPPPAASGSARAFRELPPLGAEFDGLGLFTPERSPLDGLPE  
GEAAFFPPPAAPEDCALRPFPRAPYAPTELSRDPGGCYGAPLAEAL  
RTAPPAAPLAGLYYGLTGTGPGYPGPLSPPEAPPLESAEPLGPA  
ADLWADVLDLTFDQYLNCSRTRPDAPGLPYHVALAKLGPAMSCP  
EESLISALS DASSAVVYSACISG

2. PB-M2rtTA sequences (regulatory element  
for DOX-inducible SOX18 expression) (SEQ ID NO. : 18)  
ACTCTTCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTG  
TCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAACA  
AATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTAAATT  
GTAAGCGTTAATATTTTGTAAAATTCGCGTTAAATTTTTGTAA  
ATCAGCTCATTTTTTAAACCAATAGGCCGAAATCGGCAAAATCCCT  
TATAAATCAAAAAGATAGACCGAGATAGGGTTGAGTGTGTTCCA  
GTTTGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTC  
AAAGGGCGAAAACCGTCTATCAGGGCGATGGCCACTACGTGAA  
CCATCACCCCTAATCAAGTTTTTTGGGGTCGAGGTGCGCTAAAGCA  
CTAAATCGGAACCTAAAGGGAGCCCCGATTTAGAGCTTGACGG  
GGAAGCCGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCGAAA  
GGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACAGTGC  
GTAACCACCACACCCCGCCGCTTAATGCGCCGCTACAGGGCGCG  
TCCCATTCGCCATTGAGGCTGCGCAACTGTTGGGAAGGGCGATCG

- continued

GTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAAGGGGATGT  
GCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCA  
CGACGTTGTAAAACGACGGCCAGTGAGCGGCCCTCGTTCATTAC  
GTTTTTGAACCCGTGGAGGACGGGCAGACTCGCGGTGCAAAATGTG  
TTTTTACAGCGTGATGGAGCAGATGAAGATGCTCGACACGCTGCAG  
AACACGCAGCTAGATTAACCCTAGAAGATAATCATATTGTGACG  
TACGTTAAAGATAATCATGCGTAAAATGACGCATGTGTTTTATC  
GGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAG  
TTTTATTATATTTACACTTACATACTAATAATAAATCAACAAC  
AATTTATTTATGTTTTATTTATTTATTAATAAAAAACAAAACTCA  
AAATTTCTTCTATAAAGTAAACAAACTTTTATCGAATTCCTGCAG  
CCCGGGGATCCACTAGTCTTAGAGGGACAGCCCCCCCCAAAGC  
CCCCAGGGATGTAATACGTCCCTCCCCGCTAGGGGGCAGCAGC  
GAGCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCCGCATCC  
CCGAGCCGCGAGCGTCCGGGACAGCCCGGGCACGGGGAAGGTGG  
CACGGGATCGCTTTCCTCTGAACGCTTCTCGCTGCTCTTGAGCC  
TGCAGACACCTGGGGGATACGGGAAAAGGCTCCAAGGCCAGC  
TTCCACAATAAGTTGGTGAATTTTGGTCTATTCCTCTTTCTA  
TAGGATTGAGGTGAGAGCTTTGTGATGGGAATCTGTGGAATGTG  
TGTCAAGTTAGGGTGTGAAAGTCCCGCGATCGTCGAGAGATTTCT  
AGCTTCGTGAGGCTCCGGTCCCGTCACTGGGCAGAGCGCACATC  
GCCACAGTCCCGAGAAGTTGGGGGAGGGTCCGCAATTGAAC  
CGGTGCC TAGAGAAGGTGGCGGGGTAACCTGGGAAAGTGATGT  
CGTGTACTGGCTCCGCTTTTTCCGAGGTTGGGGGAGAACCCTA  
TATAAGTGCAGTAGTCCCGTGAACGTTCTTTTTCGCAACGGGTT  
TGCCGCCAGAACACAGGTAAGTCCGCTGTGTGGTTCCCGCGGGCC  
TGGCCTCTTTACGGGTTATGGCCCTTGCCTGCTTGAATTAATTC  
CACCTGGCTCCAGTACGTGATTTCTGATCCCGAGCTGGAGCCAGG  
GGGGGGCCTTGCCTTTAGGAGCCCTTCGCTCGTGTGAGTT  
GAGGCCCTGGCCTGGGCGCTGGGGCCCGCGTGCGAATCTGGTGG  
CACCTTCGCGCCTGTCTCGCTGCTTTGATAAGTCTCTAGCCATT  
TAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGAT  
AGTCTTGTAATCGCGCCAGGATCTGCACACTGGTATTTCCGGTT  
TTTGGCCCGCGCCGCGCAGCGGGCCGCTGCGTCCAGCGCAC  
TGTTCCGGCAGGCGGGCCCTGCGAGCGCGGCCACCGAGAATCGGA  
CGGGGGTAGTCTCAAGCTGGCCGGCTGCTCTGGTGCCTGGCCTC  
GCGCCGCGGTGTATCGCCCCGCTTGGGCGCAAGGCTGGCCCGG  
TCGGCACAGTTGCGTGAGCGAAAGATGGCCGCTTCCCGCCCT  
GCTCCAGGGGCTCAAAATGGAGGACGCGCGCTCGGGAGAGCGG

- continued

GCGGGTGAGTCACCACACAAAAGGAAAAGGGCCTTCCGTCTCTCA  
GCCGTGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGG  
CACCTCGATTAGTTCTGGAGCTTTTGGAGTACGTCGCTTTTAGGT  
TGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGG  
GTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATCTCG  
TTGGAATTTGCCCTTTTGGAGTTGGATCTTGGTTCATTTCTCAAG  
CCTCAGACAGTGGTTCAAAGTTTTTTCTTCCATTTCAAGTGTG  
TGAACACGTGGTCCGGCCATCTAGAGCAACTAGCTTGGGCCCGG  
GAATTAATTCACCATGTCTAGACTGGACAAGAGCAAGTCATAAA  
CGGCGCTCTGGAATTAAGTCAATGGAGTCCGATTCGAAAGCCTGAC  
GACAAGGAACTCGCTCAAAAGCTGGGAGTTGAGCAGCCTACCCCT  
GTACTGGCACGTGAAGAACAAGCGGGCCCTGCTCGATGCCCTGCC  
AATCGAGATGCTGGACAGGCATCATAACCACTTCTGCCCTGGA  
AGGCGAGTCATGGCAAGACTTTCTGCGGAACAACGCCAAGTCATT  
CCGCTGTGCTCTCCTCTCACATCGCGACGGGGCTAAAGTGCATCT  
CGGCACCCGCCAACAGAGAAACAGTACGAAACCTTGGAAAATCA  
GCTCGGTTCTGTGTGAGCAAGGCTTCTCCCTGGAGAACGCACT  
GTACGCTCTGTCGCCGTTGGCCACTTTACACTGGGCTGCGTATT  
GGAGGAACAGGAGCATCAAGTAGCAAAAGAGGAAAGAGAGACACC  
TACCACCGATTCTATGCCCCACTTCTGAGACAAGCAATTGAGCT  
GTTCCGACCGCGAGGAGCCGAACTGCTTCTTTTCCGCTGGA  
ACTAATCATATGTGGCTGGAGAAACAGCTAAAGTGCAGAAAGCGG  
CGGGCCGGCCGACGCCCTTACGATTTTACTTAGACATGCTCCC  
AGCCGATGCCCTTGACGACTTTGACCTTGATATGCTGCTGCTGA  
CGCTCTTGACGATTTGACCTTGACATGCTCCCGGGGAAATTCGG  
AAGCGGAGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGA  
GGAGAATCCCGCCCTGATATCATGACCGAGTACAAGCCACGGT  
GCGCCTCGCCACCCGCGACGAGTCCCGGGCCGTACGCACCT  
CGCCCGCGGTTCCCGGACTACCCGCCACCGCCACACCGTCTGA  
CCCGGACCGCCACATCGAGCGGGTACCGAGCTGCAAGAACTCTT  
CCTCACGCGCGTCCGGCTCGACATCGGCAAGGTGTGGGTCGCGGA  
CGACGGCGCCGGTGGCGGTCTGGACCACCGCGGAGAGCGTCTGA  
AGCGGGGGCGGTGTTCCCGGAGATCGGCCCGGCATGGCCGAGTT  
GAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCT  
GGCGCCGACCGGCCAAGGAGCCCGGCTGGTTCCTGGCCACCGT  
CGGCGTCTCGCCGACACACAGGGCAAGGGTCTGGGCGAGCGCGT  
CGTGTCTCCCGGAGTGGAGGCGCGGAGCGCGCGGGGTGCCCGC  
CTTCTGGAGACTCCGCGCCCCGCAACCTCCCTTCTACGAGCG  
GCTCGGCTTACCGTACCGCCGAGCTCGAGTGGCCGAGGACC

- continued

GCGCACCTGGTGCATGACCCGCAAGCCCGGTGCTGATAAGATAC  
 ATTGATGAGTTTGGACAAACCACAACCTAGAAATGACAGTGAAAAAAA  
 TGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACC  
 ATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTGAT  
 TTTATGCTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGC  
 AAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATCCGGCT  
 GCCTCGCGCGGTACGACGCGTATACCTAGGATCCGGCCGGCCGTC  
 AGGTGTCCTCACAGGAACGAAGTCCCTAAAGAAACAGTGGCAGCC  
 AGGTTTAGCCCCGGAATTGACTGGATTCTTTTTTAGGGCCATT  
 GGATATGGCTTTTTCCCGTATCCCCCAGGTGCTGACAGGCTCAA  
 AGAGCAGCGAGAAGCGTTTACAGAGAAAGCGATCCCGTGGCCACTT  
 CCCCCTGCCCCGGGTGTCCTCCGACGCTGCCGGCTCGGGGATGCG  
 GGGGAGCGCCGACCGGAGCGGAGCCCGGGCGGCTCGCTGCTG  
 CCCCCTAGCGGGGAGGGACGTAATTACATCCTCGGGGCTTTGG  
 GGGGGGCTGTCCTCTAGAGCGGCCGCCACCGGCTGGAGCTCC  
 AGCTTTTGTTCCTTTAGTGGGGTTAATTAGATCTTAATACGAC  
 TCACTATAGGGCAATTGGGTACCGGGCCCCCTCGAGGTCGAC  
 GGATCGATAAGCTTGATACTATAACAAGAAAATATATATATAA  
 TAAGTTATACGTAAGTAGAACATGAAATAACAATATAATTATCG  
 TATGAGTTAAATCTTAAAGTACCGTAAAAGATAATCATGCGTCA  
 TTTTACTCACGGGTGCTTATAGTTCAAATCAGTGACACTTAC  
 CGCATTGACAAGCACGCTCACGGGAGCTCCAAGCGCGACTGAG  
 ATGTCCTAAATGCACAGCGACGGATTCCGCTATTTAGAAAGAGA  
 GAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCACTATC  
 TTTCTAGGGTTAATCTAGCTGCATCAGGATCATATCGTGGGTCT  
 TTTTTCCGGCTCAGTTCATCGCCCAAGCTGGCGCTATCTGGGCATC  
 GGGGAGGAGAAGCCCGTGCCTTTTCCCGGAGGTTGAAGCGGCA  
 TGGAAAGAGTTTGCAGGATGACTGCTGCTGATTGACGTTGAG  
 CGAAAACGCAAGTTTACCATGATGATTGGGAAGGTGTGGCCATG  
 CACGCTTTAACGGTGAACGTTGCTGTTTACGGCCACCTGGGATAC  
 AGTTCGTGCGGGCTTTTCCGGACACAGTTCCGGATGGTCAGCCG  
 AAGCGCATCAGCAACCGAACAAATACCGGCGACAGCCGGAATGC  
 CGTGCCTGTGTCAGATTAATGACAGCGGTGCGGCGCTGGGATAT  
 TACGTGAGCGAGGACGGGTATCTGGCTGGATGCCGAGAAATGG  
 ACATGGATACCCCGTGGTTACCGGGGGCGGCTTGGCGTAAT  
 CATGGTCATAGCTGTTTCCCTGTGTGAAATGTTTATCCGCTCACAA  
 TTCCACACAACATACGAGCCGGAGCATAAAGTGTAAAGCCTGGG  
 GTGCCAATGAGTGAGCTAACTCACATTAATGCGTTGCGCTCAC  
 TGCCCGCTTCCAGTCGGGAAACCTGCTGTCAGCTGCATTAAT

- continued

GAATCGGCCAACGCGCGGGGAGAGCGGTTTGCCTATTGGGCGCT  
 CTTCCGCTTCCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGGC  
 TGCGGCGAGCGGTATCAGCTCACTCAAAGCGGTAATACGGTTAT  
 CCACAGAATCAGGGGATAACGCAAGAAAGAACATGTGAGCAAAAG  
 GCCAGCAAAAGGCCAGGAACCGTAAAAGGCCCGGTTGCTGGCGT  
 TTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGAC  
 GCTCAAGTCAGAGGTGGGAAACCCGACAGGACTATAAAGATACC  
 AGGCGTTTTCCCTGGAGCTCCCTCGTGGCTCTCTGTTCGGA  
 CCCTGCGCTTACCGGATACCTGTCGCCCTTCTCCCTTCGGGAA  
 GCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCCG  
 TGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCG  
 TTCAGCCCCGACCGCTGCGCTTATCCGGTAACTATCGTCTTGAGT  
 CCAACCCGGTAAGACCGACTTATCGCCACTGGCAGCAGCCACTG  
 GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT  
 TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTAT  
 TTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAG  
 TTGGTAGCTCTTGATCCGGCAAAACAAACCCCGTGGTAGCGGTG  
 GTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGAT  
 CTCAGAAGATCCTTGATCTTTTACGGGGTCTGACGCTCAGT  
 GGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAA  
 AAAGGATCTTACCTAGATCCTTTTAAATTAATAATGAAGTTTAA  
 AATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACC  
 AATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT  
 GTTCATCCATAGTTGCTGACTCCCGCTCGTGTAGATAACTACGA  
 TACGGGAGGGCTTACCATCTGGCCCCAGTCTGCAATGATACCGC  
 GAGACCCACGCTCACCGCTCCAGATTTATCAGCAATAAACCCAGC  
 CAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCG  
 CCTCCATCCAGTCTATTAATTGTTCCGGGAAGCTAGAGTAAGTA  
 GTTCGCCAGTTAATAGTTTGCACAACGTTGTTGCCATTGCTACAG  
 GCATCGTGGTGTGACGCTCGTCTGTTTGGTATGGCTTCATTGAGCT  
 CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGT  
 GCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAA  
 GTAAGTTGGCCGAGTGTATCACTCATGTTTATGGCAGCACTGC  
 ATAATTCTCTACTGTGATGCCATCCGTAAGATGCTTTTCTGTGA  
 CTGGTGAGTACTCAACCAAGTCATCTGAGAATAGTGTATCGGGC  
 GACCGAGTTGCTCTTGCCTGGGCTCAATACGGGATAATAACCGCGC  
 CACATAGCAGAACTTTAAAGTGTCTCATTTGAAAAACGTTCTT  
 CGGGGCGAAAACCTCAAGGATCTTACCCTGTTGAGATCCAGTT  
 CGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTA

-continued

CTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATG

CCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCA

T

**[0096]** All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, and patent application was specifically and individually indicated to be incorporated by reference.

**[0097]** While some embodiments have been illustrated and described in detail in the appended drawings and the foregoing description, such illustration and description are to be considered illustrative and not restrictive. Other variations to the disclosed embodiments can be understood and effected in practicing the claims, from a study of the drawings the disclosure, and the appended claims. The mere fact that certain measures or features are recited in mutually different dependent claims does not indicate that the combination of these measures or features cannot be used. Any reference signs in the claims should not be construed as limiting the scope.

SEQUENCE LISTING

Sequence total quantity: 20	
SEQ ID NO: 1	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 1	
ccgaacaggg acttgtttag ag	22
SEQ ID NO: 2	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 2	
aggttggctc tggcatttat ag	22
SEQ ID NO: 3	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 3	
cagtgggcag cgaagctaca	20
SEQ ID NO: 4	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 4	
acaggcagtg gtagccatcc tc	22
SEQ ID NO: 5	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 5	
ggactatcgg agtttgggat tt	22
SEQ ID NO: 6	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 6	
tgggaagcgt agttgttgag	20
SEQ ID NO: 7	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 7	
aggtccagga tcagggtatt	20
SEQ ID NO: 8	moltype = DNA length = 22
FEATURE	Location/Qualifiers
source	1..22

-continued

---

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 8
ccagagaagg agggattgat tc 22

SEQ ID NO: 9      moltype = DNA length = 20
FEATURE          Location/Qualifiers
source          1..20
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 9
gcgcttctc tgaaaacaat 20

SEQ ID NO: 10     moltype = DNA length = 20
FEATURE          Location/Qualifiers
source          1..20
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 10
cagttccagg gtctggtgtt 20

SEQ ID NO: 11     moltype = DNA length = 20
FEATURE          Location/Qualifiers
source          1..20
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 11
gcaaagagtg gtcggaagaa 20

SEQ ID NO: 12     moltype = DNA length = 24
FEATURE          Location/Qualifiers
source          1..24
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 12
cgctctcgag taaggtacat attg 24

SEQ ID NO: 13     moltype = DNA length = 23
FEATURE          Location/Qualifiers
source          1..23
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 13
cctggaggag aagaggaaag aga 23

SEQ ID NO: 14     moltype = DNA length = 25
FEATURE          Location/Qualifiers
source          1..25
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 14
ttgaggacct ctgtgtat ttgtca 25

SEQ ID NO: 15     moltype = DNA length = 9190
FEATURE          Location/Qualifiers
source          1..9190
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 15
actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata 60
catatttgaa tgtatttaga aaaataaaca aatagggggt ccgcgacat ttcccgaaa 120
agtgccacct aaattgtaag cgtaaatatt ttgttaaaat tcgcgtaaa tttttgtaa 180
atcagctcat ttttaacca ataggccgaa atcggcaaaa tccctataa atcaaaagaa 240
tagaccgaga tagggttgag tgtgttcca gtttggaaac agagtccact attaaagaac 300
gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg gcgatggccc actacgtgaa 360
ccatcacctc aatcaagttt tttggggtcg aggtgccgta aagcaactaa tcggaaccct 420
aaaggagacc cccgatttag agcttgacgg gaaaagccgg cgaacgtggc gagaaagaa 480
ggaaagaaag cgaagagagc gggcgctagg gcgctggcaa gtgtagcggc cacgctgccc 540
gtaaccacca caccgcccgc gcttaatgag ccgctacagg gcgctgccc ttcgccattc 600
aggctgccc actgttgga agggcgatcg gtgcccggct cttecgatt acgccagctg 660
gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgcagggtt ttcccagtca 720
cgacgttgta aaacgacggc cagtgagcgc gctcgttca ttcacgttt tgaaccctg 780
gaggacgggc agactcggc tgcaaatgtg ttttacagcg tgatggagca gatgaagatg 840
ctcgacacgc tgcagaacac gcagctagat taacctaga aagataatca tattgtgacg 900
tacgttaaag ataactatgc gtaaaattga cgcattgtgt ttatcggtct gtatcagag 960
gtttatatt taatttgaat agatattaag ttttattata tttacactta cataactaata 1020

```

-continued

---

ataaattcaa	caaacaattt	atztatgtt	atztatttat	taaaaaaaa	caaaaactca	1080
aaatttcttc	tataaagtaa	caaaaactttt	atcgaattcc	tgcagcccg	gggatccact	1140
agttctagag	ggacagcccc	cccccaaac	ccccaggat	gtaattacgt	ccctcccccg	1200
ctagggggca	gcagcagacc	gcccggggct	ccgctccggt	ccggcgtcc	ccccgcatec	1260
ccgagccggc	agcgtgctgg	gcagcccg	gcacgggaa	ggtagcagc	gatcgcttc	1320
ctctgaacgc	ttctcgctgc	tctttgagcc	tgcagacac	tggggggata	cggggaaaag	1380
gcctccaagg	ccagcttccc	acaataagtt	gggtgaatt	tggtcattc	ctccttcta	1440
taggattgag	gtcagagctt	tgtagtggga	atctctggt	atgtgtgtca	gtaggggtg	1500
ggaaagtccc	gcatcgatc	ctagcttcgt	gaggctccg	tgcccgtcag	tgggcagagc	1560
gcacatcgcc	cacagctccc	gagaagtgg	ggggaggggt	cggaattga	accgggtgct	1620
agagaaggtg	gctgctggga	aaactgggaa	gtgatgtcgt	gtactggctc	cgctttttc	1680
ccgaggggtg	gggagaaccg	tataaagtg	cagtagctgc	cgtagaacgt	ctttttcgca	1740
acgggtttgc	gcgcagaaca	cggttaagt	ccgtgtgtg	ttcccgggg	cctggcctct	1800
ttacgggtta	gtgcctttgc	gtgccttgaa	ttacttccac	ctggctccag	tacgtgattc	1860
ttgatcccg	gctggagcca	ggggcggggc	ttgcgctta	ggagcccctt	cgctcgtgc	1920
ttgagttgag	gctggcctg	ggcgtgggg	ccgcccgtg	cgaactggt	ggcaccttcg	1980
cgctgtctc	gctgcttctg	ataagtctc	agccattta	aatttttgat	gactgctgc	2040
gacgctttt	ttctgggaag	atagcttgt	aaatgcccc	caggatctgc	acactgggat	2100
ttcggtttt	gggcccggg	ccggcgacgg	ggcccgtgc	ttccagcgc	catgttggc	2160
gaggcggggc	ctgcgagcgc	ggccaccgag	aatcggagc	gggtagctc	aaactggccg	2220
gcctgctctg	gtgctggcc	ctgcgccc	gtgatccgc	ccgcccggg	cggaaggct	2280
ggcccgtcg	ggcaccagt	ctgagccga	aaagtggcc	cttcccggc	ctgctccagg	2340
gggtcaaaa	tgaggagcgc	ggcgtccgg	agagcggcg	ggtgagctc	ccacacaaa	2400
gaaaagggcc	tttccgtcct	cagccgtcgc	ttcatgtgac	ttccagcag	accggcgccc	2460
gtccaggcac	ctcgatagtt	ctggagctt	ttggagtag	ctgcttttag	gtagggggga	2520
ggggttttat	gcatggaggt	ttcccacac	tgagtggtg	gagactgaa	ttaggccagc	2580
ttggcaactg	atgtaattct	cgttggaatt	tgccccttt	gagttggat	cttgggtcat	2640
tctcaagcct	cagacagtgg	ttcaagttt	ttttctcca	tttccaggt	cgtgaaacgc	2700
tggtcgcgcc	catctagagc	aaactagcat	gccaagtga	ccagtgccgt	ttccggtctc	2760
accgcccggc	acgtcggcgg	agcggctgag	ttctggacc	accggctcgg	gtagctcccg	2820
gactctgtgg	aggacagctg	gcgggtgtg	gtccgggag	acgtgacct	gtagctcagc	2880
gcggtccagg	accaggtggt	gcccgaaca	accctggcct	gggtgtgggt	gtagcggcct	2940
ccagagctgt	acgcccagtg	ctcggaggtc	gtgtccacga	acttccggga	cgctccggg	3000
gcggccatga	ccgagatcgg	cgagcagcc	tgggggcggg	agtccgccct	gcgcaaccgc	3060
gcccggcaact	gctgctcact	cgtggccgag	gagcaggact	gaatcctaaga	tacattgatg	3120
agtttgagca	aaaccacaact	agaatcgagt	gaaaaaaatg	ctttatttgt	gaaatttgtg	3180
atgctattgc	tttatttcta	accattataa	cgtgcaataa	acaagttaac	aacaacaatt	3240
gcattcatct	tatgttccag	ggtcaggggg	agggtgtggga	ggttttttaa	agcaagtaaa	3300
acctctacaa	atgtggtatg	gctgattatg	atcccggctgc	ctcgcgcag	catcatagag	3360
cccaccgcat	ctcgctattg	tcttcccaat	cctcccctct	gctgctcctg	3420	
cccacccac	cccccagaat	agaatgacac	ctactcagac	aatgcatgc	aatttctca	3480
ttttataggg	aaaggacagt	gggagtgcca	ccttccaggg	tcaaggaagg	caagggggag	3540
gggcaaacaa	cgatggctg	gcaactagaa	ggcacagctc	aggctgatca	gcccagatt	3600
actttgacag	ctcgtccatg	ccgagagtg	tcccggcggc	ggtagcagac	ttccagcagg	3660
ccatgtgac	gctctctcct	ttgggtctct	tgctcagggc	ggactggtag	ctcaggtagt	3720
ggttgtcggg	cagcagcagc	gggcccgtgc	cgatgggggt	gtagctgctg	tagtggctcg	3780
cgagctgac	gcccgccttc	tcgatgtgt	ggcggatct	gaagtggcc	ttgatgccc	3840
tcttctgctt	gtcggcgggt	atatagacgt	tggtgctgt	gtagttgtac	ttcagcttgt	3900
gcccagggat	gttgcctgct	tccttgaagt	cgatgcccct	cagctcgatg	cggttccaca	3960
gggtgtcggc	ctcgaacttc	accctggcgc	gggtcttcta	gtagcctcgc	ttccttgaaga	4020
agatgggtgc	ctcctggagc	tagccttcgg	gcatggcgga	cttgaagaag	tcgtgctgct	4080
tcatgtggtc	gggttagcgg	gcaagcact	gcaggcccta	gcccaggggt	gtcacgaggg	4140
tgggccaggg	caaggggcagc	ttgcccgtgg	tgacagatcag	cttcagggtc	agctttccgt	4200
aggtggcatc	gcccctcgccc	ctgcgggaca	cgctgaactt	gtggcggttt	acgttcgccc	4260
ccagctcgac	caggatgggc	accaccccgg	tgaacagctc	ctcgccttg	ctcacatagg	4320
gtccaggggt	ctcctccacg	tctccagcct	gcttcagcag	gctgaagtt	gtagctcctc	4380
ttccgtatac	gcccggagatg	cacgcgctgt	aatagaccgc	gctgctggcg	ttccgacagc	4440
cggagatcag	cgtgctctcc	cttggcagg	acatggcgcg	cgggcccag	ttggccagtg	4500
ccacgtggta	cgggagcccc	ggggcgtcgg	gcccagctcg	gctgcagttg	aggtactggt	4560
cgaaactcgg	gaggtccaag	ctggcccaca	gatcggcggc	gggcccagc	ggctcggcgc	4620
tctccagcgg	cggggcctcg	gcggcggcgg	acagcgggcc	gggtagcggg	ccgggctgct	4680
ccaggggtg	gtagtacagg	ccagcagcgg	gcgcccgggg	ggggcggctg	ctgagcgcct	4740
ccgcccaggg	agcccctgag	caaccgccc	ggtcccgcga	caactcgggt	ggcgcgtagg	4800
gcccgcggaa	ggggcgcagc	gcccagctct	ccggcgcgc	ggggcgtggg	aaagaaggcag	4860
cctcgcgggg	ctccagggcg	ttccagaggg	agcgcctcgg	cgtgggcagc	cccagggcct	4920
cgaaactcgg	gcccagcggg	ggcagctcgc	ggaaggcgcg	agccagacca	gacgcccggg	4980
ggaaaggctc	ggggcgtggc	tgcccggggc	ctaatcccgg	gagcaggagg	ccgggctcca	5040
gcccgcgggg	cttgcggcgc	tgcttcttgc	ggcgcggccg	gtactttag	ttgggggtg	5100
cgcgcaagt	ctgcacgcgc	agccgttcgg	cttctccac	gaaggccgc	ttctccgccc	5160
cgttcagctc	cttccacgct	ttgcccagca	tcttctgtag	caccgcgttg	tgaggtccg	5220
ggttctgctg	agccagccgc	ttgcccctct	cctttgccc	caccatgaag	gcttcatagg	5280
gcccgcggat	gcccagactc	tctgcccct	ggcgttccc	gcccgcggcc	gggctgaggg	5340
catagcgc	cggctcgggg	ctgcccgggg	gactgcccgt	ccggcgtggc	ggcagggcgg	5400
gcccgcgggg	cgcggcggag	gcccgcgggg	ccggcggcag	gcccgcctg	ttcagcggcg	5460
cccctgtccc	ccggggccat	gcacagctgc	ggcggggcgg	ccggctcctc	ttgtgcccct	5520
agcccggcgg	cgatctctgc	atggatcctc	caggcagctc	gacggttcc	taaacagact	5580



-continued

```

ctgcttatat aggcctccca cctgacacgc ctacctcgac atacgttctc taccactgat 5640
agggagtaaa ctcgacatac gttctctatc actgataggg ataaactcga catacgttct 5700
ctatcactga tagggagtaa actcgacata cgttctctat cactgatagg gagtaaaactc 5760
gacatacgtt ctctatcact gatagggagt aaactcgaca tcgttctcta tcaactgatag 5820
ggagtaaaact cgcacatacgt tctctatcac tgatagggag taaactcgct cgagacgcgt 5880
ataccatagga tccggcggcg ctcgagggtg cctcacagga acgaagtccc taaagaaca 5940
gtggcagcca ggttagccgc cggaaatgac tggattcctt ttttagggcc cattggtatg 6000
gctttttccc cgtatccccc caggtgtctg caggctcaaa gagcagcgag aagcgttcag 6060
aggaagcgga tcccggtcca ccttccccct gcccgggctg tccccgcacg ctgcccggctc 6120
gggatgcggg ggggagcgcg ggaccggagc ggagcccggg gcggctcgct gctgcccct 6180
agcgggggag ggacgttaatt acatcccctg gggctttggg ggggggctgt cctctagag 6240
cggccggcac cgcggtggag ctccagcttt tgttccctt agtgagggtt aattagatct 6300
taatacgaact caactatagg gcaatgggtt accgggcccc cctcagagt cgaaggatc 6360
gataagcttg atatctataa caagaaaata tatatataat aagttatcac gtaagtagaa 6420
catgaaataa caataataat atcgtatgag ttaaatctta aaagtacagt aaaagataat 6480
catgctgatc tttgactcac gcggtcgtta tagttcaaaa tcagtgcacac ttaccgcat 6540
gacaagcacg cctcacggga gctccaagcg gcgactgaga tgtcctaact gcacagcgac 6600
ggattcggcg ttttagaaaa gagagagcaa tatttcaaga atgcatgcgt caattttacg 6660
cagactatct tcttagggtt aatctagctg catcaggatc atatgctcgg gcttttttc 6720
cggctcagtc atcgcccagg ctggcgctat ctgggcatcg gggaggaaga agcccgtgc 6780
ttttcccggc aggttgaagc ggcattggaa gagtttggc aggatgactg ctgctgcat 6840
gacgttgagc gaaaacgcac gtttaccatg atgatcggg aaggtgtggc catgcaagcc 6900
tttaacggtg aactgtctgt tcaggccacc tgggatacca gttcgtcgcg gcttttccg 6960
acacagttcc ggatggctcag cccgaagcgc atcagcaacc cgaacaatac cggcgacagc 7020
cggaaactgc gtcgctgtgt gcagattaat gacagcggtg cggcgtgagg atattacgtc 7080
agcgaggacg ggtatcctgg ctggatgccc cagaaaatgga catggatacc cctgtagtt 7140
cccggcgggc ggccttggcg taatcatggt catagctgtt tccctgtgta aattgttatc 7200
cgctcacaat tccacaCaac atacgagcgg gaagcataaa gtgtaaaacc tggggtgcct 7260
aatgagtgag tcaactcaca ttaattgctg tgcgctcact gcccgcttc cagtcgggaa 7320
acctgtcgtc ccagctgcat taatgaatcg gccaacgcgc ggggagaggg ggtttgctga 7380
ttggcgctc ttcgcttcc ctgctcactg actcgtcgcg ctcggtcgtt cggctgccc 7440
gagcgggtac agctcactca aaggcggtaa tacggttacc cacagaatca ggggataaacg 7500
caggaagaaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 7560
tgctggcgtt tttctatagg ctccgcccc ctgacgagca tcacaaaaat cgacgctcaa 7620
gtcagaggtg gcaaaacccc acaggactat aaagatacca ggcgtttccc cctggaagct 7680
cctcgtgccc ctctcctggt ccgaccctgc cgttaccgg atacctgccc gcctttctcc 7740
cttcgggaaag cgtggcgtt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 7800
tcgttcgctc caagctgggc tgtgtgcacg aacccccctg tcagcccgc cgtgcccct 7860
tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccaactggcag 7920
cagccactgg taaccaggatt acacagcga ggtatgtagg cggatcaca gagttcttga 7980
agtgggtggc taactacggc tacactagaa ggacagtatt tggatctgc gctctgctga 8040
agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaaaa accaccgctg 8100
gtagcgggtg ttttttggtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 8160
aagatccttt gatctttctt acggggtctg acgctcagtg gaacgaaaa tcacgttaag 8220
ggattttggt catgagatata tcaaaaagga tcttcaacta gatcctttaa aatataaaat 8280
gaagttttaa atcaatctaa agtatatatg agtaaaactg gctcagcagt taccaatgct 8340
taactcagtg ggcacctatc tcagcgatct gcttatttcg ttcaccata gttgctgac 8400
tcccctgctg gtatagatac acgatacggg agggcttacc atctggcccc agtgcgcaa 8460
tgataccgcg agaccacgc tcaccggctc cagatttacc agcaataaac cagccagccg 8520
gaaggggcga ggcgagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaat 8580
gttgcgggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgtgcca 8640
ttgctacagg caactgggtg tcacgctcgt cgtttggtat ggttctctc agctccggtt 8700
cccaacgatc aaggcagatt acatgatccc ccatgttctg caaaaaagcg gttagctcct 8760
tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 8820
cagcactgca taattctctt actgtcatgc catccgtaag atgctttctt gtgactggtg 8880
agtactcaac caagtcattc tgagaatagt gtatgcccgc accgagttgc tcttgcccgg 8940
cgtcaatacgg gataataacc gcgccacata gcagaacttt aaaaagtctc atcattggaa 9000
aacgttcttc gggcgcaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 9060
aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 9120
gagcaaaaaac aggaagcaca aatgccgcaa aaaagggaat aaggcgacac cggaaatggt 9180
gaatactcat
    
```

```

SEQ ID NO: 16      moltype = DNA length = 1152
FEATURE
source            Location/Qualifiers
                  1..1152
                  mol_type = other DNA
                  organism = synthetic construct
    
```

```

SEQUENCE: 16
atgcagagat cgccgccccg ctacggcgca caggacgacc cgcccgcccc cgcgactgt 60
gcatgggccc cgggacacg ggcggccgct gacacgcgcg gcctcgcgc cggccccgc 120
gccctcggcg cggccgcgcg tcccgcctcg ccgcccagcc cgcagcgcag tcccccgcg 180
agccccgagc cggggcgcta gggcctcagc ccggccggcc ggggggaaac ccaggcggca 240
gacgagtcgc gcatccggcg gcccatgaac gccctcatgg tgtgggcaaa ggacgagcgc 300
aagcggctgg ctcagcagaa cccggacctg cacaacgcgg tgctcagcaa gatgctgggc 360
aaagcgtgga aggagctgaa ccggcgggag aagcggccct tcgtggagga agccgaacgg 420
ctgcccgtgc agcacttgcg cgaccacccc aactacaagt accggcccgc ccgcaagaag 480
    
```

-continued

```

caggcgcgca aggcccgggc gctggagccc ggccctctgc tcccgggatt agcgcceccg 540
cagccaccgc ccgagccttt ccccgcgggc tctggctcgg ctcgcgcctt cgcgagctg 600
cccccgctgg gcgcccagtt cgacggcctg ggggtgccc cgcgccagcg ctgcctctg 660
gacggcctgg agcccgcgca ggctgccttc ttcccaccgc cgcgcgccc cgaggactgc 720
gcgctgcggc ccttccgcgc gccctacgcg cccaccgagt tgtcgcggga ccccgcggt 780
tgctacgggg ctccccctggc ggaggcgctc aggaccgcgc cccccggcg gccgctcgt 840
ggcctgtact acggcacccc gggcacgccc ggcccgtacc cggcccctgt gtcgcccgcg 900
cccagggccc cgcgctgga gagcgcgag ccgctggggc cgcgcgcga tctgtgggc 960
gacgtggacc tcaccagatt cgaccagtac ctcaactgca gccggactcg gcccgagcc 1020
ccccggctcc cgtaccagct ggcactggcc aaactggggc cgcgcgccat gtcctgcca 1080
gaggagagca gctgatctc cgcgctgctg gacgcccaga gcgcggtcta ttacagcgcg 1140
tgcattctcg gc 1152

```

```

SEQ ID NO: 17      moltype = AA length = 384
FEATURE           Location/Qualifiers
source            1..384
                  mol_type = protein
                  organism = synthetic construct

```

```

SEQUENCE: 17
MQRSPPGYGA QDDPPARRDC AWAPGHGAAA DTRGLAAGPA ALAAPAAPAS PPSPPQRSPPR 60
SPEPGRYGLS PAGRGERQAA DESRIIRPMN AFMVWAKDER KRLAQONPDL HNAVLKMLG 120
KAWKELNAAE KRPFVEAER LRVQHLRDHP NYKYRPRRKK QARKARLEP GLLPLGLAPP 180
QPPPEPPPAA SGSARAFREL PPLGAEFDGL GLPTPERSPL DGLEPGEAAF FPPPAAPEDC 240
ALRPFPRAPYA PTELSRDPGG CYGAPLAEAL RTAPPAAPLA GLYYGTGLTP GPYPGPLSP 300
PEAPPLESAE PLGPAADLWA DVLTEFDQY LNCSRTRPDA PGLPYHVALA KLGPAMSC 360
ESSLISALS DASSAVVYSA CISG 384

```

```

SEQ ID NO: 18      moltype = DNA length = 7741
FEATURE           Location/Qualifiers
source            1..7741
                  mol_type = other DNA
                  organism = synthetic construct

```

```

SEQUENCE: 18
actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata 60
catatttgaa tgtatttaga aaaataaaca aataggggtt cgcgcacat ttcccgaaa 120
agtgccacct aaattgtaag cgttaatat ttgttaaaa tcgcgtaaa tttttgtaa 180
atcagctcat ttttaacaa ataggccgaa atcggcaaaa tccctataa atcaaaaga 240
tagaccgaga tagggttgag tgttgtcca gtttggaaca agagtccact attaaagaac 300
gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg gcgtagccc actacgtgaa 360
ccatcacctc aatcaagttt ttggggctg aggtgcccga aagcaactaa tcggaacct 420
aaagggagcc cccgatttag agcttgacc ggaaaagcgg cgaacgtggc gagaaagaa 480
gggaagaaa cgaaaggagc gggcgctagg gcgctggcaa gtgtagcggc cacgctgccc 540
gtaaccacca caccgcccgc gcttaatgcg ccgctacagg gcgctccca ttccgattc 600
aggctgcgca actggtggga agggcgatcg gtgcccctt cttcgctatt acgccagctg 660
gcgaaggggg gatgtgctgc aaggcgatta agttgggtaa cgcaggggtt ttcccagta 720
cgacgttgta aaacgacggc cagtgagcgc gcctcgttca ttcacgtttt tgaacctcg 780
gaggacgggc agactcggc tgcaaatgtg ttttacagcg tgatggagca gatgaagatg 840
ctcgacacgc tgcagaacac gcagctagat taacctaga aagataatca tattgtgacg 900
tacgttaaag ataactatga gtaaatgta cgcattgtgt ttatcggtct gatatcgag 960
gtttatztat taatttgaat agatattaag ttttattata tttacactta cataactaata 1020
ataaatcoa caaacaattt atttatggtt atttattat taaaaaaaaa caaaaactca 1080
aaatttcttc taaaagttaa caaaaactttt atcgaattcc tgcagcccgg gggatccact 1140
agttctagag ggacagcccc ccccaaaagc ccccagggat gtaattacgt cctcccccg 1200
ctagggggca gcagcagagc gccccgggct ccgctccggt cgcgctccc ccccgcatcc 1260
ccgagccggc agcgtgcccg gcacgcccgg gcacggggaa ggtggcacgg gatcgcttc 1320
ctctgaacgc ttctcgctgc tctttgagcc tgcagacacc tgggggggaa cggggaaaag 1380
gcctccaagg ccagcttccc acaataagtt ggggtgaatt tggctcattc ctctttcta 1440
taggatgag gtcagagctt tgtgatggga atctgtgga atgtgtgca gttagggtg 1500
ggaaagtccc gcgactgctg agagatctct agcttcgta ggctccggtg cccgtcagtg 1560
ggcagagcgc acatcgcccga cagtccccga gaagtgggg ggaggggtcg gcaattgaa 1620
cggtgcttag agaaggtggc cgggggtaaa ctgggaaaag gatgtcgtg actggctccg 1680
cctttttccc gaggggtggg gagaaccgta tataagtgca gtatgcccg tgaacgttct 1740
tttcgcaac gggtttgccc ccagaacaca ggtaagtgcc gtgtgtggtt ccccggggcc 1800
tggcctcttt acgggttatg gccctgctg gccttgaatt acttccactt ggctccagta 1860
cgtgattctt gatcccagc tggagcccgg gcgcccctt gcgctttag agccccttcg 1920
cctcgtgctt gagttgagc ctggcctggg cgtggggccc gcccgctgcg aatctggtgg 1980
caccttcgcg cctgtctcgc tgccttcgat aagtctctag ccatttaaaa tttttgatga 2040
cctgctgca gcgttttttt ctggcaagat agtcttgtaa atcggggcca ggatctgca 2100
actggtatct cggtttttgg gcccgcgccc ggccgacgggg cccgtgcctc ccagcgcaca 2160
tgctcggcga ggcggggcct gcgagcggc ccaccgagaa tcggacgggg gtatgctcaa 2220
gctggcccggc gctcctggtt gcctgcctc gcgcccgcgt gatcgcccc gccctggcg 2280
gcaaggctgg cccggtggcc accagttgct tgagcggaaa gatggccgct tcccggcct 2340
gctccagggg gctcaaatg gaggacggcg cgtcgggag agcggggcgg tgagtcaccc 2400
acacaaggga aaagggcctt tccgtcctca gccgtcgtt catgtgactc cacggagtac 2460
cgggcccgtc caggccacct cgattagttc tggagctttt ggagtaactc gctcttaggt 2520
tggggggagg ggttttatgc gatggagttt ccccaactg agtgggtgga gactgaagtt 2580

```

-continued

aggccagctt	ggcacttgat	gtaattctcg	ttggaatttg	ccctttttga	gtttggatct	2640
tggttcattc	tcaagcctca	gacagtgggt	caaagttttt	ttcttccatt	tcagggtgctg	2700
tgaacacogtg	gtcgcggcca	tctagagcaa	ctagcttggg	cccggaatt	aatccaccat	2760
gtctagactg	gacaagagca	aagtcataaa	cggcgctctg	gaattactca	atggagtctg	2820
tatcgaagcg	ctgacgacaa	ggaaactcgc	tcaaaagctg	ggagttgagc	agcctaccct	2880
gtactggcac	gtgaagaaca	agcgggccc	gctcgatgcc	ctgccaatcg	agatgctgga	2940
caggcatcat	accactctc	gccccctgga	aggcgagtca	tggcaagact	ttctgcgaa	3000
caacgccaa	ttctccgct	gtgctctcct	ctcacatcgc	gacgggcta	aagtgcattct	3060
cggcaccgcg	ccaacagaga	aacagtacga	aacctggaa	aatcagctcg	cgttcctgtg	3120
tcagcaagcg	ttctccctgg	agaacgcact	gtacgctctg	tccgcccctg	gccactttac	3180
actgggtgctg	gtattggagg	aacaggagca	tcaagtagca	aaagaggaaa	gagagacacc	3240
taccaccgat	tctatgcccc	cacttctgag	acaagcaatt	gagctgtctg	accggcaggg	3300
agccgaacct	gcccctcttt	tcggcctgga	actaatcata	tgtggcctgg	agaaacagct	3360
aaagtgcgaa	agcggcgggc	cggccgacgc	ccttgacgat	tttgacttag	acatgctccc	3420
agccgatgcc	cttgacgact	ttgacctga	tatgctgccc	gctgacgctc	ttgacgattt	3480
tgaccttgac	atgctcccc	gggaattctg	aagcggagag	ggcagaggaa	gtcttctaac	3540
atgctggtgac	gtggagagaga	atccccgccc	tgatatcatg	accgagtaca	agccccaggt	3600
gctcctcgcc	accgcgacg	acgtcccccg	ggcgtacgc	accctcgcg	ccgcttctgc	3660
cgactaccoc	gccacgcgcg	acaccgtcga	cccgaccgc	cacatcgagc	gggtcaaccga	3720
gctgcaagaa	cttctctca	cgcgctcgg	gctcgacatc	ggcaaggtgt	gggtcgcgga	3780
cgacggcgc	gctgctggcg	tctggaccac	gcccggagagc	gtcgaagcgg	ggcggtggt	3840
cgccgagatc	ggccgcgcga	ggcccgagtt	gagcggttcc	cggtggccg	cgacgcaaca	3900
gatggaagcg	ctctctggcg	cgcaccggcc	caaggagccc	gctggttcc	tggccaccgt	3960
cgctgctcgc	cccgaccacc	agggcaaggg	tctgggcagc	gcccgtctgc	tccccggagt	4020
ggaggcggcc	gagcgcgcg	gggtgcccgc	cttctggag	acctccgcg	cccgaacct	4080
ccccctctac	gagcggctcg	gcttaccctg	caccgcgcac	gtcagagtg	ccgaaggacc	4140
gctgacccctg	gtcatgacc	gcaagcccgg	tgctgataaa	gatcacatga	tgagtttggga	4200
caaaaccacaa	tagaatgca	gtgaaaaaaa	tgctttattt	gtgaaatttg	tgatgctatt	4260
gctttatttg	taaccattat	aagctgcaat	aaacaagtta	acaacaacaa	ttgcattcat	4320
tttatgcttc	aggttcaggg	ggaggtgtgg	gaggtttttt	aaagcaagta	aaacctctac	4380
aaatgtggta	tggctgatta	tgatccggct	gcctcgcgcg	gtacgacgcg	tatacctagg	4440
atccggcggc	cttcgagggt	tcctcacagg	aacgaagtcc	ctaagaaaac	agtggcagcc	4500
aggttttagcc	ccggaattga	ctggattcct	tttttagggc	ccattggtat	ggctttttcc	4560
ccgtatcccc	ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaagcgc	4620
atccccgtcc	accttccccg	tgccccgggt	gtccccgcac	gctgcccgtc	cggggatgcg	4680
gggggagcgc	cggaccggag	cgagcccccg	ggcggctcgc	tgctgcccc	tagcggggga	4740
gggacgtaat	tacatccctg	ggggctttgg	ggggggctg	tccctctaga	gcccgcgcca	4800
ccgctggtga	gctccagctt	ttgttccctt	tagtgagggg	taattagatc	ttaatcgcac	4860
tactataggg	gcaaatggg	taccgggccc	ccccctgagg	tcgacggtat	cgataagctt	4920
gatattctata	acaagaaaaa	atatataaa	taagttatca	cgtaaagta	acatgaaata	4980
acaatataat	tatcgtatga	gttaaatctt	aaaagtcacg	taaaagataa	tcatgcgtca	5040
ttttgactca	cgcggctggt	atagttcaaa	atcagtgaca	cttaccgcat	tgacaagcac	5100
gctcaccggg	agctccaagc	ggcgcactgag	atgtcctaaa	tgacacagcga	cggatctcgc	5160
ctattttagaa	agagagagca	atatttcaag	aatgcactgc	tcaattttac	gcagactatc	5220
ttcttagggg	taactcagct	gcatcaggat	catatcgtcg	ggctttttt	ccggtctcagt	5280
catcgcccaa	cctggcgcata	ctgggcatc	ggggaggaag	aagcccgtgc	cttttccccg	5340
gaggttgaag	ggcactggaa	agagtttggc	gaggtgact	gctgctgcat	tgacgttgag	5400
cgaaaacgca	cgtttaccat	gatgattcgg	gaaggtgtgg	ccatgcaacg	ctttaaccgt	5460
gaactgtctg	ttcagggccc	ctgggatacc	agttcgtcgc	ggcttttccg	gacacagttc	5520
cggtatggtca	gcccggaagc	catcagcaac	ccgaacaata	ccggcgacag	ccggaactgc	5580
cgtgcccgtg	tgacagataa	tgacagcggg	gcccgcgctg	gatattacgt	cagcggagac	5640
gggtatcctg	gctggatgcc	gcagaaatgg	acatggatc	cccgtgagtt	accocggcgg	5700
cgcgcttggc	gtaactcagg	tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacia	5760
ttccacacaa	catcagagcc	ggaagcataa	agtgtaaacg	ctgggggtgc	taatgagtga	5820
gctaactcac	attaattggc	ttgcccctac	tgcccgcttt	ccagtcggga	aacctgtcgt	5880
gccagctgca	ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgctg	attgggcgct	5940
cttccgcttc	ctcgcctcact	gactcgcctc	gctcggctg	tcggctcggg	cgagcgggat	6000
cagctcactc	aaaggcggta	ataccggtat	ccacagaatc	aggggataac	gcaggaaaga	6060
acatgtgagc	aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgcg	ttgctggcgt	6120
ttttccatag	gctcccgcct	cctgacgagc	atcacaaaaa	tcgacgctca	agtcagaggt	6180
ggcgaaaacc	gacagggacta	taaaataacc	agggctttcc	ccctggaaagc	tccctcgtgc	6240
gctctcctgt	tccgaccctg	ccgcttaccg	gatacctgtc	cgcttttctc	ccttcgggaa	6300
gctggtggct	ttctcatagc	tcacgctgta	ggtatctcag	ttcgggttag	gtcgttcgct	6360
ccaagctggg	ctgtgtgca	gaaccccccg	ttcagcccga	ccgctgcgcc	ttatccggta	6420
actatcgtct	tgagtccaac	ccggtaaacg	acgacttatc	gccactggca	gcagccaactg	6480
gtaacaggat	tagcagagcg	aggtatgtag	gctggtgctac	agagttcttg	aagtgggtggc	6540
ctaactacgg	ctacacataga	aggacagtat	ttggatctctg	cgctctgctg	aagccagtta	6600
ccttcggaaa	aaaggttggg	agctcttgat	ccggcaaaaa	aaccaccgct	ggtagcgggtg	6660
gtttttttgt	ttgcaagcag	cagattacgc	gcagaaaaaa	aggatctcaa	gaagatcctt	6720
tgatcttttc	tacggggtct	gacgctcagt	ggaacgaaaa	ctcactgtaa	gggattttgg	6780
tcatgagatt	atcaaaaagg	atcttcaact	agatcctttt	aaattaaaaa	tgaagtttta	6840
aatcaatcta	aagtatatag	gagtaaacct	ggtctgacag	ttaccaatgc	ttaatcagtg	6900
agggcacctat	ctcagcgatc	tgtctatctt	gttcatccat	agttgcctga	ctccccgctg	6960
tgtagataac	tacgatccg	gagggcttac	catctggccc	cagtgctgca	atgataaccgc	7020
gagaccacgc	ctcaccggct	ccagatttat	cagcaataaa	ccagcccagc	ggaagggcgg	7080
agcgcagaag	tggtcctgca	actttatccg	cctccatcca	gtctattaat	tgttgcggg	7140

-continued

---

```

aagctagagt aagtagttcg ccagttaata gtttgcgcaa cgttggtgcc attgctacag 7200
gcatcgtggt gtcacgctcg tcgtttggtg tggttcatt cagctccggg tcccaacgat 7260
caaggcgagt tacatgatcc cccatggtgt gcaaaaaagc ggtagctcc ttcggctctc 7320
cgatcgttgt cagaagtaag ttggcgcag tgttatcact catggtatg gcagcactgc 7380
ataattctct tactgtcatg ccatccgtaa gatgcttttc tgtgactggt gactactcaa 7440
ccaagtcatc ctgagaatag tgtatgcggc gaccgagttg ctcttgccc gcgtcaatac 7500
gggataatac cgcgccacat agcagaactt taaaagtgct catcattgga aaacgttctt 7560
cggggcgaaa actctcaagg atcttaccgc tgttgagatc cagttcgatg taaccactc 7620
gtgcacccaa ctgatcttca gcattcttta ctttcaccag cgtttctggg tgagcaaaaa 7680
caggaaggca aaatgccgca aaaaagggaa taaggcgac acggaatgt tgaataactca 7740
t

```

```

SEQ ID NO: 19          moltype = DNA length = 17
FEATURE              Location/Qualifiers
source               1..17
                    mol_type = other DNA
                    organism = synthetic construct

```

```

SEQUENCE: 19
ggcaaagcgt ggaagga 17

```

```

SEQ ID NO: 20          moltype = DNA length = 19
FEATURE              Location/Qualifiers
source               1..19
                    mol_type = other DNA
                    organism = synthetic construct

```

```

SEQUENCE: 20
cggccggtac ttgtagttg 19

```

---

1. A method for increasing stem cell differentiation to produce erythromyeloid progenitors, megakaryocytes and NK cells but not differentiated T cells, comprising enforcing expression of SOX18 in hemogenic endothelial cells during their transition from endothelium to hematopoietic cell progenitors.

2. The method of claim 1, wherein the cells produced are enriched in NK progenitor cells.

3. The method of claim 2, wherein NK progenitor cell production is increased about five-fold compared with hemogenic epithelial cells wherein SOX18 production is not enforced.

4. The method of claim 1, wherein SOX18 expression is enforced after introduction of an inducible SOX18-encoding expression construct.

5. The method of claim 1, wherein SOX18 expression is enforced after introduction of a constitutively expressing SOX18-encoding expression construct.

6. The method of claim 1, wherein SOX18 expression is enforced by introducing a modified mRNA (mmRNA) encoding SOX18 into the cell.

7. The method of claim 4, wherein the inducible SOX18 expression construct is enforced by contacting the cells with doxycycline.

8. A NK progenitor cell culture produced by the method of claim 1.

9. A composition comprising the NK progenitor cell culture of claim 8.

10. A pharmaceutical composition comprising differentiated NK cells produced from the composition of claim 9 and pharmaceutically (or therapeutically) acceptable excipients and adjuvants.

11. An immunotherapeutic method comprising administration of a therapeutically effective amount of the pharmaceutical composition of claim 10 to an individual in need thereof.

12. A megakaryocyte progenitor cell culture produced by the method of claim 1.

13. A composition comprising the megakaryocyte progenitor cell culture of claim 12.

14. A pharmaceutical composition comprising differentiated megakaryocyte cells produced from the composition of claim 13 and pharmaceutically (or therapeutically) acceptable excipients and adjuvants.

15. An immunotherapeutic method comprising administration of a therapeutically effective amount of the pharmaceutical composition of claim 14 to an individual in need thereof.

16. A composition comprising a differentiated NK cell population produced according to the method of claim 1.

17. A composition comprising a differentiated megakaryocyte cell population produced according to the method of claim 1.

\* \* \* \* \*