Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2835825 C 2021/03/23

(11)(21) **2 835 825**

(12) **BREVET CANADIEN**
CANADIAN PATENT

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 2012/05/11
(87) Date publication PCT/PCT Publication Date: 2012/11/22
(45) Date de délivrance/Issue Date: 2021/03/23
(85) Entrée phase nationale/National Entry: 2013/11/12
(86) N° demande PCT/PCT Application No.: US 2012/037643
(87) N° publication PCT/PCT Publication No.: 2012/158561
(30) Priorités/Priorities: 2011/05/13 (US61/486,004); 2011/08/30 (US61/529,055)

(51) Cl.Int./Int.Cl. *C12N 5/10* (2006.01), *C12N 15/00* (2006.01), *C12N 15/85* (2006.01), *C12N 5/07* (2010.01)

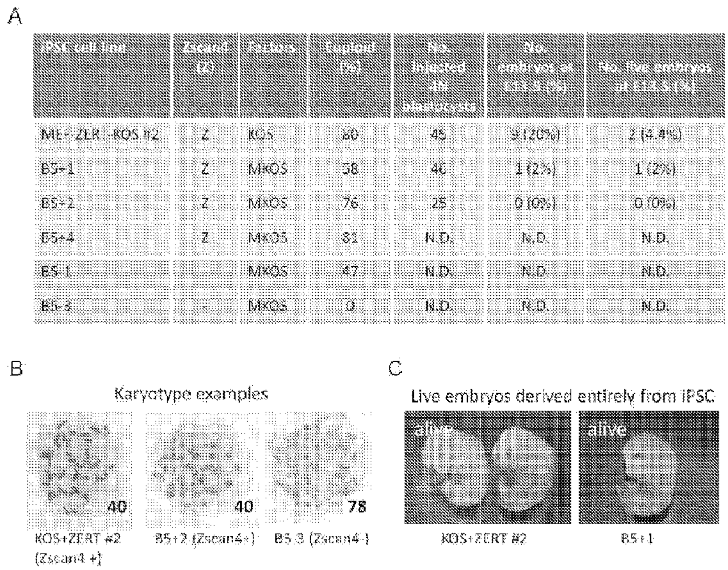
(72) Inventeurs/Inventors:
KO, MINORU S.H., US;
HIRATA, TETSUYA, US

(73) Propriétaire/Owner:
ELIXIRGEN, LLC, US

(74) Agent: BENOIT & COTE INC.

(54) Titre : UTILISATION DE ZSCAN4 ET GENES DEPENDANT DE ZSCAN4 POUR REPROGRAMMATION DIRECTE DE CELLULES SOMATIQUES

(54) Title: USE OF ZSCAN4 AND ZSCAN4-DEPENDENT GENES FOR DIRECT REPROGRAMMING OF SOMATIC CELLS



(57) Abrégé/Abstract:
Disclosed herein is the finding that Zscan4 is an early embryonic factor that facilitates cellular reprogramming. In particular, Zscan4 can replace the oncogenic reprogramming factor c- Myc to produce induced pluripotent stem cells when coexpressed with Klf4, Oct4 and Sox2. In addition, several Zscan4-dependent genes were identified that promote iPSC formation when co- expressed with known reprogramming factors. Thus, the present disclosure provides an ex vivo method of producing an iPS cell by reprogramming of a somatic cell. The method includes contacting the somatic cell with a Zscan4, or a Zscan4- dependent gene, and at least one reprogramming factor. Also provided are iPS cells produced by the disclosed method and non- human animals generated from such iPS cells.

CA 02835825 2013-11-12

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 November 2012 (22.11.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/158561 A1

(51) International Patent Classification:
C12N 5/00 (2006.01) *C12N 15/63* (2006.01)
C12N 15/00 (2006.01) *C12N 5/10* (2006.01)

(21) International Application Number:
PCT/US2012/037643

(22) International Filing Date:
11 May 2012 (11.05.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/486,004 13 May 2011 (13.05.2011) US
61/529,055 30 August 2011 (30.08.2011) US

(71) Applicant *(for all designated States except US)*: **THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY, DEPT. OF HEALTH AND HUMAN SERVICES** [US/US]; Office Of Technology Transfer, National Institutes, of Health, 6011 Executive Boulevard, Suite 325 MSC 7660, Bethesda, MD 20892-7660 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **KO, Minoru, S.H.** [US/US]; 10516 Willow Vista Way, Cockeysville, MD 21030 (US).

(74) Agents: **LITTLEFIELD, Otis** et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).

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

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

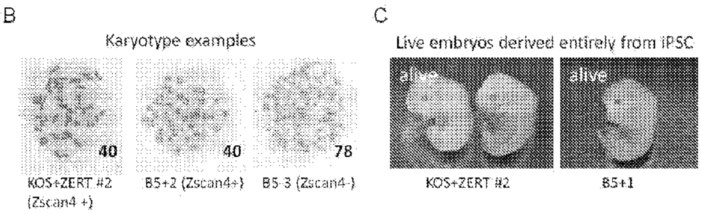
[Continued on next page]

(54) Title: USE OF ZSCAN4 AND ZSCAN4-DEPENDENT GENES FOR DIRECT REPROGRAMMING OF SOMATIC CELLS

FIGURE 18

A

iPSC cell line	Zscan4 (Z)	Factors	Euploid (%)	No. injected 4N blastocysts	No. embryos at E13.5 (%)	No. live embryos at E13.5 (%)
MEF-ZERT-KOS #2	Z	KOS	80	45	9 (20%)	2 (4.4%)
B5+1	Z	MKOS	58	46	1 (2%)	1 (2%)
B5+2	Z	MKOS	76	25	0 (0%)	0 (0%)
B5+4	Z	MKOS	81	N.D.	N.D.	N.D.
B5-1	-	MKOS	47	N.D.	N.D.	N.D.
B5-3	-	MKOS	0	N.D.	N.D.	N.D.



(57) Abstract: Disclosed herein is the finding that Zscan4 is an early embryonic factor that facilitates cellular reprogramming. In particular, Zscan4 can replace the oncogenic reprogramming factor c- Myc to produce induced pluripotent stem cells when co-expressed with Klf4, Oct4 and Sox2. In addition, several Zscan4-dependent genes were identified that promote iPSC formation when co- expressed with known reprogramming factors. Thus, the present disclosure provides an ex vivo method of producing an iPS cell by reprogramming of a somatic cell. The method includes contacting the somatic cell with a Zscan4, or a Zscan4- dependent gene, and at least one reprogramming factor. Also provided are iPS cells produced by the disclosed method and non- human animals generated from such iPS cells.

WO 2012/158561 A1

WO 2012/158561 A1

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

USE OF ZSCAN4 AND ZSCAN4-DEPENDENT GENES FOR DIRECT REPROGRAMMING OF SOMATIC CELLS

FIELD

[0001] This disclosure concerns the identification of Zscan4 as an early embryonic factor required for direct reprogramming of somatic cells. This disclosure further concerns the use of Zscan4 and Zscan4-dependent genes for production of induced pluripotent stem (iPS) cells, iPS cells produced by this method, and methods of using the generated iPS cells.

BACKGROUND

[0002] Induced pluripotent stem cells hold great promise as a source of patient-specific cells in regenerative medicine, but there are many challenges that must be overcome before this technology can be applied effectively in clinical therapies (Hanna *et al.*, *Cell* 143:508-525, 2010; Yamanaka, *Cell* 137:13-17, 2009; Stadtfeld *et al.*, *Genes Dev* 24:2239-2263, 2010). One critical issue is the use of an oncogene, c-Myc (M), together with other three factors KOS (Klf4, Oct4, Sox2) to generate induced pluripotent stem (iPS) cells. The ectopic overexpression of KOS alone leads to a low efficiency of iPS cell formation (Takahashi and Yamanaka, *Cell* 126:663-676, 2006). However, the use of oncogenes raises serious concern about tumorigenicity of iPS cells and long-term safety in potential clinical use of iPS cells. Indeed, recent reports show that even after successful reprogramming, iPS cells tend to show low genome stability and premature cellular senescence upon differentiation (Feng *et al.*, *Cell Stem Cell* 4:301-312, 2009; Hu *et al.*, *Proc Natl Acad Sci USA* 107:4335-4340, 2010). However, increasing cell proliferation and suppressing genome stability by c-Myc seems to be inseparable from efficient induced pluripotent stem cell (iPSC) formation (Zhao *et al.*, *Cell Stem Cell* 3:475-479, 2008). Thus, a fundamental challenge that must be addressed is how to increase efficiency of iPS cell generation without sacrificing genome stability. Without resolving this issue, iPS cells may never be usable in clinical practice.

[0003] The Zscan4 (zinc finger and scan domain-containing protein 4) gene was identified by expression profiling of all preimplantation stages of mouse embryos using a large-scale cDNA sequencing project (Ko *et al.*, *Development* 127:1737-1749, 2000; Sharov *et al.*, *PLoS Biol* 1:E74, 2003) and DNA microarray analysis (Hamatani *et al.*, *Dev Cell* 6:117-131, 2004). In mice, Zscan4 consists of 6 paralog genes (Zscan4a to Zscan4f) and 3 pseudogenes (Zscan4-ps1

to Zscan4-ps3) clustered on an approximately 850 kb region of chromosome 7. Among the six paralogs, the open reading frames of Zscan4c, Zscan4d, and Zscan4f encode a SCAN domain as well as all four zinc finger domains, suggesting their potential role as transcription factors. A high expression peak of Zscan4 marks the late 2-cell stage of mouse embryos. Zscan4 expression, normally below detection threshold in blastocysts, is reactivated *in vitro* in a small fraction of embryonic stem (ES) cells in culture. It has previously been demonstrated that Zscan4 acts critically in the formation of proper blastocysts (Falco *et al.*, *Dev Biol* 307:539-550, 2007; PCT Publication No. WO 2008/118957) and is required for the maintenance of genome stability and normal karyotype in ES cells (Zalzman *et al.*, *Nature* 464:858-863, 2010; PCT Publication No. WO 2011/028880).

SUMMARY

[0004] Disclosed herein is the finding that Zscan4 initiates direct reprogramming of somatic cells by reactivating early embryonic genes. Forced expression of Zscan4 in somatic cells, along with other previously described reprogramming factors, leads to the efficient production of high quality induced pluripotent stem (iPS) cells. It is also disclosed herein that expression of the Zscan4-dependent genes Patl2, Pramel6, Piwil2 and D5Ertd577e in somatic cells promotes induction of iPS cells.

[0005] Provided herein is an *ex vivo* method of producing an iPS cell by reprogramming of a somatic cell. The method includes contacting the somatic cell with a Zscan4, or a Zscan4-dependent gene, and at least one reprogramming factor, thereby producing an iPS cell. In some embodiments, the Zscan4-dependent gene is selected from Patl2, Pramel6, Piwil2 and D5Ertd577e. In some embodiments, the method includes contacting the somatic cell with at least two, at least three, or at least four reprogramming factors. The reprogramming factors for use in the disclosed methods include, but are not limited to, c-Myc, Klf4, Oct4, Sox2, Lin28 and Nanog. In some embodiments, the method includes contacting the somatic cell with a Zscan4, at least one Zscan4-dependent gene and at least one reprogramming factor. Also provided are isolated iPS cells produced according to the methods disclosed herein. Use of the isolated iPS cells for therapeutic applications is further provided by the present disclosure.

[0006] Also provided herein are methods of identifying mature and/or high quality iPSCs in a cell population by transfecting the cell population with an expression vector comprising a Zscan4 promoter operably linked to a reporter gene, wherein expression of the reporter gene in a cell of the cell population identifies the cell as a mature and/or high-quality iPSC. Further

provided is a method of isolating mature iPSCs from a cell population, comprising transfecting the cell population with an expression vector comprising a Zscan4 promoter operably linked to a reporter gene, and separating cells expressing the reporter gene from the cell population, thereby isolating mature iPSCs.

[0007] The foregoing and other objects and features of the disclosure will become more apparent from the following detailed description, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] **FIG. 1** shows that Zscan4 enhances iPSC formation without Myc. **FIG. 1A** depicts a schematic of the structure of pCAG-Zscan4cERT2 plasmid used to make murine embryonic fibroblast (MEF)-ZERT cells. **FIG. 1B** depicts a schematic representation of experimental procedures for iPSC generation. **FIG. 1C** depicts representative pictures of 6-well plates stained for ALP 20 days after the doxycycline (Dox) induction. Alkaline phosphatase (ALP) positive colonies were counted (mean±S.E.M.) and the results are shown in the graph on the right. **FIG. 1D** depicts the efficiency of iPSC formation examined after different Tamoxifen (Tmx) treatments. **FIG. 1E** also depicts the efficiency of iPSC formation examined after different Tamoxifen (Tmx) treatments. ALP-positive colonies were counted 20 days after the induction (mean±S.E.M.). Different letters denote significant differences between groups ($P < 0.05$).

[0009] **FIG. 2** shows the generation of secondary MEFs and induction of secondary iPSCs. **FIG. 2A** depicts a schematic representation of procedures for secondary MEFs generation. **FIG. 2B** depicts representative cell morphologies during the first 6 days of Dox and Tmx treatments. Pictures of cells after ALP-staining on day 17 are shown. **FIG. 2C** depicts the efficiency of iPSC formation examined after different Tmx treatments. ALP-positive colonies were counted 13 days after Dox induction. Different letters denote significant differences between groups ($P < 0.05$).

[0010] **FIG. 3** shows microarray analysis of the early phase of iPSC formation from the MEF-KOS-ZERT^{2nd} cells. **FIG. 3A** depicts scatter-plots showing pair-wise comparison between Dox+ Tmx-(KOS factors) and Dox- Tmx- (No factor); Dox+ Tmx+ (ZKOS factors) and Dox- Tmx+ (No factor); Dox- Tmx+ (No factor) and Dox- Tmx- (No factor); and Dox+ Tmx+ (ZKOS factors) and Dox+ Tmx-(KOS factors) conditions. Cells were harvested on day 1, 3, and 6 after beginning the Dox or Tmx treatment. Figures in each scatter plot represent the number of genes that showed statistically significant differences between the conditions (FDR≤0.05, fold-

change≥2). A list of non-redundant 231 genes were obtained by combining 12 (day 1), 90 (day 3), and 178 (day 6) genes that were more highly expressed in Dox+ Tmx+ (ZKOS) condition than in Dox+ Tmx- (KOS) condition. **FIG. 3B** depicts a heatmap showing the fold-difference of expression levels of 231 genes between Tmx+ and Tmx-conditions. The fold difference for each gene was calculated by dividing the expression level (Tmx+) by the expression level (Tmx-). Among 231 genes, Pramel6 showed the highest fold-difference: 10.2-fold on day 3 (see Table 1). Results obtained by searching the EST database for 231 genes are shown as symbols after gene names: Red circle, genes expressed predominantly in oocytes; blue circle, genes expressed predominantly in preimplantation embryos (1-cell to blastocysts); pink square, genes expressed predominantly in testes or ovaries. **FIG. 3C** depicts a heatmap showing expression patterns of 201 genes (a subset of 231 genes) found in the NIA Gene Expression Atlas (Sharov *et al.*, *BMC Genomics* 12:102, 2011) (22 different adult organs/tissues and cultured cells, from left to right: brain, cerebellum, eyes, skeletal muscle, heart, bone, liver, kidney, bladder, skin, visceral fat, lung, small intestine, large intestine, stomach, placenta, ovary, oocyte, testis, MEF cells, ESCs, and iPSCs). Bar graphs show the gene expression levels of two representative genes (Patl2 and D13Ertd608e) among these tissues. **FIG. 3D** depicts a heatmap showing the expression patterns of 99 genes (a subset of 231 genes) found in the GNF database (Su *et al.*, *Proc Natl Acad Sci USA* 99:4465-4470, 2002) (62 different organs/tissues). Bar graphs show the gene expression levels of two representative genes (Pramel6 and D5Ertd577e) among these tissues. **FIG. 3E** depicts a summary diagram showing events occurring during Zscan4-mediated iPSC formation.

[0011] **FIG. 4** depicts the characterization of MEF-ZERT cells and summary diagrams. **FIG. 4A** depicts growth curves of MEF-WT and MEF-ZERT cultured and passaged in Tmx+ and Tmx- conditions. **FIG. 4B** depicts scatter-plots showing global gene expression differences between Tmx+ and Tmx- conditions 48 hours after Tmx treatment. **FIG. 4C** depicts DNA methylation patterns of the promoter region of Oct4 gene. **FIG. 4D** depicts a summary diagram of the Zscan4 effect on MEFs.

[0012] **FIG. 5** depicts the generation and characterization of MEF-ZERT cell lines. **FIG. 5A** depicts a schematic representation of procedures for MEF-ZERT generation. A pCAG-Zscan4cERT2 vector was transfected into V6.5 embryonic stem (ES) cells to make Zscan4ERT2 ES cells (ES-ZERT). ES-ZERT cells were microinjected into blastocysts from the ICR mice to generate male chimeric mice, which were subsequently mated with the ICR female mice. E13.5 embryos were dissected out from the pregnant ICR female mice and used to generate mouse embryo fibroblasts (MEFs). MEFs were subjected to genotyping and

quantitative RT-PCR. MEFs that carried pCAGZscan4cERT2 DNA and expressed the exogenous Zscan4c were designated as MEF-ZERT and MEFs that did not were designated MEF-WT (wild type). **FIG. 5B** depicts expression levels of Zscan4c in a series of MEFs (MEF-ZERT and MEF-WT) were examined by qRT-PCR. **FIG. 5C** depicts expression levels of Zscan4c in a different series of MEFs (MEF-ZERT and MEF-WT) were examined by qRT-PCR. Data in triplicate were represented as mean±S.E.M. after calculating a ratio between the expression levels of Zscan4c and those of Gapdh. The following MEF lines were used in the studies disclosed herein: MEF-ZERT (#A2, #A7, #B5); MEF-WT (#A3).

[0013] **FIG. 6** shows that Zscan4 enhances MKOS-mediated iPSC colony formation from the MEF-ZERT cells. **FIG. 6A** depicts a schematic presentation of experimental procedures. A piggyBac vector (PB-TET-MKOS) carrying doxycycline (Dox)-inducible Myc (M), Klf4 (K), Oct4 (O), and Sox2 (S), was transfected into MEF-ZERT and MEF-WT, respectively. The cells were cultured under the Dox+ Tmx- or Dox+ Tmx+ condition for 14 days, fixed, and stained for alkaline phosphatase (ALP). **FIG. 6B** depicts phase-contrast microscopic images showing morphological changes of MEF-ZERT cells during the MKOS-mediated iPSC colony formation. **FIG. 6C** depicts representative pictures of 6-well plates stained for ALP. **FIG. 6D** depicts iPSC colonies that were scored based on their authentic ES cell morphology and ALP-staining (mean $\pm$S.E.M.). *, $P < 0.05$.

[0014] **FIG. 7** depicts the characterization of iPSCs generated from the MEF-WT and MEF-ZERT cells with a PB-TET-MKOS vector. **FIG. 7A** shows several iPSC colonies that were picked from the wells prepared in parallel with the experiment shown in FIG. 6C and propagated in the ES cell culture condition on feeder cells: one iPSC clone from the MEF-WT (Dox+ Tmx-); three iPSC clones from the MEF-WT (Dox+ Tmx+); two iPSC clones from the MEF-ZERT (Dox+ Tmx-); and three iPSC clones from the MEF-ZERT (Dox+ Tmx+). These iPSC clones, MEF-WT cells, MEF-ZERT cells, and V6.5 ES cells were subjected to RT-PCR analysis with pluripotency gene markers: endogenous Oct4 (Pou5f1), endogenous Sox2, Nanog, Zfp42 (Rex1), and Dax1 (Nr0b1). Gapdh was used as a control. **FIG. 7B** depicts a representative phase-contrast image of the iPSC clone (#5B+1) derived from the MEF-ZERT cells with MKOS under Tmx+ condition (*i.e.*, with MKOS factors plus Zscan4 (ZMKOS factors)). **FIG. 7C** depicts a microscopic image of the iPSC clone (#5B+1) after staining with ALP. **FIG. 7D** depicts fluorescence microscopic images of the iPSC clone (#5B+1) after staining with antibodies against SSEA-1 and NANOG. Pictures (right) are the same images after merging with DAPI-staining. **FIG. 7E** depicts a microscopic image showing embryoid bodies (day 4) generated from

the iPSC clone (#5B+1); x200. **FIG. 7F** depicts fluorescence images of the iPSCs after *in vitro* differentiation from the embryoid body shown in FIG. 7E, and stained with antibodies against αSMA (mesoderm), AFP (endoderm), GATA4 (endoderm), and βIII-tubulin (ectoderm). Pictures (right) are the same images after merging with DAPI-staining. Scale bar, 100 mm. **FIG. 7G** depicts an E13.5 embryo derived from the iPSC by the 4N complementation.

[0015] **FIG. 8** shows that Zscan4 enhances MKOS-mediated iPSC colony formation from the MEF-WT cells. **FIG. 8A** depicts a schematic representation of experimental procedures. A PB-MKOS vector together with a PB-TET-DsRed (control), PB-TET-Zscan4, or PB-TET-Zscan4ERT2 vector was transfected into the MEF-WT cells (C57BL/6J x 129S6/SvEvTac). The cells were cultured for 14 days under the Dox+ condition (for the PB-TET-Zscan4ERT2 vector, Dox+ Tmx- or Dox+ Tmx+ condition), fixed, and stained for the ALP. **FIG. 8B** depicts a representative pictures of 6-well plates stained for ALP. **FIG. 8C** depicts ALP-positive colonies that were scored. Transfections and Dox inductions were performed in triplicate. Data from two independent experiments are shown. Data are represented as mean±S.E.M. (triplicate wells); *, $P < 0.01$.

[0016] **FIG. 9** depicts the characterization of iPSC clones derived from the MEF-WT with MKOS and Zscan4. **FIG. 9A** depicts several iPSC colonies that were picked from the wells prepared in parallel with the experiment shown in FIG. 8B and propagated in the ES cell culture condition on feeder cells: five iPSC clones from the MEF-WT (MKOS: #A2, #A3, #A4, #A5, #A6); four iPSC clones from the MEF-WT (ZMKOS: #B1, #B3, #B5, #B6). These iPSC clones and MEF-WT cells were subjected to RT-PCR analysis with pluripotency gene markers: endogenous Oct4 (Pou5f1), endogenous Sox2, Nanog, Zfp42 (Rex1), and Dax1 (Nr0b1). Gapdh was used as a control. **FIG. 9B** depicts a representative phase-contrast image of the iPSC clone (#B5) after staining with ALP. **FIG. 9C** depicts a microscopic image showing embryoid bodies (day 4) generated from the iPSC clone (#B5). **FIG. 9D** depicts fluorescence microscopic images of the iPSC clone (#B5) after *in vitro* differentiation from the embryoid body shown in FIG. 9C, and stained with antibodies against αSMA (mesoderm), AFP (endoderm), GATA4 (endoderm), and βIII-tubulin (ectoderm). Pictures (right) are the same images after merging with DAPI-staining. Scale bar, 100 μm.

[0017] **FIG. 10** depicts that characterization of iPSC clones derived from the MEF-WT with MKOS and Zscan4ERT2 (Tmx+). **FIG. 10A** depicts several iPSC colonies that were picked from the wells prepared in parallel with the experiment shown in FIG. 8B and propagated in the ES cell culture condition on feeder cells: five iPSC clones from the MEF-

WT (MKOS+Zscan4ERT2, Tmx- condition: #C1, #C2, #C3, #C5, #C6); four iPSC clones from the MEF-WT (MKOS+Zscan4ERT2, Tmx+: #D2, #D3, #D5, #D6). These iPSC clones and control MEF-WT cells were subjected to RT-PCR analysis with pluripotency gene markers: endogenous Oct4 (Pou5f1), endogenous Sox2, Nanog, Zfp42 (Rex1), and Dax1 (Nr0b1). Gapdh was used as a control. **FIG. 10B** depicts a representative phase-contrast image of the iPSC clone (#D3) after staining with ALP. **FIG. 10C** depicts a microscopic image showing embryoid bodies (day 4) generated from the iPSC clone (#D3). **FIG. 10D** depicts fluorescence microscopic images of the iPSC clone (#D3) after *in vitro* differentiation from the embryoid body shown in FIG. 10C, and stained with antibodies against α SMA (mesoderm), AFP (endoderm), GATA4 (endoderm), and βIII-tubulin (ectoderm). Pictures (right) are the same images after merging with DAPI-staining. Scale bar, 100 μm.

[0018] **FIG. 11** shows human ZSCAN4 functions in the same manner as mouse Zscan4. By contrast to mouse Zscan4 that consists of 6 paralogous genes and 3 pseudogenes (Falco *et al.*, *Dev Biol* 307:539-550, 2007), the human genome has only one ZSCAN4 gene copy. To investigate if human ZSCAN4 has similar functions to mouse Zscan4, the efficiency of iPSC formation was compared between mouse Zscan4 and human ZSCAN4. **FIG. 11A** depicts a schematic representation of a piggyBac vector (PB-TET-hZSCAN4) carrying human ZSCAN4 open reading frame (ORF) under the Dox-inducible promoter. A PB-TET-MKOS vector together with either PB-TET-hZSCAN4 or a control PB-DsRed vector were cotransfected to MEF-WT cells (C57BL/6J x 129S6/SvEvTac). The experimental design was essentially the same as that shown in FIG. 8A. The cells were cultured for 14 days, fixed, and stained for ALP. **FIG. 11B** depicts representative pictures of 6-well plates stained for ALP. **FIG. 11C** depicts additional representative pictures of 6-well plates stained for ALP. ALP+ colonies were scored. Data are represented as mean $\pm$S.E.M. in triplicate; *, $P < 0.01$. (B) and (C) are representatives of three independent experiments.

[0019] **FIG. 12** shows that Zscan4 enhances iPSC colony formation without Myc. The charts shown represent two of the three replications for the experiments shown in FIG. 1B. The third data set is shown in FIG. 1C. A piggyBac vector (PB-TET-KOS) carrying Klf4 (K), Oct4 (O), and Sox2 (S) was transfected into the MEF-ZERT. The cells were cultured for 20 days under Dox+ Tmx- or Dox+ Tmx+ condition, fixed, and stained for ALP. ALP+ colonies were scored and presented in the bar chart. For each experiment, the transfection and Dox-induction were performed in triplicate. Data are represented as mean±S.E.M. (triplicate wells); *, $P < 0.01$.

[0020] **FIG. 13** depicts the characterization of iPSC clones derived from the MEF-ZERT with KOS factors. **FIG. 13A** depicts several iPSC colonies that were picked from the wells prepared in parallel with the experiment shown in FIG. 1C and propagated in the ES cell culture condition on feeder cells: four iPSC clones from the MEF-ZERT (KOS factors and Tmx+ condition, *i.e.*, ZKOS factors: #2, #4, #7, #8). These iPSC clones and control MEF-ZERT cells were subjected to RT-PCR analysis with pluripotency gene markers: endogenous Oct4 (Pou5f1), endogenous Sox2, Nanog, Zfp42 (Rex1), and Dax1 (Nr0b1). Gapdh was used as a control. **FIG. 13B** depicts a representative phase-contrast image of the iPSC clone (#7) after staining with ALP. **FIG. 13C** depicts an E13.5 embryo derived from the iPSC clone (#7) by the 4N complementation.

[0021] **FIG. 14** depicts the generation and characterization of iPSC clones derived from the MEF-WT with the KOS factors and Zscan4ERT2 under the Tmx+ condition. **FIG. 14A** shows that following the experimental procedure shown in FIG. 2A, iPSC colonies were generated by transfecting the MEF-WT (C57BL/6Jx129S6/SvEvTac) with piggyBac vectors (PB-TET-KOS and PB-TET-Zscan4ERT2-IRES-HisDsRed) and culturing the cells for 30 days under the Dox+ Tmx+ condition. Under fluorescence microscope, Zscan4ERT2^{+} iPS colonies could be identified by red-fluorescence. The two red iPSC colonies were picked from the wells and propagated in the ES cell culture condition on feeder cells, resulting in the establishment of two iPSC clones (MEF-WT with the ZKOS factors: #2, #4). These iPSC clones and control MEF-WT cells were subjected to RT-PCR analysis with pluripotency gene markers: endogenous Oct4 (Pou5f1), endogenous Sox2, Nanog, Zfp42 (Rex1), and Dax1 (Nr0b1). Gapdh was used as a control. **FIG. 14B** depicts a representative phase-contrast image of the iPSC clone (#2). **FIG. 14C** depicts a representative phase-contrast image of the iPSC clone (#2) after staining with ALP. **FIG. 14D** depicts fluorescence microscopic images of the iPSC clone (#2) after staining with antibodies against SSEA-1 and NANOG. Pictures (right) are the same images after merging with DAPI-staining. **FIG. 14E** depicts a microscopic image showing embryoid bodies (day 4) generated from the iPSC clone (#2). **FIG. 14F** depicts fluorescence images of the iPSC clone (#2) after *in vitro* differentiation from the embryoid body shown in FIG. 14E, and stained with antibodies against α SMA (mesoderm), AFP (endoderm), GATA4 (endoderm), and βIII-tubulin (ectoderm). Pictures (right) are the same images after merging with DAPI-staining. Scale bar, 100 μm. **FIG. 14G** depicts E13.5 embryos derived from the iPSC clone (#2) by the 4N complementation. These embryos were used to generate the secondary MEFs (MEF-KOS-ZERT^{2nd}) as described in FIG. 2A.

[0022] **FIG. 15** shows a comparison of global gene expression profiles between iPSC, ESC and MEF. Global expression profiles of the iPSC clone (ZKOS#2), V6.5 ESC, and MEF-WT were generated by using DNA microarrays. **FIG. 15A** depicts a scatter-plot showing pair-wise comparison between iPSC (ZKOS#2) and MEF-WT. **FIG. 15B** depicts a scatter-plot showing pair-wide comparison between iPSC (ZKOS#2) and V6.5 ESC. Spots in color represent genes whose expression show statistically significant difference between samples (FDR=0.05, fold-change>2).

[0023] **FIG. 16** depicts expression levels of pluripotency genes – Nanog, Zfp42, and Dppa5a during initial phase of induction in the secondary MEFs. Gene expression levels were obtained from DNA microarray analysis and plotted by the NIA ANOVA tool (Sharov *et al.*, *Bioinformatics* 21:2548-2549, 2005).

[0024] **FIG. 17** depicts a list of primers used in the studies disclosed herein (SEQ ID NOs: 15-37).

[0025] **FIG. 18** shows that iPSCs generated with Zscan4 are of high quality based on the karyotype and tetraploid complementation assay. **FIG. 18A** depicts a table of the results of a tetraploid complementation assay. **FIG. 18B** depicts images of karyotypes. **FIG. 18C** depicts images of live embryos derived from iPSC. Karyotype analysis of randomly selected iPSC lines clearly showed that iPSCs generated with Zscan4 were of higher quality than iPSCs generated without Zscan4. In addition, iPSCs generated with Zscan4 could form entire live embryos by the tetraploid complementation assay, which is the most stringent test for the pluripotency of iPS cells.

[0026] **FIG. 19** shows that Zscan4 is not expressed during early phase of iPSC formation, but reactivated later in iPSC cells. **FIG. 19A** depicts a schematic representation of procedures to examine Zscan4 expression during iPSC formation. TA1 ES cells, F1 hybrid strain (C57BL/6J x 129S6/SvEvTac). A piggyBac transfection involves a main vector PB-TET-MKOS (shown), PB-CAG-rtTA (a tetracycline transactivator), and pCyL43 (transposase). **FIG. 19B** depicts phase-contrast microscopic images during the formation of cell colonies with authentic ES-like morphology (denoted MOR+). Day 0 is set when doxycycline (Dox) is added to the complete ES medium 24 hours after a piggyBac transfection. **FIG. 19C** depicts fluorescence images (left), fluorescence images merged with phase-contrast images (middle), and flow cytometry charts (right) of two representative cell clones established from MOR+ colonies and cultured in the

absence of Dox. **FIG. 19D** depicts the appearance of EM+ cells (represented as "+") in the culture. Fraction of EM+ cells was measured by the flow cytometry on day 28.

[0027] **FIG. 20** depicts the validation of a MOR+ ALP+ phenotype for scoring authentic iPSC colonies generated by a piggyBac vector system. **FIG. 20A** depicts nanog-immunohistochemistry on ESCs and MEFs. Nanog-immunohistochemistry with a DAB (3,3'-diaminobenzidine) reporter showed staining patterns comparable to Nanog-immunohistochemistry with an Alexa-fluorescence reporter: absence of Nanog in MEFs and presence of Nanog in ESCs. When Nanog is not used as one of the exogenous iPSC factors, the activation of Nanog expression has been used as an indication of reprogramming to authentic iPSCs. **FIG. 20B** depicts a comparison of scoring methods for reprogrammed cells by MKOS factors between MOR+ ALP+ phenotype and NANOG+ phenotype. Similar number of colonies was obtained by both methods, indicating that MOR+ ALP+ can be used to score iPSC colonies reprogrammed by MKOS factors on a piggyBac vector. MEFs (B6DBA1F1) plated on gelatin-coated 6-well plates at a density of $1x10^5$ cells/well were used. **FIG. 20C** depicts additional support for the generation of authentic iPSC colonies reprogrammed by the MKOS factors on a piggyBac vector from MEFs carrying a GFP reporter driven by the Oct4 promoter (Stemgent, USA). Images of three representative MOR+ colonies are shown. **FIG. 20D** depicts a comparison of scoring methods for reprogrammed cells by the ZMKOS factors between MOR+ ALP+ phenotype and NANOG+ phenotype. Similar number of colonies was obtained by both methods, indicating that MOR+ ALP+ can be used to score iPSC colonies reprogrammed by ZMKOS factors on piggyBac vectors. MEFs (B6DBA1F1) plated on gelatin-coated 6-well plates at a density of $1x10^5$ cells/well were used. **FIG. 20E** depicts additional support for the generation of authentic iPSC colonies reprogrammed by ZMKOS factors on piggyBac vectors from MEFs carrying a GFP reporter driven by the Oct4 promoter (Stemgent, USA). Images of three representative MOR+ colonies are shown.

[0028] **FIG. 21** depicts a schematic of the experimental procedure and constructs used for testing the effect of several Zscan4-dependent genes on the efficiency of iPSC formation.

[0029] **FIG. 22** shows the efficiency of iPS colony formation by Zscan4 and Zscan4-dependent genes. **FIG. 22A** depicts a bar graph showing increased iPS colony formation from mouse embryonic fibroblast (MEF) by Zscan4c, Patl2 and Pramel6. **FIG. 22B** depicts a bar graph showing increased iPS colony formation from MEF by Piwil2. **FIG. 22C** depicts a bar graph showing increased iPS colony formation from MEF by Zscan4c, Patl2 and D5Ertd577e. The efficiency of iPS colony formation was scored based on their authentic ES cell morphology

and alkaline phosphatase (ALP) staining 14 days after transfection (mean ± S.E.M.). The experiments were performed in triplicate. Different letters (a or b) denote significant differences between groups ($P < 0.05$).

SEQUENCE LISTING

[0030] The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids, as defined in 37 C.F.R. 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. The Sequence Listing is submitted as an ASCII text file, created on August 26, 2011, 212 KB. In the accompanying sequence listing:

[0031] **SEQ ID NOs: 1 and 2** are nucleotide and amino acid sequences of human ZSCAN4.

[0032] **SEQ ID NOs: 3 and 4** are nucleotide and amino acid sequences of mouse Zscan4a.

[0033] **SEQ ID NOs: 5 and 6** are nucleotide and amino acid sequences of mouse Zscan4b.

[0034] **SEQ ID NOs: 7 and 8** are nucleotide and amino acid sequences of mouse Zscan4c.

[0035] **SEQ ID NOs: 9 and 10** are nucleotide and amino acid sequences of mouse Zscan4d.

[0036] **SEQ ID NOs: 11 and 12** are nucleotide and amino acid sequences of mouse Zscan4e.

[0037] **SEQ ID NOs: 13 and 14** are nucleotide and amino acid sequences of mouse Zscan4f.

[0038] **SEQ ID NOs: 15-37** are primer sequences.

[0039] **SEQ ID NOs: 38 and 39** are the nucleotide and amino acid sequences of mouse Patl2 (GenBank Accession No. NM_026251).

[0040] **SEQ ID NOs: 40 and 41** are the nucleotide and amino acid sequences of mouse Pramel6 (GenBank Accession No. NM_178249).

[0041] **SEQ ID NOs: 42 and 43** are the nucleotide and amino acid sequences of mouse Piwil2 (GenBank Accession No. NM_021308).

[0042] **SEQ ID NOs: 44 and 45** are the nucleotide and amino acid sequences of mouse D5Ertd577e (GenBank Accession No. NM_177187).

[0043] **SEQ ID NOs: 46 and 47** are the nucleotide and amino acid sequences of human Patl2 respectively (GenBank Accession No. NM_001145112).

[0044] **SEQ ID NOs: 48 and 49** are the nucleotide and amino acid sequences of human Piwil2 (GenBank Accession No. NM_001135721).

[0045] **SEQ ID NO: 50** is the nucleotide sequence of human PRAMEF1 (GenBank Accession No. NM_023013).

[0046] **SEQ ID NO: 51** is the nucleotide sequence of human PRAMF12 (GenBank Accession No. NM_001080830).

[0047] **SEQ ID NO: 52** is the nucleotide sequence of human PRAMEF2 (GenBank Accession No. NM_023014).

[0048] **SEQ ID NO: 53** is the nucleotide sequence of human PRAMEF15 (GenBank Accession No. NM_001098376).

[0049] **SEQ ID NO: 54** is the nucleotide sequence of human PRAMEF8 (GenBank Accession No. NM_001012276).

[0050] **SEQ ID NO: 55** is the nucleotide sequence of human PRAMEF10 (GenBank Accession No. NM_001039361).

[0051] **SEQ ID NO: 56** is the nucleotide sequence of human PRAMEF20 (GenBank Accession No. NM_001099852).

[0052] **SEQ ID NO: 57** is the nucleotide sequence of human PRAMEF17 (GenBank Accession No. NM_001099851).

[0053] **SEQ ID NO: 58** is the nucleotide sequence of human PRAMEF19 (GenBank Accession No. NM_001099790).

[0054] **SEQ ID NO: 59** is the nucleotide sequence of human PRAMEF14 (GenBank Accession No. NM_001099854).

[0055] **SEQ ID NO: 60** is the nucleotide sequence of human PRAMEF21 (GenBank Accession No. NM_001100114).

[0056] **SEQ ID NO: 61** is the nucleotide sequence of human PRAMEF16 (GenBank Accession No. NM_001045480).

[0057] **SEQ ID NO: 62** is the nucleotide sequence of human PRAMEF18 (GenBank Accession No. NM_001099850).

[0058] **SEQ ID NO: 63** is the nucleotide sequence of human PRAMEF13 (GenBank Accession No. NM_001024661).

[0059] **SEQ ID NO: 64** is the nucleotide sequence of human PRAMEF9 (GenBank Accession No. NM_001010890).

[0060] **SEQ ID NO: 65** is the nucleotide sequence of human PRAMEF5 (GenBank Accession No. NM_001013407).

[0061] **SEQ ID NO: 66** is the nucleotide sequence of human PRAMEF3 (GenBank Accession No. NM_001013692).

[0062] **SEQ ID NO: 67** is the nucleotide sequence of human PRAMEF22 (GenBank Accession No. NM_001100631).

[0063] **SEQ ID NO: 68** is the nucleotide sequence of human PRAMEF7 (GenBank Accession No. NM_001012277).

[0064] **SEQ ID NO: 69** is the nucleotide sequence of human PRAMEF11 (GenBank Accession No. NM_001146344).

[0065] **SEQ ID NO: 70** is the nucleotide sequence of human PRAME (Ensembl: WI2-2994D6.2; ENSG00000229571).

[0066] **SEQ ID NO: 71** is the nucleotide sequence of human PRAMEF6 (GenBank Accession No. NM_001010889).

[0067] **SEQ ID NO: 72** is the nucleotide sequence of human PRAMEF4 (GenBank Accession No. NM_001009611).

[0068] **SEQ ID NO: 73** is the nucleotide sequence of the Zscan4-Emerald expression vector (9396 bp). The starting nucleotide of the Zscan4c promoter sequence is 906 and the ending nucleotide is 4468.

DETAILED DESCRIPTION

I. Abbreviations

ALP	alkaline phosphatase
D5Ertd577e	DNA segment, Chr 5, ERATO Doi 577, expressed
Dox	doxycycline
ES	embryonic
hCG	human chorionic gonadotropin
iPS	induced pluripotent stem
iPSC	induced pluripotent stem cell
IRES	internal ribosomal entry site
KOS	Klf4, Oct4, Sox2
LIF	leukemia inhibitory factor
MEF	murine embryonic fibroblast
MKOS	Myc, Klf4, Oct4, Sox2
NT	nuclear transplantation
ORF	open reading frame
Patl2	protein associated with topoisomerase II homolog 2
PCR	polymerase chain reaction
Piwil2	piwi-like homolog 2
PMSG	pregnant mare serum gonadotropin
PRAME	preferentially expressed antigen in melanoma
Pramel6	preferentially expressed antigen in melanoma like 6
qPCR	quantitative PCR
RT-PCR	reverse transcriptase PCR
Tmx	Tamoxifen
WT	wild type
ZKOS	Zscan4, Klf4, Oct4, Sox2
Zscan4	zinc finger and scan domain-containing protein 4

II. Terms and Methods

[0069] Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

[0070] In order to facilitate review of the various embodiments of the disclosure, the following explanations of specific terms are provided:

[0071] **Agent**: Any protein, nucleic acid molecule, compound, small molecule, organic compound, inorganic compound, or other molecule of interest. In some embodiments, the "agent" is any agent that increases expression of Zscan4. In particular examples, the agent is a retinoid or an agent that induces oxidative stress.

[0072] **c-Myc:** A transcription factor that plays a role in cell cycle progression, apoptosis and cellular transformation, and has been previously reported to function as a reprogramming factor. Nucleic acid and protein sequences of c-Myc for a number of different species are known in the art. For example, mouse c-Myc nucleic acid and protein sequences include GenBank accession numbers NM_010849 and NP_034979, respectively; and human c-Myc nucleic acid and protein sequences include GenBank accession numbers NM_002467 and NP_002458, respectively. c-Myc is also known as v-myc myelocytomatosis viral oncogene homolog, MYC and myelocytomatosis oncogene.

[0073] **Cell-penetrating peptide (CPP):** A type of polypeptide that facilitates transduction of proteins, nucleic acids or other compounds across membranes in a receptor-independent manner (Wadia and Dowdy, *Curr Protein Pept Sci 4(2)*:97-104, 2003). Typically, CPPs are short polycationic sequences that can facilitate cellular uptake of compounds to which they are linked into endosomes of cells. Examples of CPPs include poly-arginine tags and protein transduction domains (such as HIV-1 Tat).

[0074] **Contacting:** Placement in direct physical association; includes both in solid and liquid form. As used herein, "contacting" is used interchangeably with "exposed." In some cases, "contacting" includes transfecting, such as transfecting a nucleic acid molecule into a cell.

[0075] **Degenerate variant:** A polynucleotide encoding a polypeptide, such as a Zscan4 polypeptide, that includes a sequence that is degenerate as a result of the genetic code. There are 20 natural amino acids, most of which are specified by more than one codon. Therefore, all degenerate nucleotide sequences are included as long as the amino acid sequence of the polypeptide encoded by the nucleotide sequence is unchanged.

[0076] **Differentiation:** Refers to the process by which a cell develops into a specific type of cell (for example, muscle cell, skin cell etc.). Differentiation of pluripotent stem cells refers to the development of the cells toward a specific cell lineage. As a cell becomes more differentiated, the cell loses potency, or the ability to become multiple different cell types.

[0077] **DNA segment, Chr 5, ERATO Doi 577, expressed (D5Ertd577e):** A gene located on chromosome 5 of the mouse genome. Nucleotide and amino acid sequences of D5Ertd577e are available, such as in the NCBI database under Gene ID 320549. Exemplary mRNA and protein sequences are provided herein as SEQ ID NOs: 44 and 45, respectively (GenBank Accession No. NM_177187). Predicted human orthologs of D5Ertd577e include members of the preferentially expressed antigen in melanoma family (PRAMEF), for example PRAMEF1, PRAMF12, PRAMEF2, PRAMEF15, PRAMEF8, PRAMEF10, PRAMEF20, PRAMEF17, PRAMEF19, PRAMEF14, PRAMEF21, PRAMEF16, PRAMEF18, PRAMEF13, PRAMEF9, PRAMEF5, PRAMEF3, PRAMEF22, PRAMEF7, PRAMEF11, WI2-2994D6.2, PRAMEF6 and PRAMEF4 (nucleotide sequences for each of the orthologs is set forth herein as SEQ ID NOs: 50-72).

[0078] **Encapsulated:** As used herein, a molecule "encapsulated" in a nanoparticle refers to a molecule (such as a Zscan4 nucleic acid or protein) that is either contained within the nanoparticle or attached to the surface of the nanoparticle, or a combination thereof.

[0079] **ERT2:** A protein comprising a mutated ligand binding domain of the human estrogen receptor that does not bind its natural ligand (17β-estradiol) at physiological concentrations, but is highly sensitive to nanomolar concentrations of tamoxifen or its metabolite 4-hydroxy-tamoxifen (4OHT) (Feil *et al.*, *Biochem Biophys Res Commun* 237(3):752-757, 1997).

[0080] **Fusion protein:** A protein containing at least a portion of two different (heterologous) proteins. In some examples such proteins are generated by expression of a nucleic acid sequence engineered from nucleic acid sequences encoding at least a portion of two different (heterologous) proteins. To create a fusion protein, the nucleic acid sequences must be in the same reading frame and contain no internal stop codons.

[0081] **Heterologous:** A heterologous polypeptide or polynucleotide refers to a polypeptide or polynucleotide derived from a different source or species.

[0082] **Host cells:** Cells in which a vector can be propagated and its DNA expressed. The term also includes any progeny of the subject host cell. It is understood that all progeny may not be identical to the parental cell since there may be mutations that occur during replication. However, such progeny are included when the term "host cell" is used.

[0083] **Introducing:** In the context of the present disclosure, "introducing" a nucleic acid molecule or a protein into a cell encompasses any means of delivering the nucleic acid molecule or protein into the cell. For example, nucleic acid molecules can be transfected, transduced or electroporated into a cell. Delivery of proteins into cells can be achieved, for example, by fusing the protein to a cell-penetrating peptide, such as a peptide with a protein transduction domain (such as HIV-1 Tat), or a poly-arginine peptide tag (Fuchs and Raines, *Protein Science* 14:1538-1544, 2005).

[0084] **Induced pluripotent stem (iPS) cells**: A type of pluripotent stem cell artificially derived from a non-pluripotent cell, such as an adult somatic cell, by inducing a "forced" expression of certain transcription factor genes (generally referred to in the art as "reprogramming factors," "nuclear reprogramming factors," or "somatic cell reprogramming factors"). iPS cells can be derived from any organism, such as a mammal. In some embodiments, iPS cells are produced from mice, rats, rabbits, guinea pigs, goats, pigs, cows, non-human primates or humans. Human and murine derived iPS cells are exemplary.

[0085] iPS cells are similar to ES cells in many respects, such as the expression of certain stem cell genes and proteins, chromatin methylation patterns, doubling time, embryoid body formation, teratoma formation, viable chimera formation, and potency and differentiability. Methods for producing iPS cells are known in the art (exemplary methods are discussed below in section V). For example, iPS cells are typically derived by delivery of certain stem cell-associated genes (such as Oct-3/4 (Pouf51) and Sox2) into non-pluripotent cells, such as adult fibroblasts. Delivery can be achieved through viral vectors, such as retroviruses, lentiviruses, or adenoviruses; transfection of plasmid vectors; or delivery of reprogramming factor mRNA or protein. For example, cells can be transfected with Oct3/4, Sox2, Klf4, and c-Myc using a retroviral system or with OCT4, SOX2, NANOG, and LIN28 using a lentiviral system. After 3-4 weeks, small numbers of transfected cells begin to become morphologically and biochemically similar to pluripotent stem cells, and are typically isolated through morphological selection,

doubling time, or through a reporter gene and antibiotic selection. Methods of producing iPS cells from adult human cells have been previously described (see for example, Yu *et al.*, *Science* 318(5854):1224, 2007; Takahashi *et al.*, Cell 131(5):861-72, 2007; U.S. Patent Application Publication Nos. 2008/0280362, 2009/0068742, 2009/0227032, 2009/0047263 and 20100279404).

[0086] **Isolated:** An isolated nucleic acid has been substantially separated or purified away from other nucleic acid sequences and from the cell of the organism in which the nucleic acid naturally occurs, *i.e.*, other chromosomal and extrachromosomal DNA and RNA. The term "isolated" thus encompasses nucleic acids purified by standard nucleic acid purification methods. The term also embraces nucleic acids prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids. Similarly, "isolated" proteins have been substantially separated or purified from other proteins of the cells of an organism in which the protein naturally occurs, and encompasses proteins prepared by recombination expression in a host cell as well as chemically synthesized proteins. Similarly, "isolated" cells have been substantially separated away from other cell types.

[0087] **Klf4:** A transcription factor previously reported to function as a reprogramming factor. Nucleic acid and protein sequences of Klf4 for a number of different species are known in the art. For example, mouse Klf4 nucleic acid and protein sequences include GenBank accession numbers NM_010637 and NP_034767, respectively; and human Klf4 nucleic acid and protein sequences include GenBank accession numbers NM_004235 and NP_004226, respectively. Klf4 is also known as kruppel-like factor 4.

[0088] **Lin28:** A transcription factor previously reported to function as a reprogramming factor. Nucleic acid and protein sequences of Lin28 for a number of different species are known in the art. For example, mouse Lin28 nucleic acid and protein sequences include GenBank accession numbers NM_145833 and NP_665832, respectively; and human Lin28 nucleic acid and protein sequences include GenBank accession numbers NM_024674 and NP_078950, respectively. Lin28 is also known as lin-28 homolog A (LIN28A) and Lin-28.

[0089] **Linker:** One or more nucleotides or amino acids that serve as a spacer between two molecules, such as between two nucleic acid molecules or two peptides (such as in a fusion protein). In some examples a linker is 1 to 100 amino acids, such as 1 to 50 or 5 to 10 amino acids.

[0090] **Nanog:** A transcription factor previously reported to function as a reprogramming factor. Nucleic acid and protein sequences of Nanog for a number of different species are known in the art. For example, mouse Nanog nucleic acid and protein sequences include GenBank accession numbers NM_028016 and NP_082292, respectively; and human Nanog nucleic acid and protein sequences include GenBank accession numbers NM_024865 and NP_079141, respectively. Nanog is also known as Nanog homeobox.

[0091] **Nanoparticle:** A particle less than about 1000 nanometers (nm) in diameter. Exemplary nanoparticles for use with the methods provided herein are made of biocompatible and biodegradable polymeric materials. In some embodiments, the nanoparticles are PLGA nanoparticles. As used herein, a "polymeric nanoparticle" is a nanoparticle made up of repeating subunits of a particular substance or substances. "Poly(lactic acid) nanoparticles" are nanoparticles having repeated lactic acid subunits. Similarly, "poly(glycolic acid) nanoparticles" are nanoparticles having repeated glycolic acid subunits.

[0092] **Non-human animal:** Includes all animals other than humans. A non-human animal includes, but is not limited to, a non-human primate, a farm animal such as swine, cattle, and poultry, a sport animal or pet such as dogs, cats, horses, hamsters, rodents, such as mice, or a zoo animal such as lions, tigers or bears. In one example, the non-human animal is a mouse.

[0093] **Oct4:** A transcription factor that plays a role in embryonic development, especially during early embryogenesis. Oct4 is necessary for embryonic stem cell potency and has been previously reported to function as a reprogramming factor. Nucleic acid and protein sequences of Oct4 for a number of different species are known in the art. For example, mouse Oct4 nucleic acid and protein sequences include GenBank accession numbers NM_013633 and NP_038661, respectively; and human Oct nucleic acid and protein sequences include GenBank accession numbers NM_002701 and NP_002692, respectively. Oct4 is also known as POU domain class 5 transcription factor 1 (Pou5f1), Oct3 and Oct3/4.

[0094] **Operably linked:** A first nucleic acid sequence is operably linked to a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked nucleic acid sequences are contiguous and where necessary to join two protein coding regions, in the same reading frame.

[0095] **Piwi-like homolog 2 (Piwil2):** A gene belonging to the Argonaute family of proteins, which function in development and maintenance of germline stem cells. Nucleotide and amino acid sequences of Piwil2 are available, such as in the NCBI database under Gene ID 57746 (mouse) and Gene ID 55124 (human). Exemplary mouse mRNA and protein sequences are provided herein as SEQ ID NOs: 42 and 43, respectively (GenBank Accession No. NM_021308). Exemplary human mRNA and protein sequences are provided herein as SEQ ID NOs: 48 and 49, respectively (GenBank Accession No. NM_001135721).

[0096] **Pluripotent/pluripotency:** A "pluripotent" cell is a cell that can form all of an organism's cell lineages (endoderm, mesoderm and ectoderm), including germ cells. Pluripotent cells can give rise to any fetal or adult cell type, but cannot form an entire organism autonomously due to the inability to form extraembryonic tissue (such as placenta).

[0097] **Poly-arginine peptide tag:** A short peptide (generally 7 to 11 residues) comprised of arginine residues that facilitates delivery of larger molecules (such as proteins and nucleic acid molecules) into cells (see, for example, Fuchs and Raines, *Protein Science* 14:1538-1544, 2005).

[0098] **Polypeptide:** A polymer in which the monomers are amino acid residues which are joined together through amide bonds. When the amino acids are alpha-amino acids, either the L-optical isomer or the D-optical isomer can be used, the L-isomers being preferred. The terms "polypeptide" or "protein" as used herein are intended to encompass any amino acid sequence and include modified sequences such as glycoproteins. The term "polypeptide" is specifically intended to cover naturally occurring proteins, as well as those which are recombinantly or synthetically produced.

[0099] The term "polypeptide fragment" refers to a portion of a polypeptide which exhibits at least one useful epitope. The term "functional fragments of a polypeptide" refers to all fragments of a polypeptide that retain an activity of the polypeptide, such as a Zscan4. Biologically functional fragments, for example, can vary in size from a polypeptide fragment as small as an epitope capable of binding an antibody molecule to a large polypeptide capable of participating in the characteristic induction or programming of phenotypic changes within a cell, including affecting cell proliferation or differentiation. An "epitope" is a region of a polypeptide capable of binding an immunoglobulin generated in response to contact with an antigen. Thus, smaller peptides containing the biological activity of Zscan4, or conservative variants of Zscan4, are thus included as being of use.

[0100] The term "substantially purified polypeptide" as used herein refers to a polypeptide which is substantially free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In one embodiment, the polypeptide is at least 50%, for example at least 80% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In another embodiment, the polypeptide is at least 90% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In yet another embodiment, the polypeptide is at least 95% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated.

[0101] Conservative substitutions replace one amino acid with another amino acid that is similar in size, hydrophobicity, etc. Examples of conservative substitutions are shown below:

Original Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln; Glu
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

[0102] Variations in the cDNA sequence that result in amino acid changes, whether conservative or not, should be minimized in order to preserve the functional and immunologic identity of the encoded protein. Thus, in several non-limiting examples, a Zscan4 polypeptide includes at most two, at most five, at most ten, at most twenty, or at most fifty conservative substitutions. The immunologic identity of the protein may be assessed by determining whether it is recognized by an antibody; a variant that is recognized by such an antibody is immunologically conserved. Any cDNA sequence variant will preferably introduce no more than twenty, and preferably fewer than ten amino acid substitutions into the encoded polypeptide.

[0103] **Preferentially expressed antigen in melanoma like 6 (Pramel6):** A gene located on chromosome 2 of the mouse genome. Nucleotide and amino acid sequences of Pramel6 are available, such as in the NCBI database under Gene ID 347711. Exemplary mRNA and protein sequences are provided herein as SEQ ID NOs: 40 and 41, respectively (GenBank Accession No. NM_178249). Predicted human orthologs of Pramel6 include members of the preferentially expressed antigen in melanoma family (PRAMEF), for example PRAMEF1, PRAMF12, PRAMEF2, PRAMEF15, PRAMEF8, PRAMEF10, PRAMEF20, PRAMEF17, PRAMEF19, PRAMEF14, PRAMEF21, PRAMEF16, PRAMEF18, PRAMEF13, PRAMEF9, PRAMEF5, PRAMEF3, PRAMEF22, PRAMEF7, PRAMEF11, WI2-2994D6.2, PRAMEF6 and PRAMEF4 (nucleotide sequences for each of the orthologs is set forth herein as SEQ ID NOs: 50-72).

[0104] **Progenitor cells:** Oligopotent or unipotent cells that differentiate into a specific type of cell or cell lineage. Progenitor cells are similar to stem cells but are more differentiated and exhibit limited self renewal.

[0105] **Promoter:** Nucleic acid control sequences which direct transcription of a nucleic acid. A promoter includes necessary nucleic acid sequences near the start site of transcription. A promoter also optionally includes distal enhancer or repressor elements. A "constitutive promoter" is a promoter that is continuously active and is not subject to regulation by external signals or molecules. In contrast, the activity of an "inducible promoter" is regulated by an external signal or molecule (for example, a transcription factor).

[0106] **Protein associated with topoisomerase II homolog 2 (Patl2):** A gene conserved in human, mouse, chimpanzee, dog, cow, chicken and zebrafish. Nucleotide and amino acid sequences of Patl2 are available, such as in the NCBI database under Gene ID 67578 (mouse) and Gene ID 197135 (human). Exemplary mouse mRNA and protein sequences are provided herein as SEQ ID NOs: 38 and 39, respectively (GenBank Accession No. NM_026251). Exemplary human mRNA and protein sequences are provided herein as SEQ ID NOs: 46 and 47, respectively (GenBank Accession No. NM_001145112).

[0107] **Protein transduction domains:** Small cationic peptides that facilitate entry of larger molecules (proteins, nucleic acid molecules etc.) into a cell by a mechanism that is independent of classical endocytosis.

[0108] **Recombinant:** A recombinant nucleic acid or polypeptide is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by

chemical synthesis or by the artificial manipulation of isolated segments of nucleic acids, for example, by genetic engineering techniques.

[0109] **Reprogramming:** In the context of the present disclosure, "reprogramming" a somatic cell refers to the process of converting a partially or fully differentiated somatic cell into a pluripotent cell (*i.e.*, an iPS cell). Because reprogrammed somatic cells possess ES cell-like properties, it is believed that iPS cells can replace ES cells in a number of regenerative medicine applications.

[0110] **Reprogramming factor:** A gene or gene product that when exogenously expressed or introduced into a somatic cell is capable of promoting a pluripotent state. A number of reprogramming factors have been described in the art including, but not limited to, c-Myc, Klf4, Oct4, Sox2, Lin28 and Nanog. Although c-Myc, Klf4, Oct4, Sox2, Lin28 and Nanog are currently the most commonly used reprogramming factors, the term also encompasses other genes and gene products with the same functional effect of promoting the pluripotent state. For example, in some cases, Sox1, Sox3, L-myc, N-myc, and Klf2 can be used as reprogramming factors. Reprogramming factors are also referred to in the art as "nuclear reprogramming factors," "somatic cell reprogramming factors" and "pluripotency genes."

[0111] **Retinoids:** A class of chemical compounds that are related chemically to vitamin A. Retinoids are used in medicine, primarily due to the way they regulate epithelial cell growth. Retinoids have many important and diverse functions throughout the body including roles in vision, regulation of cell proliferation and differentiation, growth of bone tissue, immune function, and activation of tumor suppressor genes. Examples of retinoids include, but are not limited to, all-trans retinoic acid (atRA), 9-cis retinoic acid (9-cis RA), 13-cis RA and vitamin A (retinol).

[0112] **Quality of an iPS cell:** As used herein, an iPS cell of high quality refers to an iPS cell having a normal karyotype and/or possessing the ability to form a live embryo (such as in a tetraploid complementation assay). High quality iPS cells are highly pluripotent.

[0113] **Sequence identity/similarity**: The identity/similarity between two or more nucleic acid sequences, or two or more amino acid sequences, is expressed in terms of the identity or similarity between the sequences. Sequence identity can be measured in terms of percentage identity; the higher the percentage, the more identical the sequences are. Sequence similarity can be measured in terms of percentage similarity (which takes into account conservative amino acid substitutions); the higher the percentage, the more similar the sequences are. Homologs or

orthologs of nucleic acid or amino acid sequences possess a relatively high degree of sequence identity/similarity when aligned using standard methods. This homology is more significant when the orthologous proteins or cDNAs are derived from species which are more closely related (such as human and mouse sequences), compared to species more distantly related (such as human and *C. elegans* sequences).

[0114] Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; Higgins & Sharp, *Gene*, 73:237-44, 1988; Higgins & Sharp, *CABIOS* 5:151-3, 1989; Corpet *et al.*, *Nuc. Acids Res.* 16:10881-90, 1988; Huang *et al.* *Computer Appls. in the Biosciences* 8, 155-65, 1992; and Pearson *et al.*, *Meth. Mol. Bio.* 24:307-31, 1994. Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990, presents a detailed consideration of sequence alignment methods and homology calculations.

[0115] The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990) is available from several sources, including the National Center for Biological Information (NCBI, National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. Additional information can be found at the NCBI web site.

[0116] **Somatic cell:** Any cell of the body. In standard terminology, "somatic cell" generally excludes germ cells. However, in the context of the present disclosure, the "somatic cell" that can be used in the disclosed methods to produce an iPS cell is any cell, including germ cells present during development and early embryonic cells. In some embodiments, the somatic cell is a tissue stem cell, progenitor cell or differentiated cell. Fibroblasts (including embryonic fibroblasts, adult fibroblasts and cardiac fibroblasts) have commonly been used for generating iPS cells. However, a number of other cell types have been described, including oral mucosa, cord blood cells, lymphocytes (*e.g.*, T cells and B cells), stromal cells, neural progenitor cells, adipose cells, keratinocytes, neural stem cells, meningiocytes, adipose stem cells, hepatocytes, gastric cells, pancreatic beta cells, peripheral blood cells, fetal hepatocytes, adipocytes and limbal cells (Parameswaran *et al.*, *Stem Cells* 29(7):1013-1020, 2011; U.S. Patent Application Publication No. 2010/0279404). In some embodiments, the somatic cell is a neural stem cell, a hematopoietic stem cell, a mesenchymal stem cell, an adipose stem cell, a fibroblast, a lymphocyte (such as a T cell or B cell), a hepatocyte, an epithelial cell, a muscle cell, an adipose

cell, a cardiomyocyte, a pancreatic β cell, a keratinocyte, an amniotic cell, a peripheral blood cell, a platelet, or an astrocyte.

[0117] **Sox2:** A transcription factor involved in the regulation of embryonic development and in the determination of cell fate. Sox 2 (sex determining region Y-box 2) has been previously reported to function as a reprogramming factor. Nucleic acid and protein sequences of Sox2 for a number of different species are known in the art. For example, mouse Sox2 nucleic acid and protein sequences include GenBank accession numbers NM_011443 and NP_035573, respectively; and human Sox2 nucleic acid and protein sequences include GenBank accession numbers NM_003106 and NP_003097, respectively.

[0118] **Subject:** Living multi-cellular vertebrate organisms, a category that includes human and non-human mammals.

[0119] **Tissue stem cell:** Undifferentiated cells found throughout the body after embryonic development that multiply by cell division to replenish dying cells and regenerate damaged tissues. Tissue stem cells are also known as somatic stem cells or adult stem cells.

[0120] **Transfecting or transfection:** Refers to the process of introducing nucleic acid into a cell or tissue. Transfection can be achieved by any one of a number of methods, such as, but not limited to, liposomal-mediated transfection, electroporation and injection.

[0121] **Vector:** A nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector may include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication (DNA sequences that participate in initiating DNA synthesis). For example, an expression vector contains the necessary regulatory sequences to allow transcription and translation of inserted gene or genes. A vector may also include one or more selectable marker genes and other genetic elements known in the art. Vectors include, for example, virus vectors and plasmid vectors.

[0122] **Zscan4:** A group of genes that have previously been identified as exhibiting 2-cell-specific expression and ES cell-specific expression (PCT Publication No. WO 2008/118957) and have been shown to promote telomere elongation and genome stability (Zalzman *et al.*, *Nature* 464(7290):858-863, 2010; PCT Publication No. WO 2011/028880). In the context of the present disclosure, "Zscan4" includes both human ZSCAN4 and mouse Zscan4. In the mouse, the term "Zscan4" refers to a collection of genes including three pseudogenes (Zscan4-ps1, Zscan4-ps2 and Zscan4-ps3) and six expressed genes (Zscan4a, Zscan4b, Zscan4c, Zscan4d, Zscan4e and

Zscan4f). Among the six paralogs, the open reading frames of Zscan4c, Zscan4d, and Zscan4f encode a SCAN domain as well as all four zinc finger domains, suggesting their potential role as transcription factors. Zscan4 refers to Zscan4 polypeptides and Zscan4 polynucleotides encoding the Zscan4 polypeptides. Exemplary Zscan4 sequences are set forth herein as SEQ ID NOs: 1-14, and are disclosed in PCT Publication Nos. WO 2008/118957 and WO 2011/028880.

[0123] **Zscan4-dependent gene:** A gene whose expression is regulated by Zscan4. In some embodiments of the present disclosure, a Zscan4-dependent gene refers to any of the 231 genes (listed in Table 1) that were identified as upregulated during the early phase (day 1 – day 6) of iPSC formation in a Zscan4-dependent manner (see Example 1 below). In particular examples, the Zscan4-dependent gene is Patl2, Pramel6, Piwil2 or D5Ertd577e. In non-limiting examples, the Zscan4-dependent gene is mouse Patl2, mouse Pramel6, mouse Piwil2 or mouse D5Ertd577e; or human PATL2 or human PIWIL2; or a human ortholog of mouse Pramel6 or mouse D5Ertd577e (such as a member of the PRAME family).

[0124] Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. Hence "comprising A or B" means including A, or B, or A and B. It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

III. Introduction

[0125] Reprogramming of somatic cells by nuclear transplantation (NT) occurs within a few cell divisions (Egli *et al.*, *Curr Biol* 19:1403-1409, 2009), whereas reprogramming of somatic cells into induced pluripotent stem (iPS) cells by the forced expression of four factors, Myc, Klf4, Oct4, and Sox2 (MKOS) requires about 3 weeks (Takahashi and Yamanaka, *Cell* 126:663-676, 2006; Hanna *et al.*, *Cell* 143:508-525, 2010; Stadtfeld and Hochedlinger, *Genes Dev* 24:2239-2263, 2010; Gonzalez *et al.*, *Nat Rev Genet* 12:231-242, 2011), with their continuous

application for at least the first 8-10 days (Wernig *et al.*, *Nat Biotechnol* 26:916-924, 2008; Plath and Lowry, *Nat Rev Genet* 12:253-265, 2011). Although a requirement for additional embryonic factors present in the NT environment has been suggested for the efficient formation of iPS cells (Hanna *et al.*, *Nature* 462:595-601, 2009), no such factors have been reported yet. Previously, it was shown that Zscan4, expressed specifically in 2-cell embryos and only about 5% of ES cells at a given time (Falco *et al.*, *Dev Biol* 307:539-550, 2007), acts critically in the formation of proper blastocysts (Falco *et al.*, *Dev Biol* 307:539-550, 2007) and in the maintenance of genome stability and telomeres in ES cells (Zalzman *et al.*, *Nature* 464:858-863, 2010). It is disclosed herein that Zscan4 can replace oncogenic Myc and produce a comparable number of iPSC colonies from mouse embryo fibroblasts (MEFs) when coexpressed with Klf4, Oct4, and Sox2.

[0126] Furthermore, unlike other factors, Zscan4 was required only for the initial day during the formation of iPSCs. Global expression profiling of iPSC-derived, homogeneously-inducible secondary MEFs revealed that a small difference in the transcriptome caused by the overexpression of Zscan4 determines whether MEFs become iPS cells. Though few in number, many of these upregulated genes exhibit preimplantation embryo-specific expression. Without Klf4, Oct4, and Sox2, the overexpression of Zscan4 did not itself convert MEFs to iPSCs, but it reduced the proliferation of MEFs and caused partial but rapid DNA demethylation on the promoters of Oct4. It is further disclosed herein that at least four of the identified Zscan4-dependent genes also are capable of promoting iPSC formation when co-expressed with MKOS factors.

[0127] Taken together, the studies disclosed herein have established that Zscan4 is a previously unidentified transiently acting natural factor that facilitates the reprogramming process. The Zscan4-mediated reprogramming process recapitulates at least a part of NT-mediated reprogramming process through the reactivation of the early embryonic program.

IV. Overview of Several Embodiments

[0128] Disclosed herein is the finding that Zscan4 initiates direct reprogramming of somatic cells by reactivating early embryonic genes. In particular, it is disclosed that forced expression of Zscan4 in somatic cells, along with other previously described reprogramming factors, leads to the efficient production of high quality iPS cells. It is further disclosed herein that forced expression of any of the Zscan4-dependent genes Patl2, Pramel6, Piwil2 and D5Ertd577e, in combination with previously described reprogramming factors, promotes formation of iPS colonies.

[0129] Provided herein is an *ex vivo* method of producing an iPS cell by reprogramming of a somatic cell. The method includes contacting the somatic cell with a Zscan4 or a Zscan4-dependent gene and at least one reprogramming factor, thereby producing an iPS cell. In some embodiments, the Zscan4-dependent gene is selected from the genes listed in Table 1. In specific non-limiting embodiments, the Zscan4dependent gene is selected from Patl2, Pramel6, Piwil2 and D5Ertd577e. In some embodiments, the method includes contacting the somatic cell with at least two, at least three, or at least four reprogramming factors. Reprogramming factors (also known in the art as somatic cell reprogramming factors, nuclear reprogramming factors, pluripotency genes or stem cell-associated genes) have been described in the art and appropriate reprogramming factors can be selected by one of skill. In particular examples, the at least one, at least two, at least three or at least four reprogramming factors are selected from c-Myc, Klf4, Oct4, Sox2, Lin28 and Nanog. In some cases, the reprogramming factors include one or more of Sox 1, Sox3, L-myc, N-myc or Klf2. For example, Sox1 and Sox3 may be used to replace Sox2; L-myc or N-myc may be used to replace c-Myc; and/or Klf2 may be used to replace Klf4.

[0130] In several non-limiting examples, the somatic cell is contacted with two, three or four reprogramming factors, wherein (i) the four reprogramming factors are c-Myc, Klf4, Oct4 and Sox2; (ii) the four reprogramming factors are Lin28, Nanog, Oct4 and Sox2; (iii) the three reprogramming factors are Klf4, Oct4 and Sox2; or (iv) the two reprogramming factors are Oct4 and Sox2.

[0131] Zscan4, the Zscan4-dependent gene(s) and the reprogramming factor(s) can be introduced to the somatic cell using any suitable method known in the art that results in delivery of the Zscan4 or Zscan4-dependent gene(s) and the reprogramming factor(s) into the cell. The method of delivery of the Zscan4 and/or Zscan4-dependent gene(s) need not be the same method used for delivery of the reprogramming factor(s). In some embodiments, contacting the somatic cell with a Zscan4 comprises introducing a nucleic acid molecule encoding a Zscan4 protein into the somatic cell. In some embodiments, contacting the somatic cell with a Zscan4-dependent gene comprises introducing a nucleic acid molecule encoding a Zscan4-dependent gene protein into the somatic cell. In some embodiments, contacting the somatic cell with at least one reprogramming factor comprises introducing a nucleic acid molecule encoding at least one reprogramming factor protein into the somatic cell.

[0132] For embodiments in which at least two reprogramming factors are contacted with the somatic cell by introducing a nucleic acid molecule encoding the reprogramming factors, the reprogramming factors can be delivered using a single nucleic acid molecule (such as a single

viral vector or plasmid containing the nucleic acid molecule) or as separate nucleic acid molecules (such as a separate vector or plasmid for each reprogramming factor). Similarly, the Zscan4 nucleic acid molecule, or Zscan4-dependent gene nucleic acid molecule, can be delivered to the somatic cell as a separate nucleic acid molecule or can be included with the nucleic acid molecule encoding the reprogramming factor(s).

[0133] In some embodiments, the nucleic acid molecule encoding the Zscan4 or Zscan4-dependent gene, and/or the nucleic acid molecule encoding the at least one reprogramming factor, comprises a viral vector. Exemplary viral vectors include, but are not limited to retrovirus vectors, lentivirus vectors and adenovirus vectors. However, the choice of viral vector may vary depending upon, for example, the type of somatic cell to be used and the particular application for which the iPS cell will be used. One of skill in the art is capable of selecting an appropriate viral vector for introduction of the Zscan4 or the Zscan4-dependent gene, and reprogramming factor(s).

[0134] In other embodiments, the nucleic acid molecule comprises a plasmid vector. In specific examples, the plasmid vector is an episomal plasmid vector capable of autonomous replication.

[0135] In some examples, the nucleic acid molecule introduced to the somatic cell is encapsulated in a nanoparticle.

[0136] In other embodiments, the nucleic acid molecule encoding the Zscan4, the Zscan4-dependent gene, or the at least one reprogramming factor comprises mRNA encoding the Zscan4 protein, the Zscan4-dependent gene protein or the reprogramming factor protein.

[0137] In some embodiments, contacting the somatic cell with a Zscan4 comprises introducing a Zscan4 protein into the somatic cell. In some embodiments, contacting the somatic cell with a Zscan4-dependent gene comprises introducing a Zscan4dependent gene protein into the somatic cell. In some embodiments, contacting the somatic cell with at least one reprogramming factor comprises introducing a reprogramming factor protein into the somatic cell. In some examples, the Zscan4 protein, Zscan4-dependent gene protein or reprogramming factor protein is encapsulated in a nanoparticle. In other examples, the Zscan4 protein, Zscan4dependent gene protein or reprogramming factor protein is fused to a cell-penetrating peptide (CPP). A number of different CPPs are known in the art, and are discussed in greater detail below. In particular examples, the cell-penetrating peptide comprises a protein transduction domain, such as the human immunodeficiency virus (HIV) Tat protein. In other

examples, the cell-penetrating peptide comprises a poly-arginine peptide tag. The poly-arginine tag can vary in size, but in some cases is about 7 to about 11 arginine residues.

[0138] Delivery of the Zscan4 or Zscan4-dependent gene and the reprogramming factor(s) (regardless of whether delivery includes delivery of a nucleic acid molecule, vector, mRNA or protein) can occur simultaneously or sequentially. In some embodiments, the Zscan4 or Zscan4-dependent gene is delivered prior to delivery of the at least one reprogramming factor. In addition, in some cases, depending on the method of delivery, the Zscan4 or Zscan4-dependent gene and/or the reprogramming factor(s) is delivered at least twice, at least three times or at least four times to allow for a sufficient duration of expression to permit induction of the pluripotent state. In some embodiments, the methods disclosed herein include continuous expression of the reprogramming factor(s) (or presence of the reprogramming factor(s) protein) for at least 3 days, at least 4 days, at least 5 days, at least 6 days, at least 7 days, at least 8 days, at least 9 days, or at least 10 days. In some embodiments, the method includes continuous expression of Zscan4 (or presence of the Zscan4 protein) for at least one day, at least two days or at least three days. In some embodiments, the method includes continuous expression of the Zscan4-dependent gene (or presence of the Zscan4dependent gene protein) for at least one day, at least two days or at least three days.

[0139] In some embodiments of the disclosed method, the somatic cell is a murine cell. In other embodiments, the somatic cell is a human cell.

[0140] The somatic cell used in the disclosed method can be any type of cell, including a tissue stem cell, a progenitor cell or a differentiated cell. In some embodiments, the tissue stem cell is a neural stem cell, a hematopoietic stem cell, a mesenchymal stem cell or an adipose stem cell. In some embodiments, the differentiated cell is a fibroblast, lymphocyte (such as a T cell or a B cell), hepatocyte, epithelial cell, muscle cell, adipose cell, cardiomyocyte, pancreatic β cell, keratinocyte, amniotic cell, peripheral blood cell, platelet, or astrocyte.

[0141] In some embodiments, the method comprises contacting the somatic cell with a Zscan4 and at least one reprogramming factor. In particular embodiments, the Zscan4 is murine Zscan4, such as Zscan4c, Zscan4d or Zscan4f. In some examples, the murine Zscan4 is Zscan4c. In specific non-limiting example, the Zscan4c amino acid sequence is at least 95% identical to SEQ ID NO: 8; the Zscan4c amino acid sequence comprises SEQ ID NO: 8; or the Zscan4c amino acid sequence consists of SEQ ID NO: 8. In some examples, the Zscan4c is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 7; is encoded by a nucleotide

sequence comprising SEQ ID NO: 7; or is encoded by a nucleotide sequence consisting of SEQ ID NO: 7.

[0142] In other embodiments, the Zscan4 is human ZSCAN4. In some examples, the ZSCAN4 amino acid sequence is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 2; the ZSCAN4 amino acid sequence comprises SEQ ID NO: 2; or the ZSCAN4 amino acid sequence consists of SEQ ID NO: 2. In some examples, ZSCAN4 is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 1; ZSCAN4 is encoded by a nucleotide sequence comprising SEQ ID NO: 1; or ZSCAN4 is encoded by a nucleotide sequence consisting of SEQ ID NO: 1.

[0143] In some embodiments, the method includes contacting the somatic cell with a Patl2 and at least one reprogramming factor. In particular embodiments, the Patl2 is mouse Patl2. In some examples, the mouse Patl2 amino acid sequence is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 39; the Patl2 amino acid sequence comprises SEQ ID NO: 39; or the Patl2 amino acid sequence consists of SEQ ID NO: 39. In some examples, mouse Patl2 is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 38; Patl2 is encoded by a nucleotide sequence comprising SEQ ID NO: 38; or Patl2 is encoded by a nucleotide sequence consisting of SEQ ID NO: 38.

[0144] In particular embodiments, the Patl2 is human Patl2. In some examples, the human Patl2 amino acid sequence is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 47; the Patl2 amino acid sequence comprises SEQ ID NO: 47; or the Patl2 amino acid sequence consists of SEQ ID NO: 47. In some examples, human Patl2 is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 46; Patl2 is encoded by a nucleotide sequence comprising SEQ ID NO: 46; or Patl2 is encoded by a nucleotide sequence consisting of SEQ ID NO: 46.

[0145] In some embodiments, the method includes contacting the somatic cell with a Pramel6 and at least one reprogramming factor. In some examples, the Pramel6 amino acid sequence is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 41; the Pramel6 amino acid sequence comprises SEQ ID NO: 41; or the Pramel6 amino acid sequence consists of SEQ ID NO: 41. In

some examples, Pramel6 is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 40; Pramel6 is encoded by a nucleotide sequence comprising SEQ ID NO: 40; or Pramel6 is encoded by a nucleotide sequence consisting of SEQ ID NO: 40. In other examples, the Zscan4-dependent gene is a human ortholog of Pramel6. In particular examples, the human ortholog of Pramel6 is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to any one of SEQ ID NOs: 50-72.

[0146] In some embodiments, the method includes contacting the somatic cell with a Piwil2 and at least one reprogramming factor. In particular embodiments, the Piwil2 is mouse Piwil2. In some examples, the mouse Piwil2 amino acid sequence is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 43; the Piwil2 amino acid sequence comprises SEQ ID NO: 43; or the Piwil2 amino acid sequence consists of SEQ ID NO: 43. In some examples, mouse Piwil2 is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 42; Piwil2 is encoded by a nucleotide sequence comprising SEQ ID NO: 42; or Piwil2 is encoded by a nucleotide sequence consisting of SEQ ID NO: 42.

[0147] In particular embodiments, the Piwil2 is human Piwil2. In some examples, the human Piwil2 amino acid sequence is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 49; the Piwil2 amino acid sequence comprises SEQ ID NO: 49; or the Piwil2 amino acid sequence consists of SEQ ID NO: 49. In some examples, human Piwil2 is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 48; Piwil2 is encoded by a nucleotide sequence comprising SEQ ID NO: 48; or Piwil2 is encoded by a nucleotide sequence consisting of SEQ ID NO: 48.

[0148] In some embodiments, the method comprises contacting the somatic cell with a D5Ertd577e and at least one reprogramming factor. In some examples, the D5Ertd577e amino acid sequence is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 45; the D5Ertd577e amino acid sequence comprises SEQ ID NO: 45; or the D5Ertd577e amino acid sequence consists of SEQ ID NO: 45. In some examples, D5Ertd577e is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to

SEQ ID NO: 44; D5Ertd577e is encoded by a nucleotide sequence comprising SEQ ID NO: 44; or D5Ertd577e is encoded by a nucleotide sequence consisting of SEQ ID NO: 44. In other examples, the Zscan4-dependent gene is a human ortholog of D5Ertd577e. In particular examples, the human ortholog of D5Ertd577e is encoded by a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to any one of SEQ ID NOs: 50-72.

[0149] In alternative embodiments, the method includes contacting the somatic cell with an agent that increases expression of endogenous Zscan4 and contacting the cell with at least one reprogramming factor, thereby producing an iPS cell. It has previously been demonstrated that retinoids and oxidative stress increase transient expression of Zscan4 (see, PCT Publication No. WO 2011/028880). Thus, in particular examples, the agent is a retinoid, such as, but not limited to, all-trans retinoic acid, 9-cis retinoic acid, 13-cis retinoic acid and vitamin A. In other examples, the agent induces oxidative stress.

[0150] Further provided herein are isolated iPS cells produced according to the methods disclosed herein. Also provided are non-human animals (such as mice) produced from an iPS cell generated according to the disclosed methods.

[0151] The isolated iPS cells produced by the disclosed methods can be used for a variety of research and therapeutic purposes. For example, the iPS cells can be used in any regenerative medicine application as a replacement for ES cells or other stem cells. The number of iPS cells to be used and the mode of administration will vary depending upon the particular disease or disorder to be treated. In particular examples, about 1×10^6, about 2×10^6 or about 2×10^6 cells are injected. The iPS cells can be used either in the pluripotent state, or the cells can be differentiated to produce the desired cell type (such as neurons, muscle cells or cells of a particular organ). Methods of differentiating undifferentiated ES cells *in vitro* are known and can be applied to the differentiation of iPS cells. For example US Patent Application Publication No. 2006/0194321 describes differentiation of ES cells into endodermal cells (*e.g.*, pancreatic); US Patent Application Publication No. 2004/0014209 describes differentiation of ES cells into cardiac cells; US Patent Application Publication No. 2008/0194023 describes differentiation of ES cells into vascular smooth muscle cells; and US Patent Application Publication No. 2011/0117062 describes differentiating ES cells into retinal pigment epithelial cells.

[0152] Examples of disorders or diseases that can benefit from administration of iPS cells (or iPS cells that have been differentiated to a particular cell type) include autoimmune diseases, and

diseases in which cell regeneration is beneficial, such as neurologic injuries (such as brain or spinal cord injuries, or damage from stroke) or neurodegenerative disorders, as well as blindness, deafness, tooth loss, arthritis, myocardial infarctions, bone marrow transplants, baldness, Crohn's disease, diabetes, and muscular dystrophy. Exemplary neurodegenerative diseases include, for example, adrenoleukodystrophy (ALD), alcoholism, Alexander's disease, Alper's disease, Alzheimer's disease, amyotrophic lateral sclerosis (Lou Gehrig's Disease), ataxia telangiectasia, Batten disease (also known as Spielmeyer-Vogt-Sjögren-Batten disease), bovine spongiform encephalopathy (BSE), Canavan disease, cerebral palsy, Cockayne syndrome, Corticobasal degeneration, Creutzfeldt-Jakob disease, familial fatal insomnia, frontotemporal lobar degeneration, Huntington's disease, HIV-associated dementia, Kennedy's disease, Krabbe's disease, Lewy body dementia, neuroborreliosis, Machado-Joseph disease (Spinocerebellar ataxia type 3), Multiple System Atrophy, multiple sclerosis, narcolepsy, Niemann Pick disease, Parkinson's disease, Pelizaeus-Merzbacher Disease, Pick's disease, primary lateral sclerosis, prion diseases, progressive supranuclear palsy, Refsum's disease, Sandhoff disease, Schilder's disease, subacute combined degeneration of spinal cord secondary to Pernicious Anaemia, Spielmeyer-Vogt-Sjogren-Batten disease (also known as Batten disease), spinocerebellar ataxia, spinal muscular atrophy, Steele-Richardson-Olszewski disease, Tabes dorsalis, toxic encephalopathy. Exemplary autoimmune diseases that can benefit from the iPS provided herein include but are not limited to, rheumatoid arthritis, juvenile oligoarthritis, collagen-induced arthritis, adjuvant-induced arthritis, Sjogren's syndrome, multiple sclerosis, experimental autoimmune encephalomyelitis, inflammatory bowel disease (for example, Crohn's disease, ulcerative colitis), autoimmune gastric atrophy, pemphigus vulgaris, psoriasis, vitiligo, type 1 diabetes, non-obese diabetes, myasthenia gravis, Grave's disease, Hashimoto's thyroiditis, sclerosing cholangitis, sclerosing sialadenitis, systemic lupus erythematosis, autoimmune thrombocytopenia purpura, Goodpasture's syndrome, Addison's disease, systemic sclerosis, polymyositis, dermatomyositis, autoimmune hemolytic anemia, and pernicious anemia.

[0153] Also provided herein are methods of identifying mature and/or high quality iPSCs in a cell population by transfecting the cell population with an expression vector comprising a Zscan4 promoter operably linked to a reporter gene, wherein expression of the reporter gene in a cell of the cell population identifies the cell as a mature and/or high-quality iPSC. Further provided is a method of isolating mature iPSCs from a cell population, comprising transfecting the cell population with an expression vector comprising a Zscan4 promoter operably linked to a reporter gene, and separating cells expressing the reporter gene from the cell population, thereby isolating mature iPSCs.

[0154] In some embodiments, the cell population comprises iPSCs, embryonic fibroblasts, adult fibroblasts, or a combination thereof.

[0155] In some embodiments, the Zscan4 promoter comprises the Zscan4c promoter. In some examples, the Zscan4c promoter is at least 80%, at least 85%, 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to the nucleotide sequence set forth as nucleotides 906-4468 of SEQ ID NO: 73. In some embodiments, the reporter gene encodes a fluorescent protein, such as GFP or a derivative thereof (*e.g.* Emerald). Alternatively, the reporter gene encodes a drug (*e.g.*, antibiotic)-selectable marker, and the non-Zscan4-expressing cells are killed by adding the appropriate drug (*e.g.*, hygromycin, neomycin, etc.). In specific non-limiting examples, the nucleotide sequence of the expression vector is at least 80%, at least 85%, 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to the nucleotide sequence set forth as SEQ ID NO: 73. In specific examples, the expression vector comprises the nucleotide sequence of SEQ ID NO: 73.

V. Methods of Introducing Zscan4, Zscan4-Dependent Genes and Reprogramming Factors into Somatic Cells

[0156] A number of methods have been described in the art for delivery of reprogramming factors to somatic cells for the generation of iPSCs (for reviews, see Lai *et al.*, *J Assist Reprod Genet* 28(4):291-301, 2011; Parameswaran *et al.*, *Stem Cells* 29(7):1013-1020, 2011; Müller *et al.*, *Mol Ther* 17(6):947-953, 2009). The methods include several broad categories, including introduction of nucleic acid molecules encoding the reprogramming factors using a viral vector (such as integrating or non-integrating viral vectors) or a plasmid vector, delivery of mRNA molecules encoding the reprogramming factors, or direct delivery of the reprogramming factor proteins. Each of these methods has been described in the art and is therefore within the capabilities of one of skill in the art. A brief summary of each method that can be used to deliver Zscan4, a Zscan4-dependent gene and/or one more reprogramming factors to a somatic cell is provided below. It is not necessary for Zscan4 or the Zscan4dependent gene and each of the reprogramming factors to be delivered by the same method. For example, delivery of Zscan4 (or Zscan4-dependent gene) mRNA can be combined with vector-mediated delivery of the reprogramming factor(s).

A. Viral Vectors

[0157] The initial experiments carried out to produce iPSCs used retrovirus vectors (*e.g.*, Moloney murine leukemia virus (MMLV)-based vectors) to deliver the reprogramming factors to

murine and human somatic cells (Takahashi *et al.*, *Cell* 126:663-666, 2006; Takahashi *et al.*, *Cell* 31:861-872, 2007; Okita *et al.*, *Nature* 313-317, 2007; Park *et al.*, *Nature* 451:141-146; U.S. Patent Application Publication No. 2009/0047263). Subsequent studies utilized lentivirus vectors (Brambrink *et al.*, *Cell Stem Cell* 2:151-159, 2008; Wernig *et al.*, *Nat Biotechnol* 26:916-924, 2008; Stadtfeld *et al.*, *Science* 322:945-949, 2008), which had the advantage of being able to infect both dividing and non-dividing cells, thereby improving the rate of cell transduction. In addition, lentiviruses can be pseudotyped to expand viral tropism. For example, pseudotyping with vesicular stomatitis virus glycoprotein (VSVg) enables infection of a wide range of cell types (Lai *et al.*, *J Assist Reprod Genet* 28(4):291-301, 2011). Lentiviruses also allow for both constitutive and inducible expression of the reprogramming factors. Examples of drug-inducible lentivirus expression systems are described by Hockmeyer *et al.* (*Cell Stem Cell* 3:346-353, 2008) and Wernig *et al.* (*Nat Biotechnol* 26:916-924, 2008).

[0158] Lentiviruses include, but are not limited to, human immunodeficiency virus (such as HIV-1 and HIV-2), feline immunodeficiency virus, equine infectious anemia virus and simian immunodeficiency virus. Other retroviruses include, but are not limited to, human T-lymphotropic virus, simian T-lymphotropic virus, murine leukemia virus, bovine leukemia virus and feline leukemia virus. Methods of generating retrovirus and lentivirus vectors and their uses have been well described in the art (see, for example, U.S. Patent Nos. 7,211,247; 6,979,568; 7,198,784; 6,783,977; and 4,980,289).

[0159] Non-integrating viral vectors, such as adenovirus vectors, have also been used to deliver reprogramming factor nucleic acid molecules to cells. For example adenovirus vectors, which remain in episomal form in cells, have been successfully used to deliver to produce iPS cells from mouse fibroblasts and liver cells by delivery of Oct4, Sox2, Klf4, and c-Myc (Stadtfeld *et al.*, *Science* 322:945-949, 2008).

B. Plasmid Vectors

[0160] In some instances, it is desirable to use non-viral vectors, such as to avoid integration into the host cell genome. Thus, Zscan4, the Zscan4-dependent gene and/or one or more reprogramming factors can be delivered to a somatic cell using one or more plasmid vectors. Plasmid vectors are episomally maintained and generally exhibit a short duration of gene expression (Lai *et al.*, *J Assist Reprod Genet* 28(4):291-301, 2011). As one example, Okita *et al.* (*Science* 322:949-953, 2008) describe the use of the pCX plasmid, containing a CAG promoter, for the expression of reprogramming factors in somatic cells. In this study, fibroblasts were

serially transfected with two plasmids, one expressing c-Myc and the other expressing Oct4, Klf4 and Sox2. A later study by Gonzalez *et al.* (*Proc Natl Acad Sci USA* 106:8918-8922, 2009) successfully used a single polycistronic plasmid encoding all four reprogramming factors (c-Myc, Oct4, Klf4 and Sox2) to generate iPS cells.

[0161] Episomal plasmid vectors are a further option for introducing Zscan4 or the Zscan4-dependent gene and reprogramming factors into somatic cells. Episomal plasmid vectors are capable of replicating themselves autonomously as extrachromosomal elements, and therefore exhibit prolonged gene expression in target cells. An episomal plasmid vector derived from the Epstein Barr virus (oriP/EBNA1) has been used to reprogram human somatic cells by expression of OCT4, SOX2, NANOG, LIN28, c-MYC and KLF4 (Yu *et al.*, *Science* 324:797-801, 2009).

[0162] Selection of an appropriate vector is well within the capabilities of one of skill in the art. Expression vectors typically contain an origin of replication, a promoter, and optionally include specific genes to allow for phenotypic selection of the transformed cells (*e.g.* an antibiotic resistance cassette). Generally, the expression vector will include a promoter. The promoter can be inducible or constitutive. The promoter can also be tissue specific. Exemplary promoters include the CAG promoter, thymidine kinase promoter (TK), metallothionein I, polyhedron, neuron specific enolase, thyrosine hyroxylase, beta-actin, CMV immediate early promoter, or other promoters. Optionally, an enhancer element is also included, and can generally be located anywhere on the vector and still have an enhancing effect on gene expression.

[0163] Plasmid vectors can be introduced into somatic cells using any suitable method. In some embodiments, the vector is delivered to a cell by transfection using a lipid of cationic polymer. In particular examples, the transfection reagent is LIPOFECTAMINE™, or a similar reagent. In other examples, delivery is achieved using the nucleofection transfection technology (Amaxa, Cologne, Germany). This technology is based on an electroporation technique using the NUCLEOFECTOR™ delivery device to introduce DNA directly into the host cell nucleus (Lakshmipathy *et al.*, *Stem Cells* 22:531-543, 2004). In yet another example, the transfection reagent comprises poly-β-amino esters. Montserrat *et al.* (*J Biol Chem* 286(14):12417-12428, 2011) describe the production of iPS cells from human fibroblasts by delivery of a polycistronic CAG-promoter driven plasmid expressing Oct4, Sox2, Klf4 and c-Myc using poly-β-amino esters as the transfection reagent.

C. Excision strategies

[0164] Excision of exogenous reprogramming factors from genomic integration sites can be desirable. Two excision-based methods have been previously described, CreloxP recombination and piggyBac transposition. Soldner *et al.* (*Cell* 136:964-977, 2009) described the use of the Cre-lox system to produce iPS cells free of viral reprogramming factors. This strategy included positioning a *loxP* site in the 3' LTR of a lentivirus vector that contained a Dox-inducible minimal CMV promoter to drive expression of the reprogramming factors. During proviral replication, *loxP* was duplicated into the 5' LTR, resulting in genomic integration of the reprogramming factors flanked by two *loxP* sites. Transient expression of Cre-recombinase resulted in excision of the floxed reprogramming factors.

[0165] The piggyBac transposon is capable of excising itself without leaving any remnants of exogenous DNA in the cell genome (Elick *et al.*, *Genetica* 98:33-41, 1996; Fraser *et al.*, *Insect Mol Biol* 5:141-151, 1996). Using this method, iPS cells have been generated from fibroblasts by delivery of a polycistronic construct carrying reprogramming factor genes linked with a 2A peptide linker positioned between the piggyBac transposon 5' and 3' terminal repeats. Precise excision of the integrated reprogramming genes is observed upon expression of the transposase (Kaji *et al.*, *Nature* 458:771-775, 2009; Wang *et al.*, *Proc Natl Acad Sci USA* 105:9290-9295, 2008; Yusa *et al.*, *Nat Methods* 6:363-369, 2009).

D. mRNA

[0166] Another strategy for introducing Zscan4, a Zscan4-dependent gene and/or one or more reprogramming factors to a somatic cell is by delivery of mRNA encoding Zscan4, the Zscan4-dependent gene or the reprogramming factor(s). iPSCs have been generated from multiple different human cell types by administration of synthetic mRNA encoding reprogramming factors (Warren *et al.*, *Cell Stem Cell* 7(5):618-630, 2010). In the study by Warren *et al.*, the mRNA was modified to overcome innate antiviral responses and was delivered repeatedly to achieve pluripotency.

E. Protein

[0167] It is also possible to introduce Zscan4, the Zscan4-dependent gene and/or the reprogramming factors by directly delivering the respective proteins to the somatic cells. Protein delivery can be accomplished using, for example, electroporation, microinjection, cationic lipids or nanoparticles according to standard methods. Alternatively, the proteins can be modified by

fusion with a cell-penetrating peptide (CPP) to facilitate entry of the protein into the cell. The use of CPPs and nanoparticles is discussed in greater detail below.

1. Cell-Penetrating Peptides (CPPs)

[0168] CPPs are a family of polypeptides that facilitate transduction of proteins, nucleic acids or other compounds across membranes in a receptor-independent manner (Wadia and Dowdy, *Curr. Protein Pept. Sci. 4(2)*:97-104, 2003). Typically, CPPs are short polycationic sequences that can facilitate cellular uptake of compounds to which they are linked into endosomes of cells.

[0169] The capacity of certain peptides to deliver proteins or nucleic acids into cells was originally described for the HIV-encoded Tat protein, which was shown to cross membranes and initiate transcription. It was then discovered that the portion of the Tat protein that was required for the transduction of the protein was only an 11 amino acid polypeptide, referred to as the Tat peptide. When fused with other proteins, the Tat peptide has been demonstrated to deliver these proteins, varying in size from 15 to 120 kDa, into cells in tissue culture (Frankel and Pabo, *Cell 55(6)*:1189-93, 1988; Green and Loewenstein, *J. Gen. Microbiol. 134(3)*:849-55, 1988; Vives *et al.*, *J. Biol. Chem. 272(25)*:16010-7, 1997; Yoon *et al.*, *J. Microbiol. 42(4)*:328-35, 2004; Cai *et al.*, *Eur. J. Pharm. Sci. 27(4)*:311-9, 2006).

[0170] Other known CPPs include PENETRATIN™, a 16 amino acid peptide derived from the third helix of the Drosophila Antennapedia homeobox gene (U.S. Patent No. 5,888,762; Derossi *et al.*, *J. Biol. Chem. 269*:10444-10450, 1994; Schwarze *et al.*, *Trends Pharmacol. Sci. 21*:45-48, 2000); transportan, a 27 amino acid chimeric peptide comprised of 12 amino acids from the N-terminus of the neuropeptide galanin and the 14-amino acid protein mastoparan, connected via a lysine (U.S. Patent No. 6,821,948; Pooga, *FASEB J. 12*:67-77, 1998; Hawiger, *Curr. Opin. Chem. Biol. 3*:89-94, 1999); peptides from theVP22 protein of herpes simplex virus (HSV) type 1 (Elliott *et al.*, *Cell 88*:223-233, 1997); the UL-56 protein of HSV-2 (U.S. Pre-Grant Publication No. 2006/0099677); and the Vpr protein of HIV-1 (U.S. Pre-Grant Publication No. 2005/0287648). In addition, a number of artificial peptides also are known to function as CPPs, such as poly-arginine, poly-lysine and others (see, for example, U.S. Pre-Grant Publication Nos. 2006/0106197; 2006/0024331; 2005/0287648; and 2003/0125242; Zhibao *et al.*, *Mol. Ther.* 2:339-347, 2000; and Laus *et al. Nature Biotechnol.* 18:1269-1272, 2000).

[0171] Zhou *et al.* (*Cell Stem Cell* 4:381-384, 2009) describe the successful generation of iPS cells by fusing purified recombinant reprogramming factors OCT4, SOX2, KLF4 and c-MYC to poly-arginine peptide tags. Mouse embryonic fibroblasts were transduced with the recombinant four times and cultured in the presence of the histone deacetylase inhibitor, valproic acid (VPA) for 30-35 days. In addition, Kim *et al.* (*Cell Stem Cell* 4:472-476, 2009) describe reprogramming of human fetal fibroblasts by transduction of OCT4, SOX2, KLF4 and c-MYC proteins fused to the HIV-TAT protein transduction domain.

2. Nanoparticles

[0172] Nanoparticles are submicron (less than about 1000 nm) sized drug delivery vehicles that can carry encapsulated drugs such as synthetic small molecules, proteins, peptides, cells and nucleic acid based biotherapeutics for either rapid or controlled release. A variety of molecules (*e.g.*, proteins, peptides and nucleic acid molecules) can be efficiently encapsulated in nanoparticles using processes well known in the art.

[0173] In some examples, the Zscan4 protein, the Zscan4-dependent gene protein and/or a reprogramming factor protein is encapsulated by a nanoparticle to aid in delivery to the cells. Suitable nanoparticles for use with the disclosed methods are known in the art and are described briefly below.

[0174] The nanoparticles for use with the methods described herein can be any type of biocompatible nanoparticle, such as biodegradable nanoparticles, such as polymeric nanoparticles, including, but not limited to polyamide, polycarbonate, polyalkene, polyvinyl ethers, and cellulose ether nanoparticles. In some embodiments, the nanoparticles are made of biocompatible and biodegradable materials. In some embodiments, the nanoparticles include, but are not limited to nanoparticles comprising poly(lactic acid) or poly(glycolic acid), or both poly(lactic acid) and poly(glycolic acid). In particular embodiments, the nanoparticles are poly(D,L-lactic-*co*-glycolic acid) (PLGA) nanoparticles.

[0175] Other biodegradable polymeric materials are contemplated for use with the methods described herein, such as poly(lactic acid) (PLA) and polyglycolide (PGA). Additional useful nanoparticles include biodegradable poly(alkylcyanoacrylate) nanoparticles (Vauthier *et al.*, *Adv. Drug Del. Rev.* 55: 519-48, 2003).

[0176] Various types of biodegradable and biocompatible nanoparticles, methods of making such nanoparticles, including PLGA nanoparticles, and methods of encapsulating a variety of

synthetic compounds, proteins and nucleic acids, has been well described in the art (see, for example, U.S. Publication No. 2007/0148074; U.S. Publication No. 20070092575; U.S. Patent Publication No. 2006/0246139; U.S. Patent No. 5,753,234; U.S. Patent No. 7,081,489; and PCT Publication No. WO/2006/052285).

[0177] The following examples are provided to illustrate certain particular features and/or embodiments. These examples should not be construed to limit the disclosure to the particular features or embodiments described.

EXAMPLES

Example 1:Zscan4 is an early embryonic factor required for direct reprogramming of somatic cells

[0178] This example describes the finding that Zscan4 initiates direct reprogramming of somatic cells by reactivating early embryonic genes.

Methods

Plasmid vector construction

[0179] To construct a pCAG-Zscan4-ERT2 plasmid, an entire open reading frame (ORF: 506 a.a.) of the mouse Zscan4c gene (Falco *et al.*, *Dev Biol* 307:539-550, 2007) was fused with ERT2 (a mutated ligand-binding domain of the human estrogen receptor (Feil *et al.*, *Proc Natl Acad Sci USA* 93:10887-10890, 1996); 314 a.a.) and cloned into *Xho*I/*Not*I sites of a plasmid (pPyCAGBstXI-IP; Niwa *et al.*, *Gene* 108:193-199, 1991). The resultant plasmid vector (pCAG-Zscan4-ERT2) expresses a Zscan4c-ERT2 fusion protein and a Puromycin-resistant protein driven by a strong CAG promoter (FIG. 1A). PiggyBac vectors (PB-TET-IRES-βgeo, PB-TET-MKOS, and PB-CAG-rtTA; Kaji *et al.*, *Nature* 458:771-775, 2009; Woltjen *et al.*, *Nature* 458:766-770, 2009) were purchased from Addgene. ORFs of Klf4-Oct4-Sox2 (KOS), Zscan4c, Zscan4cERT2 or DsRed were PCR-amplified by using a high-fidelity DNA polymerase (Pfx50 or Platinum Pfx from Invitrogen) and attB1/2 primers (FIG. 17) from PB-TET-MKOS or pCAG-Zscan4-ERT2 and cloned into a pDONR221 vector, respectively (Invitrogen). Subsequently, these DNA fragments were inserted into a PB-TET-IRES- βgeo destination vector thorough the Gateway System (Invitrogen). To construct a PB-TETZscan4c-ERT2-IRES-Histidinol dehydrogenase (His)-DsRed vector, an IRES-His-DsRed fusion protein was excised from a pBR-CAG-cHA-IRES-HisDsRed vector (Niwa *et al.*, *Gene* 108:193-199, 1991) by *Apa*I and *BamH*I

(blunt-ended) and inserted into an *ApaI/EcoR*I (blunt-ended) site of a PB-TET-Zscan4c or PB-TET-Zscan4cERT2 vector, respectively. To construct a PB-TET-IRES-HisDsRed destination vector, an attR1R2ccdB cassette was excised from a PB-TET destination vector with *ApaI* and *Sac*II and inserted into an *ApaI/Sac*II site of a PB-TET-Zscan4cERT2-HisDsRed vector. An ORF of human ZSCAN4 was amplified by PCR using attB1/2 primers (FIG. 17) from pReceiver-M50-ZSCAN4 (Genecopoeia) and cloned into a pDONR221 vector (Invitrogen). This fragment was inserted into a PB-TET-IRES-HisDsRed destination vector through the Gateway System (Invitrogen).

Generation of ES-ZERT cells

[0180] V6.5 ES (Eggan *et al.*, *Proc Natl Acad Sci USA* 98:6209-6214, 2001) cells derived from an F1 hybrid strain (C57BL/6 x 129/Sv) were purchased from Thermo Scientific Open Biosystem. ES cells were cultured at 37°C in 5% CO_2 in the complete ES medium: DMEM, 15% FBS, 1000 U/ml leukemia inhibitory factor (LIF) (ESGRO, Chemicon), 1 mM sodium pyruvate, 0.1 mM non-essential amino acids, 2 mM GlutaMAX™, 0.1 mM β-mercaptoethanol, penicillin, and streptomycin. V6.5 ES cells (5 x 10^5 cells) in suspension were cotransfected with 2 μg of a linearized pCAGZscan4-ERT2 vector and 0.4 μg of PL452 vector (a neomycin-resistant gene driven by a PGK promoter; Liu *et al.*, *Genome Res* 13:476-484, 2003) using Effectene™ (QIAGEN) according to the manufacturer's protocol, and plated in 100 mm cell culture dishes. After selecting with G418 for 8 days, resulting ES cell colonies were picked, expanded, and frozen. Subsequently, an ES-ZERT cell clone was selected based on the results of genotyping, qPCR, and puromycin-resistance.

Generation of ZERT chimeric mice

[0181] ICR females (Charles River, 8-12 weeks old) were used for superovulation by pregnant mare serum gonadotropin (PMSG) (Sigma) followed by human chorionic gonadotropin (hCG; Sigma) administration 48 hours later. After hCG administration, females were mated with ICR males and 2-cell embryos were collected by flushing oviducts. Recovered embryos were cultured to the blastocyst stage in KSOM (Millipore) medium for 3 days at 37°C in 5% CO_2. ES-ZERT cells (10-15 cells) were injected into 2N blastocysts and then transferred to E2.5 recipient females. After genotyping the pups, ZERT chimeric mice carrying a pCAG-Zscan4-ERT2 DNA were established.

MEF isolation

[0182] MEF-ZERT cells and MEF-WT cells were isolated from E13.5 embryos, which were obtained by crossing male ZERT mice to female ICR mice (FIG. 5A). MEF-WT (C57BL/6J x 129S6/SvEvTac) cells were isolated from E13.5 embryos, which were obtained by crossing between male 129SvEvTac mice and female C57BL/6J mice. These cells were plated into 10-cm plate in DMEM supplemented with 10% FBS, 1mM sodium pyruvate, 0.1 mM non-essential amino acids, 2 mM GlutaMAX™, 0.1mM β-mercaptoethanol, penicillin and streptomycin.

Cell growth analysis

[0183] To determine growth rate under Tamoxifen (Tmx)- or Tmx+ conditions, MEFs were passaged at a density of 3 x 10^5 cells/10 cm-plate in triplicate every 3 days. Live cells were scored using trypan blue.

PiggyBac-transfection and iPSC generation

[0184] iPSCs were generated as previously reported (Kaji *et al.*, *Nature* 458:771-775, 2009; Woltjen *et al.*, *Nature* 458:766-770, 2009) with some modifications. Briefly, MEFs were plated on gelatin-coated 6-well plates at a density of 1 x 10^5 cells/well in complete ES medium. After 24 hour incubation, MEFs were transfected with 1.6 μg of each plasmid, which were pCyL43 (Sanger institute; Wang *et al.*, *Proc Natl Acad Sci USA* 105:9290-9295, 2008), CAG-rtTA and PB-TET-MKOS or -KOS with or without PB-TET-Zscan4c or PB-TET-Zscan4cERT2, using Xfect (Clontech). After 24 hour, cells were fed with the complete ES medium with Doxycycline (Dox: 1.5 μg/ml) and with or without 200 nM 4-hydroxytamoxifen (Tmx). Colonies were picked after 12 days post-Dox induction and propagated as iPSCs in the complete ES medium with Dox (1.5 μg/ml), until Dox-independency was observed in replicate wells. Culture medium was changed every day.

Alkaline phosphatase staining

[0185] Cells were stained using Leukocyte Alkaline Phosphatase kit (Sigma) according to the manufacturer's protocol.

In vitro differentiation

[0186] Cells were dissociated by Accutase (Chemicon), counted, and propagated using the hanging drop method (200 single cells per 25 μl) in the ES medium without LIF for 4 days.

Embryoid bodies, formed in the hanging drop, were transferred to gelatin-coated 24-well plates and cultured for 7 days, before being fixed with 4% paraformaldehyde.

Immunocytochemistry

[0187] Cells were fixed with 4% paraformaldehyde for 20 min and permeabilized with 0.2% triton-X 100 for 15 minutes. Cells were blocked with FX Signal Enhancer (Invitrogen). Primary antibody was added and incubated overnight at 4°C (SSEA-1 (500:1, Invitrogen), Nanog (500:1, BETHYL laboratories), α-smooth muscle actin (α SMA 1:1000, R&D), α -fetoprotein (AFP 1:500, R&D), GATA-4 (1:200, Santa Cruz) or β III-tubulin (1:500, Millipore)). Alexa488 anti-mouse IgG or IgM (Invitrogen) or Alexa-594 anti-Rabbit IgG (Invitrogen) were used as secondary antibodies. Nuclei were visualized with DAPI (Roche).

Nanog immunohistochemistry

[0188] Nanog expression of ESC or iPSC colonies was detected by the Envision™ system-HRP kit (Dako, USA) according to the manufacturer's instructions. In brief, cells were washed by PBS and fixed in Acetone/Methanol (1:1) for 20 minutes at 4°C. After PBS washing, peroxidase blocking was applied to remove exogenous peroxidase in cells for 5 minutes. Cells were blocked by 1% BSA for 10 minutes at room temperature and then incubated for 1 hour at room temperature with the primary anti-Nanog antibody (Bethyl, USA) diluted 1:500 in the blocking solution. The bound antibody was visualized with a peroxidase labeled polymer for 30 minutes and substrate chromogen for 5 minutes under an Axiovert™ microscope.

RT-PCR, quantitative PCR, and Genotyping PCR

[0189] One μg of total RNA was reverse-transcribed using Superscript III reverse transcriptase (Invitrogen). qPCR analysis was performed for 10 ng cDNA/well in triplicate using SYBR™ green master mix (Applied Biosystems) according to the manufacturer's protocol. Reactions were run on 7900HT or 7500 system (Applied Biosystems). Genotyping PCR was performed using TITANIUM Taq PCR kit (Clontech). Primers are shown in FIG. 17.

Karyotype analysis

[0190] iPS cells were treated with 0.1 μg/ml colcemid (Invitrogen) for 3 hours to induce the metaphase arrest, treated with 0.56% KCl and fixed with Methanol:Glacial Acetic acid (3:1). Slides were air-dried before Giemsa staining.

Tetraploid complementation

[0191] ICR females (Charles River, 8-12 week old) were used for superovulation by PMSG (Sigma), followed by hCG (Sigma) administration 48 hours later. After hCG administration, females were mated with male ICR mice and 2-cell embryos were collected by flushing oviducts. Recovered embryos were cultured in KSOM (Millipore) medium for 3 days at 37°C in 5% CO_2. Collected 2-cell embryos were directly transferred into 0.3M Mannitol solution and aligned automatically by alternate current (AC) pulse in an electrofusion chamber. Then two direct current (DC) pulses with 140 V/mm were applied for 40 μs using LF101 Electro Cell Fusion Generator. Fused embryos (4N) that had one blastomere were collected at 60 minutes of cultivation and then continued to culture in KSOM medium until they reached the blastocyst stage. iPS cells (10-15 cells) were injected into 4N blastocysts to assess their developmental potency and then transferred to E2.5 recipient females. Pups were harvested by cesarean section at E13.5.

Isolation of secondary MEFs and induction of secondary iPSCs

[0192] Secondary MEFs were isolated from E13.5 embryos, which were harvested by tetraploid complementation. Secondary MEFs were plated on gelatin-coated 6-well plates at a density of 1 x 10^5 cells/well in the complete ES medium. After 24 hour incubation, secondary MEFs were fed with complete ES medium with or without Dox (1.5 μg/ml) and with or without 200 nM Tmx. Culture medium was changed every day. Withdrawal of drugs (Dox or Tmx) was always followed by 1x washing by PBS before changing culture medium.

Microarray Data Analysis

[0193] Expression profiling was carried out as described (Nishiyama *et al.*, *Cell Stem Cell 5*, 420-433, 2009) using the whole-genome 60-mer oligonucleotide microarrays (Agilent; Carter *et al.*, *Genome Biol* 6:R61, 2005). Data analysis and visualization were carried out by the NIA Array Analysis Software (Sharov *et al.*, *Bioinformatics* 21:2548-2549, 2005). Heatmaps were generated with MultiExperiment Viewer (Mev) v4.2 (Saeed *et al.*, *Methods Enzymol* 411:134-193, 2006). The data were normalized in each row (gene) in the Gene/Row Adjustment of the program. The color scale was set from -3 to +3 based on the normalized data. The map displayed is from HCL (Hierarchical clustering) of MeV4.2. All the microarray data have been submitted to the public database GEO (accession number GSE28436).

Bisulfite sequencing

[0194] Genomic DNAs were extracted using DNeasy (Qiagen). Bisulfite conversion of genomic DNA was performed with EpiTect plus (QIAGEN) according to the manufacturer's protocol. Amplified products were purified by gel extraction and cloned into pCR 2.1 vector (Invitrogen). PCR primers are shown in FIG. 17. Randomly picked clones were sequenced with M13 forward or reverse primers. Sequencing data were analyzed using QUMA[36] (online at http://quma.cdb.riken.jp/top/index.html; Kumaki *et al.*, *Nucleic Acids Res* 36:W170-175, 2008). Data were excluded as low quality, if >5% unconverted CpHs or >10% alignment mismatches.

Statistical analysis

[0195] Data were evaluated using ANOVA with Scheffé's *post hoc* analysis for multiple comparisons and *t* tests for two groups. $P < 0.05$ was accepted as statistically significant.

Results

[0196] To investigate whether Zscan4 is reactivated during iPSC formation, mouse ES cells were generated that carry an Emerald (EM: a GFP variant) reporter driven by a 3.5 kb Zscan4 promoter (Zscan4-Emerald; SEQ ID NO: 73), which can reproduce the expression pattern of endogenous Zscan4 in mouse ES cells (Zalzman *et al.*, *Nature* 464:858-863, 2010; PCT Publication No. WO 2011/028880). Chimeric mice produced by injecting the ES cells (named ES-pZ-EM) into blastocysts, were used to generate E13.5 embryos, which were subsequently used to derive mouse embryo fibroblasts (MEFs). The MEFs confirmed for the presence of an Emerald reporter by genotyping were named MEF-pZ-EM cells (FIG. 19A). Emerald fluorescence was not detectable in the MEF-pZ-EM cells, indicating that Zscan4 is not expressed in MEFs.

[0197] A piggyBac vector (PB-TET-MKOS) (Kaji *et al.*, *Nature* 458:771-775, 2009; Woltjen *et al.*, *Nature* 458:766-770, 2009) carrying doxycycline (Dox)-inducible Myc (M), Klf4 (K), Oct4 (O), and Sox2 (S) was then transfected into the MEF-pZ-EM cells, and then the cells were cultured in ES cell media supplemented with Dox. As reported, colonies with an authentic ES-like morphology (denoted herein MOR+) were clearly visible by day 13 (FIG. 19B). The cells were observed under fluorescence microscopes every day, but no EM+ cells were found in the culture. Twenty-eight MOR+ colonies were picked and passaged into ES cell culture media without Dox 11 to 14 days after the piggyBac transfection. Two clones did not survive, but the other 26 clones proliferated to form MOR+ colonies. Colonies with EM+ cells began to appear

from the day 15 and by day 28 all the colonies showed the presence of EM+ cells in the same pattern as typical F1 hybrid ES cell lines: a small number (1.0±0.5%, S.E.M.) of EM+ cells (FIGS. 19C and 19D). The data indicate that iPSCs, once formed, gained the expression pattern of Zscan4 similar to ESCs, suggesting a close similarity between iPSC and ESCs. Importantly, Zscan4 was not activated during the early phase of iPSC formation by the MKOS factors. Consistent with this finding, reanalysis of the published microarray data (Samavarchi-Tehrani *et al.*, *Cell Stem Cell* 7:64-77, 2010; Sridharan *et al.*, *Cell* 136:364-377, 2009) did not reveal the activation of Zscan4 during the early phase of iPSC formation.

[0198] The data also indicate that, as reported previously (Kaji *et al.*, *Nature* 458:771775, 2009; Woltjen *et al.*, *Nature* 458:766-770, 2009), the production of iPSC by a piggyBac vector carrying MKOS factors was robust and the majority of MOR+ colonies indeed became the authentic iPSCs, as confirmed by the same unique expression pattern of Zscan4 as typical ES cells. This notion was further supported by the presence of a pluripotency marker NANOG (FIGS. 20A and 20B), which was not used as an exogenous iPSC factor, and the expression of a GFP marker under the control of the Oct4 promoter (Brambrink *et al.*, *Cell Stem Cell* 2:151-159, 2008) (FIG. 20C) in the majority of MOR+ colonies generated by the PB-TET-MKOS. Therefore, in the subsequent experiments, the MOR+ phenotype combined with the alkaline phosphatase staining (ALP+) was primarily used to score the number of authentic iPSCs.

[0199] Considering the critical roles of Zscan4 in 2-cell embryos (Falco *et al.*, *Dev Biol* 307:539-550, 2007) and ES cells (Zalzman *et al.*, *Nature* 464:858-863, 2010), it was hypothesized that Zscan4 could enhance the efficiency and quality of iPS cells. To test this notion, MEFs (named MEF-ZERT) were generated in which mouse Zscan4c fused to ERT2 (a mutated ligand-binding domain of the human estrogen receptor, which can be activated by the presence of tamoxifen [Tmx]; Feil *et al.*, *Proc Natl Acad Sci USA* 93:10887-10890, 1996) was constantly expressed under a strong and ubiquitous CAG-promoter (Niwa *et al.*, *Gene* 108:193-199, 1991) (FIG. 1A, FIG. 5). Next the effect of Zscan4 on iPSC formation was tested by transfecting a piggyBac vector (PB-TET-MKOS; Kaji *et al.*, *Nature* 458:771-775, 2009; Woltjen *et al.*, *Nature* 458:766-770, 2009) carrying doxycycline (Dox)-inducible Myc (M), Klf4 (K), Oct4 (O), and Sox2 (S), into control MEF-WT and MEF-ZERT cells and culturing them with or without Tmx in standard iPSC generation conditions (Dox+) (FIG. 6A). By day 13, colonies with ES-like morphologies were clearly visible (FIG. 6B). iPSC colonies were scored based on their authentic ES cell morphology and ALP-staining. In the control MEF-WT, the efficiency of iPSC formation was slightly lower in the Tmx+ condition than the Tmx- condition (FIGS. 6C and 6D).

By contrast, in the MEF-ZERT cells, Tmx+ conditions increased the number of iPSC colonies by 1.5- to 2-fold, suggesting that the continuous presence of Zscan4 can enhance the efficiency of iPSC formation by the MKOS factors (FIG. 6C and FIG. 6D). The formation of authentic iPSCs was confirmed by the RT-PCR analysis of pluripotency genes, alkaline-phosphatase (ALP) staining of colonies, immunohistochemistry of pluripotency markers, embryoid body formation, and the ability to differentiate into three germ layers (FIG. 7). Similar enhancement of iPSC formation by Zscan4 was observed on wild-type MEF cells (MEF-WT), excluding the possibility that enhancement of iPSC formation is a unique feature of MEF-ZERT cells (FIGS. 8, 9 and 10). It was also determined that the human ZSCAN4 gene functions similar to mouse Zscan4 in the iPSC formation (FIG. 11).

[0200] Previous studies have shown that oncogene Myc (Feng *et al.*, *Cell Stem Cell* 4:301-312, 2009; Hu *et al.*, *Proc Natl Acad Sci USA* 107:4335-4340, 2010), which increases cell proliferation and suppresses genome stability, is required for the efficient iPSC formation (Takahashi and Yamanaka, *Cell* 126:663-676, 2006; Hanna *et al.*, *Cell* 143:508-525, 2010). To test whether Zscan4 (Z) can enhance the iPSC formation without Myc, MEF-WT and MEF-ZERT cells were transfected with a PB-TET-KOS vector carrying Dox-inducible KOS factors and the cells were cultured in the Dox+ Tmx+ or Dox+ Tmx- conditions for 2 weeks (FIG. 1B). As expected, MEF-WT cells produced only a small number of ALP+ iPSC colonies (FIG. 1C). However, MEF-ZERT cells produced 40- o 70-fold more iPSC colonies in Dox+ Tmx+ (*i.e.*, ZKOS) conditions than control Dox+ Tmx- (*i.e.*, KOS) conditions (FIG. 1C). The result was confirmed in independent experiments (FIG. 12) and the quality of iPSCs generated with ZKOS factors was also demonstrated (FIG. 13). The number of iPSC colonies generated by ZKOS factors (FIG. 1C) was comparable to that generated by MKOS factors (FIG. 6). Thus, in combination with KOS factors, Zscan4 can replace Myc for iPSC formation.

[0201] To clarify further a role of Zscan4 in reprogramming, studies were carried out to determine when and how long Zscan4 is required. It has been shown that efficient iPSC formation requires the ectopic overexpression of MKOS factors at least for 9 days (Wernig *et al.*, *Nat Biotechnol* 26:916-924, 2008; Sridharan *et al.*, *Cell* 136:364-377, 2009). Using the Tmx-inducible system of Zscan4 expression in MEF-ZERT cells, which was independent of the Dox-inducible expression of KOS factors, it was possible to vary the time of exposure to Zscan4 (FIG. 1B). The absence of Zscan4 for the first 4 or 7 days (*i.e.*, adding the Tmx from day 4 or 7) significantly reduced the number of iPSC colonies, whereas the presence of Zscan4 only for the first 4 or 7 days (*i.e.*, removing the Tmx after day 4 or 7) was sufficient to produce the number of

iPSC colonies comparable to that obtained by the presence of Zscan4 for 20 days (FIG. 1D). Further experiments with progressive shortening of the exposure established that Zscan4 is required only for the initial day of iPSC formation (FIG. 1E). These data indicate that, unlike other factors, Zscan4 is involved in the early phase of reprogramming.

[0202] It has been shown that the secondary MEF system facilitates the analysis of early events leading to iPSC formation, because every cell carries inducible iPSC factors (Wernig *et al.*, *Nat Biotechnol* 26:916-924, 2008; Hanna *et al.*, *Nature* 462:595-601, 2009). Therefore, secondary MEF cells were generated by injecting iPSCs (clone #2, FIG. 14) produced with ZKOS factors into tetraploid (4N) blastocysts. The result was production of 2 live E13.5 embryos (FIG. 14G), which were subsequently used to derive the secondary MEFs (named MEF-KOS-ZERT^{2nd}) (FIG. 2A). As the embryos were entirely derived from the iPSCs in tetraploid complementation system (Nagy *et al.*, *Development* 110:815-821, 1990), the results demonstrated that iPSCs generated with ZKOS factors were fully pluripotent. This iPSC clone also showed >80% normal karyotype and global gene expression profiles indistinguishable from ES cells (FIG. 15). Without ZKOS induction (*i.e.*, Dox- Tmx- and Dox- Tmx+ conditions), ALP+ iPSC colonies were not formed from the MEF-KOS-ZERT^{2nd} cells by day 17 (FIG. 2B).

[0203] By contrast, the forced expression of ZKOS factors (Dox+ Tmx+ condition) produced a large number (approximately 400) of ALP+ iPSC colonies, whereas the KOS factor only (Dox+ Tmx-condition) produced a much smaller number (about 20) of ALP+ iPSC colonies (FIG. 2B). Interestingly, even the KOS factors alone caused dramatic changes in the MEF morphology by day 1: cells became flatter and lost typical spindle-shape of MEFs (FIG. 2B). However, ES colony-like structures started to form by day 6 only in the Dox+ Tmx+ condition (*i.e.*, the forced expression of ZKOS factors) (FIG. 2B). Furthermore, the secondary MEF system also confirmed that Zscan4 is required only for the initial day of iPSC formation (FIG. 2C).

[0204] Consistent with a dramatic morphological change, the forced expression of KOS alone (Dox+ Tmx-) altered the transcriptomes of MEF-KOS-ZERT^{2nd} significantly by day 1: upregulation of 1730 genes and downregulation of 947 genes (FIG. 3A, first row). By day 3 and 6, the number of expression-altered genes further increased. However, considering that these conditions (Dox+ Tmx-) produced only a small number of iPSCs after 17 days in culture, these transcriptome alterations were not sufficient to convert MEFs to iPSCs. By contrast, a comparison between the iPSC-producing ZKOS condition (Dox+ Tmx+) and the KOS condition (Dox+ Tmx-) revealed only a minor transcriptome difference (FIG. 3A, fourth row): 28 genes by day 1, 162 genes by day 3, and 237 genes by day 6. That such a small difference in transcriptome

yielded such a great difference in the number of iPSC colonies was astounding, highlighting the critical contribution of these genes to the early phase of iPSC formation. After combining lists of these genes, 231 distinct genes were obtained that were more highly expressed in the ZKOS condition than in the KOS condition (with the statistical significance of FDR≤0.05, fold-change≥2) (FIG. 3A). For these 231 genes, the fold-differences between Tmx+ (*i.e.*, Zscan4+) and Tmx- (*i.e.*, Zscan4-) were calculated, subjected to the hierarchical clustering, and presented as a heatmap (FIG. 3B, Table 1).

Table 1. Genes upregulated in the ZKOS condition

Gene Symbol	Day 1 Dox- (Tmx+/Tmx-)	Day 1 Dox+ (Tmx+/Tmx-)	Day 3 Dox- (Tmx+/Tmx-)	Day 3 Dox+ (Tmx+/Tmx-)	Day 6 Dox- (Tmx+/Tmx-)	Day 6 Dox+ (Tmx+/Tmx-)
Pramel6	1.05	1.12	0.74	10.23	0.89	9.33
E330017A01Rik	1.00	0.83	1.00	2.40	1.00	7.76
Lce3f	1.20	0.89	0.81	1.35	1.70	6.76
Trim31	0.98	1.41	0.93	6.31	0.59	1.48
BC094916	1.00	0.83	1.00	1.26	1.00	5.89
Galnt13	1.10	1.12	1.10	0.87	1.10	5.89
Podnl1	0.89	0.58	0.89	5.75	0.89	1.10
Patl2	1.00	1.20	0.91	3.80	1.17	5.50
Esx1	1.00	1.58	0.93	1.26	1.02	4.90
AF067063	0.98	1.62	0.76	2.40	1.55	4.79
9130409J20Rik	1.12	1.10	1.00	1.41	1.07	4.79
1700003E24Rik	1.41	1.70	0.91	3.47	0.89	4.68
4930558C23Rik	1.05	2.40	0.95	3.63	1.02	4.47
D5Ertd577e	1.07	1.23	0.93	4.47	1.17	4.17
Trim42	1.00	0.63	1.00	4.47	1.00	2.51
A930002C04Rik	1.05	0.56	1.55	0.60	1.10	4.47
Dppa2	1.00	1.02	0.79	2.45	1.12	4.07
Gm22	1.02	4.07	1.05	0.87	0.95	1.10
Aadac	1.00	3.98	0.83	2.19	1.02	1.86
LOC677440	0.98	0.93	0.69	3.98	1.12	2.88
Mageb7-ps (EG637027)	0.93	1.95	0.93	3.98	0.91	1.15
D13Ertd608e	1.07	1.62	0.98	3.89	0.98	1.74
Abca12	1.00	1.26	1.00	2.19	0.93	3.89
Nlrp4c	0.95	1.91	1.17	3.89	1.26	0.63
Slc6a14	1.10	1.23	0.72	1.82	0.78	3.80
9030625G05Rik	0.81	0.81	1.66	3.80	0.81	1.10
C130073F10Rik	0.98	1.20	1.05	2.88	1.05	3.72
Cphx	0.95	1.74	0.91	3.72	0.95	1.62
Lce1f	1.23	1.29	1.00	1.05	0.78	3.63
4930459C07Rik	1.00	1.05	1.00	0.87	1.00	3.55
AU018829	1.15	1.48	0.71	2.57	1.05	3.55
4930519F16Rik	1.20	1.00	1.00	0.95	1.00	3.55
Calcoco2	1.00	1.07	0.78	1.91	1.20	3.47
Cyp2a4	1.00	0.85	1.00	1.45	1.02	3.47
Mogat1	0.91	1.10	0.98	0.98	1.07	3.47
Gdpd2	1.29	3.47	1.10	1.05	0.85	1.00
Ifit5	1.00	1.51	1.00	2.57	1.00	3.39
Nr5a2	1.10	1.32	0.81	3.24	1.02	3.31
LOC434660	1.29	1.74	0.68	2.40	0.98	3.31
Trim43b (EG666747)	0.79	0.91	1.00	1.26	1.00	3.31

Gene Symbol	Day 1 Dox- (Tmx+/Tmx-)	Day 1 Dox+ (Tmx+/Tmx-)	Day 3 Dox- (Tmx+/Tmx-)	Day 3 Dox+ (Tmx+/Tmx-)	Day 6 Dox- (Tmx+/Tmx-)	Day 6 Dox+ (Tmx+/Tmx-)
Sp8	1.00	1.20	1.00	2.88	1.17	3.24
Slc28a1	1.05	1.66	0.95	3.24	1.05	2.19
Ubtfl1 (B020006M18Rik)	1.10	1.23	0.76	1.78	0.68	3.24
LOC434136	1.12	1.07	0.76	1.17	0.98	3.24
C1qtnf9	1.15	1.12	0.79	0.98	0.91	3.16
1600029D21Rik	1.00	1.23	1.07	3.09	1.12	3.02
Tcl1	0.89	1.07	0.95	2.34	1.02	3.09
Cdx2	1.70	1.23	1.07	1.95	0.85	3.09
Khdc1c	1.00	1.12	1.00	1.45	1.00	3.09
Tarm1 (9930022N03Rik)	0.89	0.95	1.00	1.26	1.00	3.09
4732457N14	1.00	1.32	0.87	2.34	1.48	3.02
AU015836	0.93	0.78	1.15	3.02	0.95	2.29
Anxa10	0.69	3.02	1.00	1.82	1.00	1.23
D630045M09Rik	0.98	1.07	1.23	1.82	1.55	3.02
Prdm13	1.17	0.91	0.71	1.70	0.87	3.02
Mx1	1.17	2.95	1.20	1.02	1.12	0.91
Obox6	0.98	1.20	0.87	2.88	1.10	2.69
LOC668206	1.12	1.55	0.87	2.19	0.95	2.88
LOC432715	0.98	1.15	0.87	2.88	1.02	2.51
Mmp8	0.89	2.88	0.79	1.55	1.15	1.35
4933411G11Rik	1.26	0.91	0.95	1.20	0.91	2.88
D730045A05Rik	1.15	0.63	1.35	1.45	0.87	2.88
Fgf20	1.00	0.65	1.00	1.10	1.35	2.88
Spink5	1.10	0.58	1.26	1.02	1.10	2.88
Gabrr3	1.00	1.00	1.00	2.24	1.00	2.82
4930430J02Rik	0.95	1.35	0.91	2.82	0.93	1.86
Crct1	1.02	1.12	1.07	1.10	1.12	2.82
Cdcp1	1.29	1.26	0.87	2.82	1.00	2.75
4933402E13Rik	0.95	1.05	0.87	2.00	1.05	2.82
Rptn	1.29	1.70	1.10	1.12	1.55	2.75
Stox1	0.93	0.91	0.91	1.86	1.02	2.75
Trim43a (EG547109)	1.23	0.49	1.41	1.86	0.91	2.75
Gpx2-ps1	1.00	0.83	1.00	1.45	1.00	2.75
LOC677115	1.15	0.79	1.15	1.17	0.76	2.75
Ankrd22	0.98	1.35	0.81	2.69	1.07	2.09
LOC625360	0.98	1.05	0.81	2.69	1.02	2.00
Trpv3	1.00	1.00	1.00	1.32	1.00	2.69
RP23-438H3.2	1.35	1.05	0.98	1.91	0.93	2.69
Kremen2	1.00	1.05	0.91	1.38	0.69	2.69
Ldhc	1.07	1.95	0.91	2.63	1.00	2.04
A530040E14Rik	1.26	1.20	0.83	2.63	1.12	2.34
Zp3	1.20	1.15	2.04	2.34	1.15	2.63
AF067061	0.91	0.71	1.07	2.34	1.10	2.63
Fam155a (AW121567)	0.95	1.48	1.10	2.63	1.29	1.48
Sftpd	0.98	1.00	0.87	1.55	1.02	2.63
Sox30	1.17	1.10	1.17	1.20	0.81	2.63
Pof1b	0.98	1.00	0.89	2.29	1.07	2.57
B930018H19	1.55	1.15	1.07	1.91	1.26	2.57
Slc39a4	0.98	1.02	1.10	1.74	1.15	2.57
Spnb3	1.07	0.98	1.07	1.70	1.10	2.57
LOC672264	0.98	0.79	1.51	1.66	0.87	2.57
D10Bwg1379e	1.12	0.79	1.23	1.45	0.91	2.57
Tmprss11d	1.00	1.00	1.10	1.23	1.00	2.57
1700065L07Rik	0.85	2.57	0.62	0.91	0.76	0.91

Gene Symbol	Day 1 Dox- (Tmx+/Tmx-)	Day 1 Dox+ (Tmx+/Tmx-)	Day 3 Dox- (Tmx+/Tmx-)	Day 3 Dox+ (Tmx+/Tmx-)	Day 6 Dox- (Tmx+/Tmx-)	Day 6 Dox+ (Tmx+/Tmx-)
Slc23a3	1.10	1.35	1.05	2.14	1.00	2.51
2210418O10Rik	1.02	0.98	0.91	2.24	1.05	2.51
Tcstv3	0.83	0.78	1.02	2.40	0.95	2.51
Slc46a2	0.93	1.29	0.74	1.51	0.74	2.51
Mszf81	0.98	1.20	1.00	1.32	1.15	2.51
Nlrp4e	1.05	1.26	0.89	2.51	1.02	1.12
4931429I11Rik	1.12	2.51	1.00	1.10	1.00	1.17
Khdc1a	1.00	0.91	1.00	1.07	1.00	2.51
Adcyap1	1.29	0.76	1.29	1.20	1.58	2.51
Mnx1	0.87	0.78	1.20	0.98	0.91	2.51
Cgn	1.02	1.70	1.05	2.45	1.10	2.40
Wfdc15a	0.55	1.35	0.71	1.95	0.91	2.45
Tcstv1	1.15	0.89	0.87	2.40	1.26	2.45
B020031M17Rik	1.10	0.93	1.17	2.29	1.12	2.45
Fam25c (2200001I15Rik)	1.00	1.62	1.05	1.55	0.95	2.45
Lce1d	0.78	1.10	1.00	2.45	1.00	2.00
AI848258	1.02	1.12	0.74	2.45	1.07	1.32
Slc44a4	0.95	1.02	1.07	1.41	1.05	2.45
Slc34a3	1.00	0.81	1.51	1.58	0.87	2.45
Nr0b1	0.98	1.41	1.00	2.40	1.00	2.09
Bex6	1.48	1.10	0.89	2.40	1.07	2.40
Adh4	0.89	1.41	0.83	1.95	1.15	2.40
C130026I21Rik	1.41	0.83	1.91	2.19	1.07	2.40
Adad2	1.10	1.32	0.87	1.55	1.10	2.40
LOC673289	0.65	0.62	1.02	2.40	0.91	1.95
Sst	0.98	0.98	0.78	1.58	0.76	2.40
E230016M11Rik	1.12	1.26	0.93	2.40	0.85	1.07
Slc38a5	1.05	1.12	0.89	1.17	1.00	2.40
Fasl	1.17	0.95	0.95	1.26	0.98	2.40
4930538E20Rik	1.00	1.12	1.00	0.91	1.12	2.40
Gpr111	1.02	1.00	1.00	1.00	1.00	2.40
Gtsf1	1.05	1.05	1.10	0.91	1.15	2.40
Cfc1	1.02	1.48	0.89	2.34	0.95	1.74
Fbxo15	1.26	1.07	0.98	1.95	1.10	2.34
Plbd1 (1100001H23Rik)	1.05	0.95	1.10	2.04	1.15	2.34
Gm8016 (EG666272)	0.91	0.91	0.89	2.34	0.85	1.86
Mageb8-ps (EG436212)	1.00	1.00	1.00	1.70	1.12	2.34
Zbtb32	0.83	1.23	0.93	1.41	1.02	2.34
Eif1a	0.98	0.81	1.00	2.34	1.12	1.78
2010109I03Rik	1.00	1.00	0.58	1.41	1.00	2.34
BC066135	0.87	1.15	1.00	1.23	0.95	2.34
Vstm2a	1.00	1.00	1.17	2.34	0.95	0.98
Prom2	0.74	0.59	1.17	1.38	1.10	2.34
Hoxc13	1.00	0.85	0.87	2.34	0.85	0.98
LOC665276	1.00	1.00	1.00	0.69	1.00	2.34
Ankrd56	0.89	1.35	0.87	1.78	1.00	2.29
Tnk1	1.07	0.83	0.87	2.29	1.32	2.14
Prss8	0.68	1.00	1.02	1.95	1.32	2.29
Shisa3	1.32	1.41	0.83	2.29	1.23	1.35
Fndc3c1 (Gm784)	1.17	1.17	1.02	1.51	1.15	2.29
Btnl9	1.29	0.51	0.98	2.09	1.32	2.29
Vmn1r15 (V1rc6)	1.41	1.35	1.00	1.00	1.26	2.29

Gene Symbol	Day 1 Dox- (Tmx+/Tmx-)	Day 1 Dox+ (Tmx+/Tmx-)	Day 3 Dox- (Tmx+/Tmx-)	Day 3 Dox+ (Tmx+/Tmx-)	Day 6 Dox- (Tmx+/Tmx-)	Day 6 Dox+ (Tmx+/Tmx-)
Mmp19	0.91	1.05	0.98	2.29	0.91	0.91
Dkk1l	1.00	1.45	0.98	2.24	1.02	1.74
Gli1	0.98	1.10	0.95	2.04	0.91	2.24
Tceal7	1.26	1.05	1.17	2.24	1.07	2.00
1700024P16Rik	1.07	1.12	1.10	1.91	1.00	2.24
Trim6	1.15	1.15	1.00	1.82	1.07	2.24
Piwil2	1.17	1.48	0.81	1.45	1.02	2.24
Isg15	1.23	1.32	0.95	1.58	1.17	2.24
Gm5576 (EG434050)	0.85	0.98	1.00	1.78	0.91	2.24
Cldn7	0.93	0.93	0.98	1.78	1.10	2.24
Tmem30b	0.93	0.59	1.07	2.24	0.93	2.09
Lce3c	1.07	1.23	1.15	1.38	1.02	2.24
D030018L15Rik	0.93	1.26	1.12	1.32	0.79	2.24
1700042O10Rik	1.02	0.95	1.10	1.32	1.05	2.24
LOC672673	0.95	1.10	0.91	1.15	0.89	2.24
1700003M02Rik	0.81	1.35	0.79	2.24	1.45	0.87
1700008A04Rik	0.91	1.20	1.23	0.95	1.74	2.24
Gm5891 (EG545929)	0.83	0.85	0.89	2.09	1.35	2.19
Gm9124 (EG668356)	0.79	1.20	1.02	2.19	1.05	1.74
Crb3	1.12	2.19	1.02	1.58	1.12	1.35
Rab25	0.78	0.91	0.98	1.91	0.87	2.19
EG226955	0.69	1.02	1.20	2.19	1.55	1.70
Gm3336 (2410018E23Rik)	1.10	0.93	1.55	1.74	1.00	2.19
2310007B03Rik	0.89	1.35	0.98	1.32	1.45	2.19
Sp110	1.00	0.98	1.00	1.66	1.07	2.19
Gldc	1.02	1.05	0.98	1.32	1.07	2.19
Lce1a1	1.20	1.23	0.87	1.00	0.83	2.19
Meox1	0.83	1.07	1.02	1.10	1.00	2.19
Pglyrp3	1.10	1.15	1.00	1.00	0.95	2.19
Gpr115	1.10	1.07	1.66	1.00	0.65	2.19
Muc4	1.00	2.19	1.00	1.00	1.00	0.98
LOC626773	1.12	0.83	1.51	1.02	0.40	2.19
B230217J21Rik	1.10	0.91	1.70	0.65	0.83	2.19
Mbl2	1.15	1.62	0.71	2.14	0.50	1.62
1700016G22Rik	1.05	1.78	0.81	1.41	1.51	2.14
Hsh2d	1.07	1.05	0.72	2.14	1.48	2.09
Usp43	1.35	1.02	0.83	2.14	1.02	1.95
Mal2	1.15	2.14	1.02	1.45	1.10	1.45
Ssxb2	0.79	0.85	0.78	2.14	0.66	1.95
Prdm1	1.12	1.15	1.12	1.58	0.98	2.14
LOC233184	1.00	1.00	0.95	2.14	1.02	1.66
1700110K17Rik	0.56	1.26	1.10	2.14	0.81	1.38
BC013672	0.81	1.26	0.95	1.35	1.62	2.14
Cldn6	0.98	1.00	1.07	1.55	1.05	2.14
RP23-67E6.3	0.87	0.93	1.45	1.58	0.95	2.14
Pdc	0.87	1.00	0.76	1.41	0.98	2.14
Speer1-ps1	0.89	1.26	1.05	1.15	1.17	2.14
4933438K21Rik	1.10	1.20	0.93	1.10	0.87	2.14
Tas2r137	1.41	0.81	0.95	2.14	0.79	1.45
Tns4	1.12	1.02	0.95	1.07	1.20	2.14
Tcte2	1.29	0.87	0.95	1.05	0.81	2.14
Gpr152	1.35	0.68	1.15	0.74	0.83	2.14
LOC623810	1.00	0.60	1.00	0.46	1.00	2.14
Rhox4c	0.95	1.74	0.74	2.09	0.95	1.45

Gene Symbol	Day 1 Dox- (Tmx+/Tmx-)	Day 1 Dox+ (Tmx+/Tmx-)	Day 3 Dox- (Tmx+/Tmx-)	Day 3 Dox+ (Tmx+/Tmx-)	Day 6 Dox- (Tmx+/Tmx-)	Day 6 Dox+ (Tmx+/Tmx-)
Tsga8	1.55	1.41	0.72	1.66	1.15	2.09
Epcam(Tacstd1)	1.02	1.15	0.93	2.09	1.07	1.86
Bex1	0.93	1.17	1.02	1.78	1.00	2.09
Spint1	1.05	1.26	1.23	1.62	0.78	2.09
Gc	1.20	1.10	0.71	1.78	1.02	2.09
D7Ertd183e	1.51	1.55	1.07	0.93	0.76	2.09
Bcl2l14	1.00	1.15	1.02	1.20	1.12	2.09
Scel	1.07	1.00	1.12	1.07	1.23	2.09
Robo4	0.76	0.83	1.12	1.20	0.98	2.09
Ttll10	0.78	0.83	0.52	0.93	1.12	2.09
Cyp26b1	1.17	0.91	0.83	0.76	0.89	2.09
BC024997	0.95	1.78	1.00	2.04	1.32	1.70
Btn1a1	1.10	1.66	1.00	1.74	1.29	2.04
9330159N05Rik	1.45	1.45	1.12	1.58	1.38	2.04
1810019J16Rik	1.05	1.12	0.93	2.04	1.07	1.91
Capsl	0.89	1.23	0.89	1.78	0.95	2.04
A630095N17Rik	1.12	1.29	0.95	1.45	1.07	2.04
Grb7	0.95	0.98	1.05	1.55	0.98	2.04
LOC671025	1.07	1.07	1.12	1.35	1.20	2.04
Gm44	1.51	0.79	1.05	2.04	1.17	1.62
Pard6b	0.93	0.98	1.17	1.20	1.07	2.04
1730030J21Rik	1.07	1.05	1.20	2.04	1.35	1.05
Gpr112	1.00	0.91	1.00	1.17	1.00	2.04
Psg27	0.95	0.95	1.00	2.04	0.93	1.07
Mctp2	1.00	1.00	1.00	1.00	1.00	2.04
Il1f9	1.00	1.05	1.00	0.85	0.95	2.04
Colec10	1.02	1.00	1.26	0.56	1.07	2.04
Syt2	1.00	0.79	1.15	0.49	1.00	2.04
1700013H16Rik	1.10	1.23	0.85	2.00	1.10	1.48
LOC673795	1.00	1.00	1.00	1.51	1.00	2.00
Rab39	1.17	0.69	1.12	1.38	1.07	2.00

[0205] Some of these genes were originally identified from only preimplantation embryos through large-scale cDNA sequencing projects (Ko *et al.*, *Development* 127:1737-1749, 2000), *e.g.*, Pramel6, D5Ertd577e, D13Ertd608e, Tcstv1(Struwe and Solter, 1998, GenBank accession AF067057.1), Trim43a (Stanghellini et al., 2009), and Trim43b (Stanghellini *et al.*, *Gene Expr Patterns* 9:595-602, 2009). Accordingly, the public Expressed Sequence Tags (ESTs) database (NCBI/NIH) was searched and d the number of EST hits for each gene was scored. A significant fraction of the 231 genes was indeed expressed predominantly in early embryos and closely related gonads (testis and ovary): 27 genes in preimplantation stage (1-cell – blastocysts); 14 genes in oocytes; and 37 genes in testis/ovary (FIG. 3B). The preimplantation- and gonad-specific expression of these genes was also confirmed by the expression patterns in the NIA Gene Expression Atlas (Sharov *et al.*, *BMC Genomics* 12:102, 2011) (FIG. 3C) and the GNF database (Su *et al.*, *Proc Natl Acad Sci USA* 99:4465-4470, 2002) (FIG. 3D). Interestingly, most of these genes showed little or no expression in MEFs, ESCs, and iPSCs (FIG. 3C), indicating that these genes are activated transiently in the early phase of iPSC formation, but then

downregulated once iPSCs are fully formed. These genes may thus represent the earliest markers for eventual iPSC formation from MEFs, because they were fully activated as early as day 1, 3, or 6, when typical pluripotency markers such as Nanog, Zfp42 (also known as Rex1), and Dppa5a (also known as Esg1) were still silent (FIG. 16).

[0206] The early activation of preimplantation-specific genes seems to be unique to Zscan4-mediated iPSC formation. Indeed, previous microarray studies of the secondary MEF system using piggyBac MKOS factors have not identified the activation of preimplantation-specific genes as the critical features of iPSC-forming transcriptome changes (Samavarchi-Tehrani *et al.*, *Cell Stem Cell* 7:64-77, 2010). For example, most of the 231 critical genes identified were not activated transiently during the early phase of MKOS-based iPSC formation. These data indicate that Zscan4-mediated iPSC formation takes a different path from standard MKOS-mediated iPSC formation: the former path is more strongly associated with the genetic program occurring in the preimplantation embryos than the latter path.

[0207] To further investigate the action of Zscan4, the effect of Zscan4 alone on the MEFs was examined using the MEF-ZERT cells. Cell growth curves showed that Tmx slowed down the proliferation of MEF-ZERT significantly, but had only a minor effect on wild-type MEF (MEF-WT) (FIG. 4A). Evidently, although Zscan4 is not usually expressed there, it can function in MEFs and negatively affect their proliferation. Zscan4 alone, however, did not convert MEFs into iPSCs. In fact, the microarray analysis showed that a short-term treatment of Tmx had almost no effect on the transcriptome of both MEF-ZERT and MEF-WT (FIG. 4B). Next, DNA methylation patterns for Oct4 promoters was examined by the bisulfite sequencing method. Consistent with previous reports, the established iPSC, similar to ES cells, showed almost no DNA methylation (FIG. 4C). Interestingly, Tmx-treatment decreased DNA methylation levels from ~60% to ~44% by day 2, indicating a partial but rapid demethylation of DNAs by Zscan4 in MEFs. These data suggest that the expression of Zscan4 quickly alters the epigenetic status of the genome (*e.g.*, open chromatin conformation) by a mechanism that is as yet unidentified, and renders MEF susceptible to the effects of the forced expression of KOS factors (FIG. 4D).

[0208] Several lines of evidence indicate that Zscan4 is the missing early factor that initiates the cellular reprogramming (FIG. 3E). First, Zscan4 is a gene expressed specifically in late 2-cell embryos, blastomeres of which has been shown to have a potent reprogramming activity in the NT cloning experiments (Egli *et al.*, *Curr Biol* 19:1403-1409, 2009). Second, Zscan4-mediated reprogramming activates other preimplantation-specific genes in MEFs, seemingly recapitulating the genetic program occurring transiently during preimplantation embryo development and NT

embryo development. Third, Zscan4 alone can induce the rapid DNA demethylation in MEFs. This is in accord with the requirement of DNA demethylation for reprogramming in iPSC (Mikkelsen *et al.*, *Nature* 454:49-55, 2008), NT (Simonsson and Gurdon, *Nat Cell Biol* 6:984-990, 2004), and heterokaryons (Bhutani *et al.*, *Nature* 463:1042-1047, 2010). Fourth, Zscan4 replaces Myc in enhancing the efficiency of iPSC formation with KOS factors. Unlike Myc, which is required for at least the first 5 days of iPSC formation (Sridharan *et al.*, *Cell* 136:364-377, 2009), Zscan4 is required for only the initial day of iPSC formation – long before the expression of core pluripotency transcription factors such as Nanog. In addition, unlike Myc and other iPSC-promoting conditions (*e.g.*, repression of P53; Tapia and Scholer, *J Exp Med* 207:2045-2048, 2010), which stimulate the proliferation of MEFs, destabilize genome integrity (Hanna *et al.*, *Cell* 143:508-525, 2010; Stadtfeld and Hochedlinger, *Genes Dev* 24:2239-2263, 2010; Nakagawa *et al.*, *Proc Natl Acad Sci USA* 107:14152-14157, 2010), and raise concerns about long-term stability of iPSCs in culture (Hu *et al.*, *Proc Natl Acad Sci USA* 107:4335-4340, 2010; Feng *et al.*, *Stem Cells* 28:704-712, 2010), Zscan4 represses the proliferation of MEFs and promotes the genome stability and maintenance of normal karyotype in ES cells (Zalzman *et al.*, *Nature* 464:858-863, 2010). Therefore, these results indicate that Zscan4 is an "initiating" factor with distinct properties that can help to reprogram cells while preserving genome stability.

Example 2: iPSCs generated with Zscan4 are of high quality

[0209] Several studies were carried out to evaluate the quality of the iPSCs generated by expression of Zscan4, including a tetraploid complementation assay, which is the most stringent test for the pluripotency of iPS cells. A karyotype analysis of iPSCs generated with or without forced expression of Zscan4 was also performed. The results are shown in FIG. 18. Karyotype analysis of randomly selected iPSC lines clearly showed that iPSCs generated with Zscan4 were of higher quality than iPSCs generated without Zscan4. In addition, iPSCs generated with Zscan4 could form entire live embryos by the tetraploid complementation assay. The success rate for iPSCs in the tetraploid complementation assay is usually much lower than what was achieved by expression of Zscan4. Furthermore, this high success rate was achieved with the standard fetal calf serum-based culture condition, whereas the reported success has been achieved by using the special cell culture condition (KSR medium), which is known to dramatically enhance the pluripotency of ES/iPSC cells (see for example, Li *et al.*, *Cell Res* 21(3):550-553, 2011).

Example 3: Zscan4-dependent genes increase the efficiency of iPSC formation

[0210] As discussed in Example 1, 231 genes were identified that are upregulated during the early phase (day 1 – day 6) of iPSC formation in a Zscan4-dependent manner. Since many of these genes are preimplantation- or germline-specifically expressed, it was hypothesized that these genes would also enhance the efficiency of iPSC formation.

[0211] To test this notion, a piggyBac vector containing MKOS factors (Myc, Klf4, Oct4, and Sox2) and a piggyBac vector containing either DsRed (Control), Zscan4c (positive control), Patl2, Pramel6, Piwil2, D5Ertd577e, or D13Ertd608e, were transfected into wild-type mouse embryo fibroblast (MEF-WT) cells (FIG. 21). The efficiency of iPSC formation was scored based on their authentic ES cell morphology and alkaline phosphatase (ALP) staining 14 days after transfection (mean $\pm$S.E.M.) (FIGS. 22A-22C).

[0212] The results showed that Patl2, Pramel6, Piwil2, and D5Ertd577e increase the efficiency of iPSC formation. These effects were very similar to what was observed with Zscan4.

[0213] In view of the many possible embodiments to which the principles of the disclosed invention may be applied, it should be recognized that the illustrated embodiments are only examples of the disclosure and should not be taken as limiting the scope of the disclosure. Rather, the scope of the disclosure is defined by the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

CLAIMS

1. An ex vivo method of producing an induced pluripotent stem (iPS) cell by reprogramming of a somatic cell, comprising introducing into the somatic cell (1) a Zscan4 or a Zscan4-dependent gene selected from Patl2, Pramel6, Piwil2 and D5Ertd577e, and (2) at least three reprogramming factors, or at least four reprogramming factors selected from c-Myc, Klf4, Oct4, Sox2, Lin28 and Nanog, thereby producing an iPS cell, wherein the Zscan4 or the Zscan4-dependent gene is introduced into the somatic cell with or before introducing at least one of the reprogramming factors into the somatic cell.

2. The method of claim 1, comprising introducing into the somatic cell the at least four reprogramming factors.

3. The method of any one of claims 1-3, wherein three or four reprogramming factors are introduced into the somatic cell, and wherein:

(i) the four reprogramming factors are c-Myc, Klf4, Oct4 and Sox2;

(ii) the four reprogramming factors are Lin28, Nanog, Oct4 and Sox2; or

(iii) the three reprogramming factors are Klf4, Oct4 and Sox2

4. The method of any one of claims 1-3, wherein introducing into the somatic cell a Zscan4 or a Zscan4-dependent gene comprises introducing a nucleic acid molecule encoding a Zscan4 protein or a Zscan4-dependent gene protein into the somatic cell and wherein the nucleic acid molecule is expressed in the somatic cell.

5. The method of any one of claims 1-4, wherein introducing into the somatic cell at least three reprogramming factors or at least four reprogramming factors comprises introducing a nucleic acid molecule encoding at least one reprogramming factor protein into the somatic cell and wherein the nucleic acid molecule is expressed in the somatic cell.

6. The method of claim 4 or claim 5, wherein the nucleic acid molecule comprises a viral vector.

7. The method of claim 6, wherein the viral vector is a retrovirus vector, a lentivirus vector or an adenovirus vector.

8. The method of claim 4 or claim 5, wherein the nucleic acid molecule comprises a plasmid vector.

9. The method of claim 4 or claim 5, wherein the nucleic acid molecule is encapsulated in a nanoparticle.

10. The method of claim 4 or claim 5, wherein the nucleic acid molecule comprises mRNA encoding the Zscan4 protein or the Zscan4-dependent gene protein, or mRNA encoding the reprogramming factor protein.

11. The method of any one of claims 1-3, wherein introducing into the somatic cell a Zscan4 or a Zscan4-dependent gene comprises introducing a Zscan4 protein or a Zscan4-dependent gene protein into the somatic cell.

12. The method of any one of claims 1-3 and 11, wherein introducing into the somatic cell at least three reprogramming factors or at least four reprogramming factors comprises introducing a reprogramming factor protein into the somatic cell.

13. The method of claim 11 or claim 12, wherein the Zscan4 protein or Zscan4-dependent protein, or reprogramming factor protein, is encapsulated in a nanoparticle.

14. The method of claim 12 or claim 13, wherein the Zscan4 protein or Zscan4-dependent gene protein, or reprogramming factor protein, is fused to a cell-penetrating peptide.

15. The method of claim 14, wherein the cell-penetrating peptide comprises a protein transduction domain.

16. The method of claim 14, wherein the cell-penetrating peptide comprises a poly-arginine peptide tag.

17. The method of any one of claims 1-16, wherein the somatic cell is a murine cell.

18. The method of any one of claims 1-16, wherein the somatic cell is a human cell.

19. The method of any one of claims 1-18, wherein the somatic cell is a tissue stem cell, a progenitor cell or a differentiated cell.

20. The method of claim 19, wherein the tissue stem cell is a neural stem cell, a hematopoietic stem cell, a mesenchymal stem cell or an adipose stem cell.

21. The method of claim 19, wherein the differentiated cell is a fibroblast, lymphocyte, hepatocyte, epithelial cell, muscle cell, adipose cell, cardiomyocyte, pancreatic β cell, keratinocyte, amniotic cell, peripheral blood cell, platelet, or astrocyte.

22. The method of claim 21, wherein the lymphocyte is a T cell or B cell.

23. The method of any one of claims 1-22, comprising introducing into the somatic cell a Zscan4 and at least three reprogramming factors or at least four reprogramming factors.

24. The method of claim 23, wherein the Zscan4 is murine Zscan4.

25. The method of claim 24, wherein murine Zscan4 is Zscan4c.

26. The method of claim 25, wherein the Zscan4c is functional and comprises an amino acid sequence that is at least 95% identical to SEQ ID NO: 8.

27. The method of claim 26, wherein the Zscan4c amino acid sequence comprises SEQ ID NO: 8.

28. The method of claim 25, wherein the Zscan4c is functional and is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 7 and wherein the nucleotide sequence is expressed in the somatic cell.

29. The method of claim 28, wherein Zscan4c is encoded by a nucleotide sequence comprising SEQ ID NO: 7.

30. The method of claim 23, wherein the Zscan4 is human ZSCAN4.

31. The method of claim 30, wherein the ZSCAN4 is functional and comprises an amino acid sequence that is at least 95% identical to SEQ ID NO: 2.

32. The method of claim 30, wherein the ZSCAN4 amino acid sequence comprises SEQ ID NO: 2.

33. The method of claim 30, wherein the ZSCAN4 is functional and is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 1 and wherein the nucleotide sequence is expressed in the somatic cell.

34. The method of claim 30, wherein ZSCAN4 is encoded by a nucleotide sequence comprising SEQ ID NO: 1 and wherein the nucleotide sequence is expressed in the somatic cell.

35. The method of any one of claims 1-22, comprising introducing into the somatic cell a Patl2 and at least three reprogramming factors or at least four reprogramming factors.

36. The method of claim 35, wherein the Patl2 is mouse Patl2.

37. The method of claim 36, wherein the Patl2 is functional and comprises an amino acid sequence that is at least 95% identical to SEQ ID NO: 39.

38. The method of claim 36, wherein the Patl2 amino acid sequence comprises SEQ ID NO: 39.

39. The method of claim 36, wherein the Patl2 is functional and is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 38 and wherein the nucleotide sequence is expressed in the somatic cell.

40. The method of claim 36, wherein Patl2 is encoded by a nucleotide sequence comprising SEQ ID NO: 38 and wherein the nucleotide sequence is expressed in the somatic cell.

41. The method of claim 35, wherein the Patl2 is human Patl2.

42. The method of claim 41, wherein the Patl2 is functional and comprises an amino acid sequence that is at least 95% identical to SEQ ID NO: 47.

43. The method of claim 41, wherein the Patl2 amino acid sequence comprises SEQ ID NO: 47.

44. The method of claim 41, wherein the Patl2 is functional and is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 46 and wherein the nucleotide sequence is expressed in the somatic cell.

45. The method of claim 41, wherein Patl2 is encoded by a nucleotide sequence comprising SEQ ID NO: 46 and wherein the nucleotide sequence is expressed in the somatic cell.

46. The method of any one of claims 1-22, comprising introducing into the somatic cell a Pramel6 and at least three reprogramming factors or at least four reprogramming factors.

47. The method of claim 46, wherein the Pramel6 is functional and comprises an amino acid sequence that is at least 95% identical to SEQ ID NO: 41.

48. The method of claim 46, wherein the Pramel6 amino acid sequence comprises SEQ ID NO: 41.

49. The method of claim 46, wherein the Pramel6 is functional and is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 40 and wherein the nucleotide sequence is expressed in the somatic cell.

50. The method of claim 46, wherein Pramel6 is encoded by a nucleotide sequence comprising SEQ ID NO: 40 and wherein the nucleotide sequence is expressed in the somatic cell.

51. The method of any one of claims 1-22, comprising introducing into the somatic cell a Piwil2 and at least three reprogramming factors or at least four reprogramming factors.

52. The method of claim 51, wherein the Piwil2 is mouse Piwil2.

53. The method of claim 52, wherein the Piwil2 is functional and comprises an amino acid sequence that is at least 95% identical to SEQ ID NO: 43.

54. The method of claim 52, wherein the Piwil2 amino acid sequence comprises SEQ ID NO: 43.

55. The method of claim 52, wherein the Piwil2 is functional and is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 42 and wherein the nucleotide sequence is expressed in the somatic cell.

56. The method of claim 52, wherein Piwil2 is encoded by a nucleotide sequence comprising SEQ ID NO: 42 and wherein the nucleotide sequence is expressed in the somatic cell.

57. The method of claim 51, wherein the Piwil2 is human Piwil2.

58. The method of claim 57, wherein the Piwil2 is functional and comprises an amino acid sequence that is at least 95% identical to SEQ ID NO: 49.

59. The method of claim 57, wherein the Piwil2 amino acid sequence comprises SEQ ID NO: 49.

60. The method of claim 57, wherein the Piwil2 is functional and is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 48 and wherein the nucleotide sequence is expressed in the somatic cell.

61. The method of claim 57, wherein Piwil2 is encoded by a nucleotide sequence comprising SEQ ID NO: 48 and wherein the nucleotide sequence is expressed in the somatic cell.

62. The method of any one of claims 1-22, comprising introducing into the somatic cell a D5Ertd577e and at least three reprogramming factors or at least four reprogramming factors.

63. The method of claim 62, wherein the D5Ertd577e is functional and comprises an amino acid sequence that is at least 95% identical to SEQ ID NO: 45.

64. The method of claim 62, wherein the D5Ertd577e amino acid sequence comprises SEQ ID NO: 45.

65. The method of claim 62, wherein the D5Ertd577e is functional and is encoded by a nucleotide sequence at least 95% identical to SEQ ID NO: 44 and wherein the nucleotide sequence is expressed in the somatic cell.

66. The method of claim 62, wherein D5Ertd577e is encoded by a nucleotide sequence comprising SEQ ID NO: 44 and wherein the nucleotide sequence is expressed in the somatic cell.

67. The method of any one of claims 1-22, comprising introducing into the somatic cell (1) the Zscan4 or the Zscan4-dependent gene selected from Patl2, Pramel6, Piwil2 and D5Ertd577e, and (2) the at least four reprogramming factors.

68. An isolated iPS cell produced according to the method of any one of claims 1-67, wherein the iPS cell exhibits enhanced pluripotency as measured by a tetraploid complementation assay.

FIGURE 1

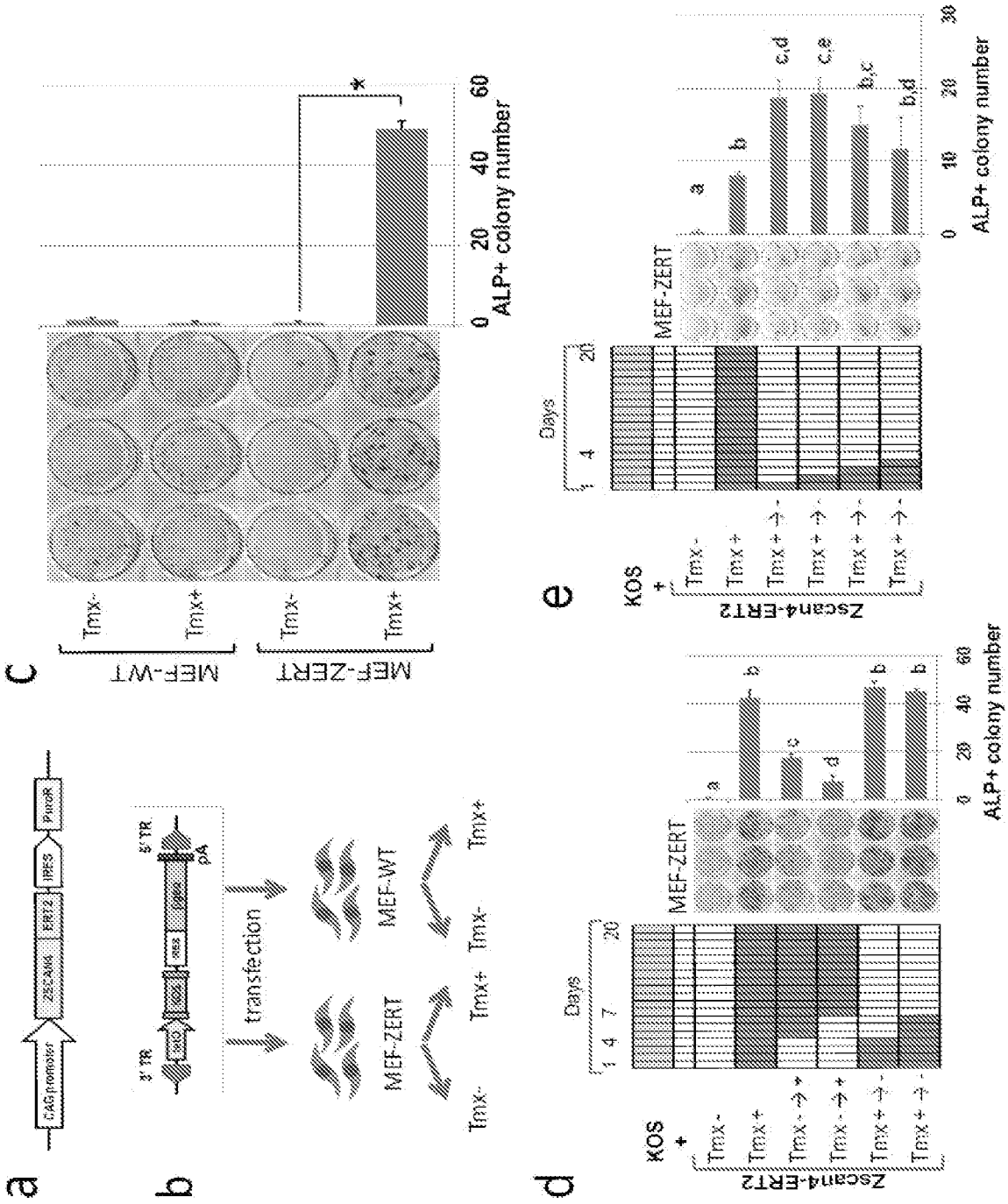

a
CAG promoter
Zscan4
ERT2
IRES
PuroR
b
5' TR
Ψ
Zscan4
IRES
rtTA
pA
3' TR
transfection
MEF-ZERT
MEF-WT
Tmx-
Tmx+
Tmx+
Tmx-
c
MEF-WT
MEF-ZERT
Tmx-
Tmx+
Tmx-
Tmx+
*
0
20
40
60
ALP+ colony number
d
KOS
+
Zscan4-ERT2
Tmx-
Tmx+
Tmx- → +
Tmx + → -
Tmx + → -
Tmx + → -
Days
1 4 7 20
MEF-ZERT
a
b
c
d
b
b
b
0
20
40
60
ALP+ colony number
e
KOS
+
Zscan4-ERT2
Tmx-
Tmx+
Tmx + → -
Tmx + → -
Tmx + → -
Tmx + → -
Days
1 4 20
MEF-ZERT
a
b
c,d
c,e
b,c
b,d
0
10
20
30
ALP+ colony number

FIGURE 2

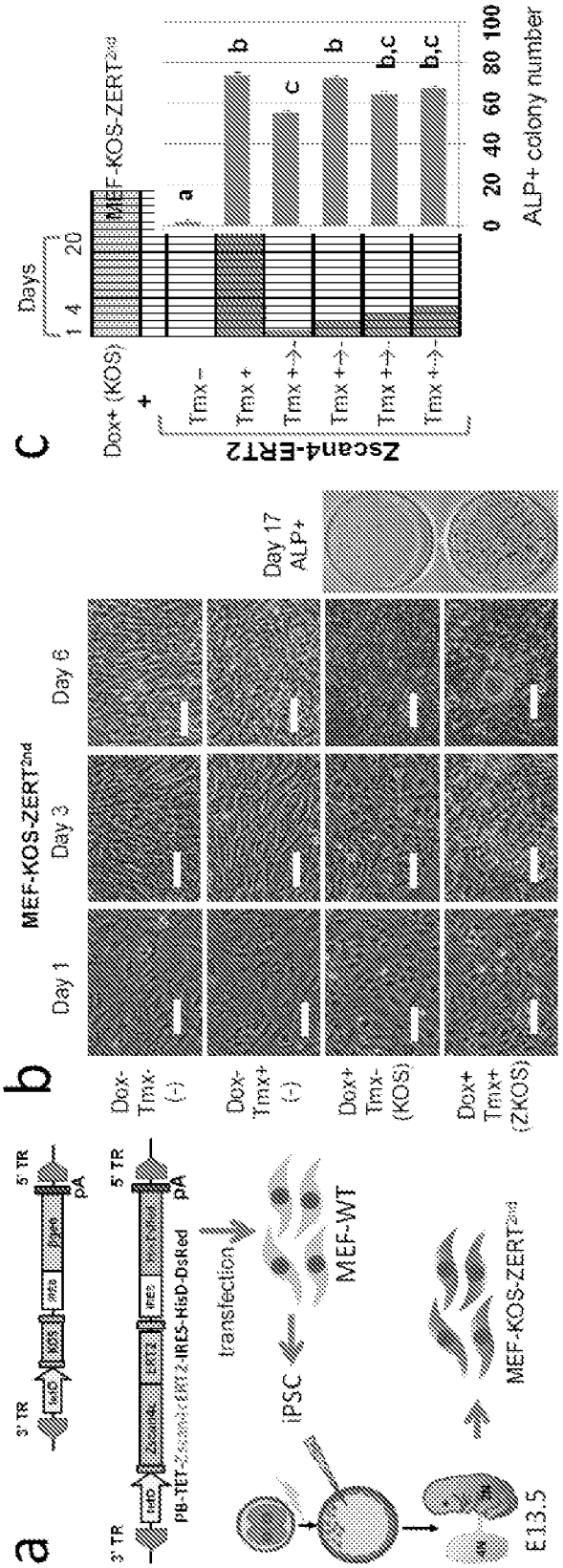

a
5' TR
pA
3' TR
transfection
iPSC
MEF-WT
MEF-KOS-ZERT2nd
E13.5
b
MEF-KOS-ZERT2nd
Day 1
Day 3
Day 6
Day 17
ALP+
Dox-
Tmx-
(-)
Dox-
Tmx+
(-)
Dox+
Tmx-
(KOS)
Dox+
Tmx+
(ZKOS)
c
Zscan4-ERT2
Dox+ (KOS)
+
Tmx -
Tmx +
Tmx +→
Tmx +→
Tmx +→
Tmx +→
Days
1 4 20
MEF-KOS-ZERT2nd
a
b
c
b
b,c
b,c
0 20 40 60 80 100
ALP+ colony number

FIGURE 3A-B

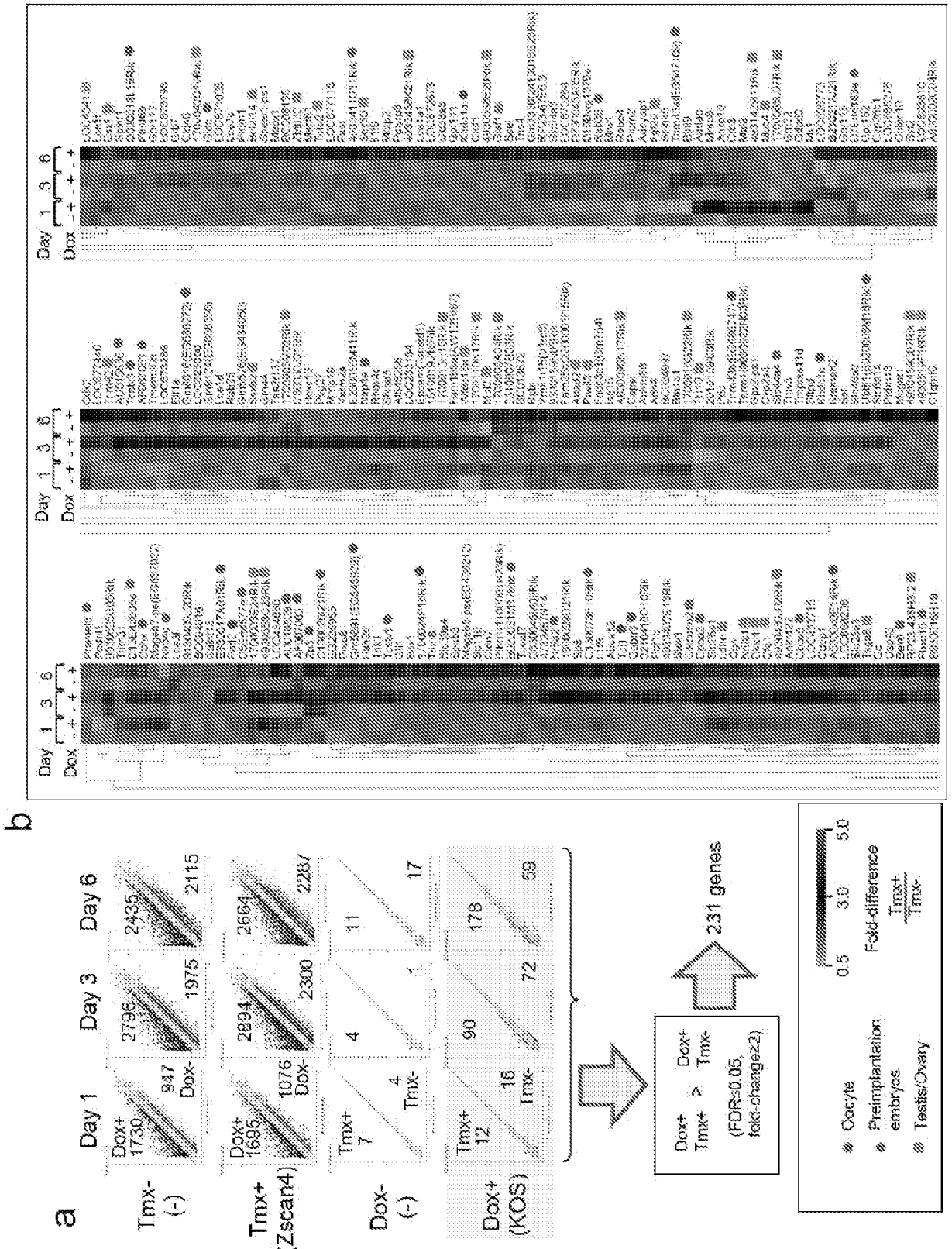

a
Day 1
Day 3
Day 6
Tmx- (-)
Tmx+ (Zscan4)
Dox- (-)
Dox+ (KOS)
Dox+ 1730
947 Dox-
2796
1975
2435
2115
Dox+ 1695
1076 Dox-
2894
2300
2664
2287
Tmx+ 7
4 Tmx-
4
1
11
17
Tmx+ 12
16 Tmx-
90
72
178
69
Dox+ Tmx+ > Dox+ Tmx-
(FDR≤0.05, fold-change≥2)
231 genes
Oocyte
Preimplantation embryos
Testis/Ovary
0.5
3.0
5.0
Fold-difference
Tmx+
Tmx-
b
Day
Dox

FIGURE 3C-E

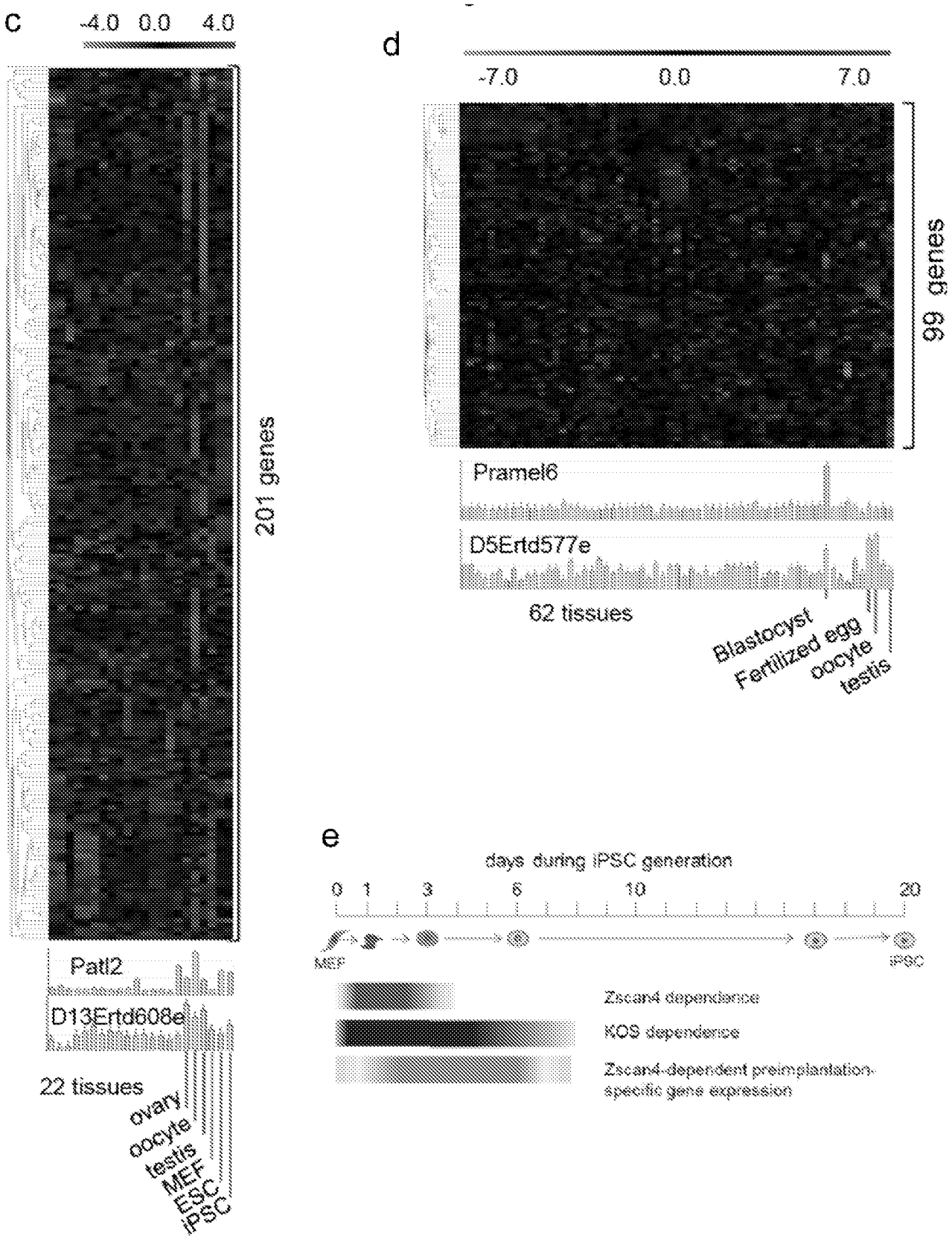

c
-4.0 0.0 4.0
201 genes
Patl2
D13Ertd608e
22 tissues
ovary
oocyte
testis
MEF
ESC
iPSC
d
-7.0 0.0 7.0
99 genes
Pramel6
D5Ertd577e
62 tissues
Blastocyst
Fertilized egg
oocyte
testis
e
days during iPSC generation
0 1 3 6 10 20
MEF
iPSC
Zscan4 dependence
KOS dependence
Zscan4-dependent preimplantation-specific gene expression

FIGURE 4

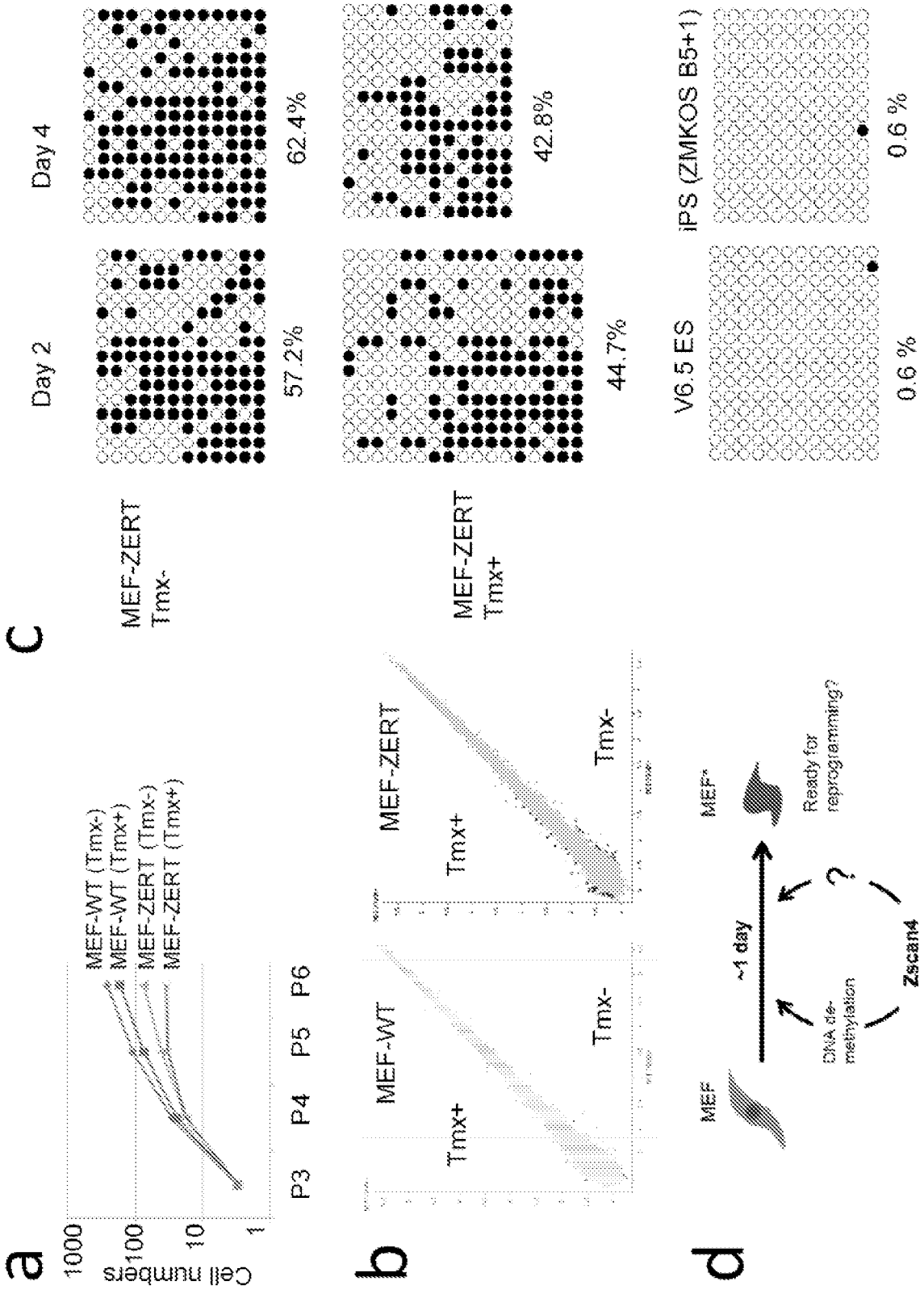

a
Cell numbers
1000
100
10
1
MEF-WT (Tmx-)
MEF-WT (Tmx+)
MEF-ZERT (Tmx-)
MEF-ZERT (Tmx+)
P3 P4 P5 P6
b
MEF-WT
Tmx+
Tmx-
MEF-ZERT
Tmx+
Tmx-
c
Day 2
Day 4
MEF-ZERT Tmx-
57.2%
62.4%
MEF-ZERT Tmx+
44.7%
42.8%
V6.5 ES
0.6 %
0.6 %
iPS (ZMKOS B5+1)
d
MEF
~1 day
MEF*
Ready for reprogramming?
DNA de-methylation
Zscan4

FIGURE 5

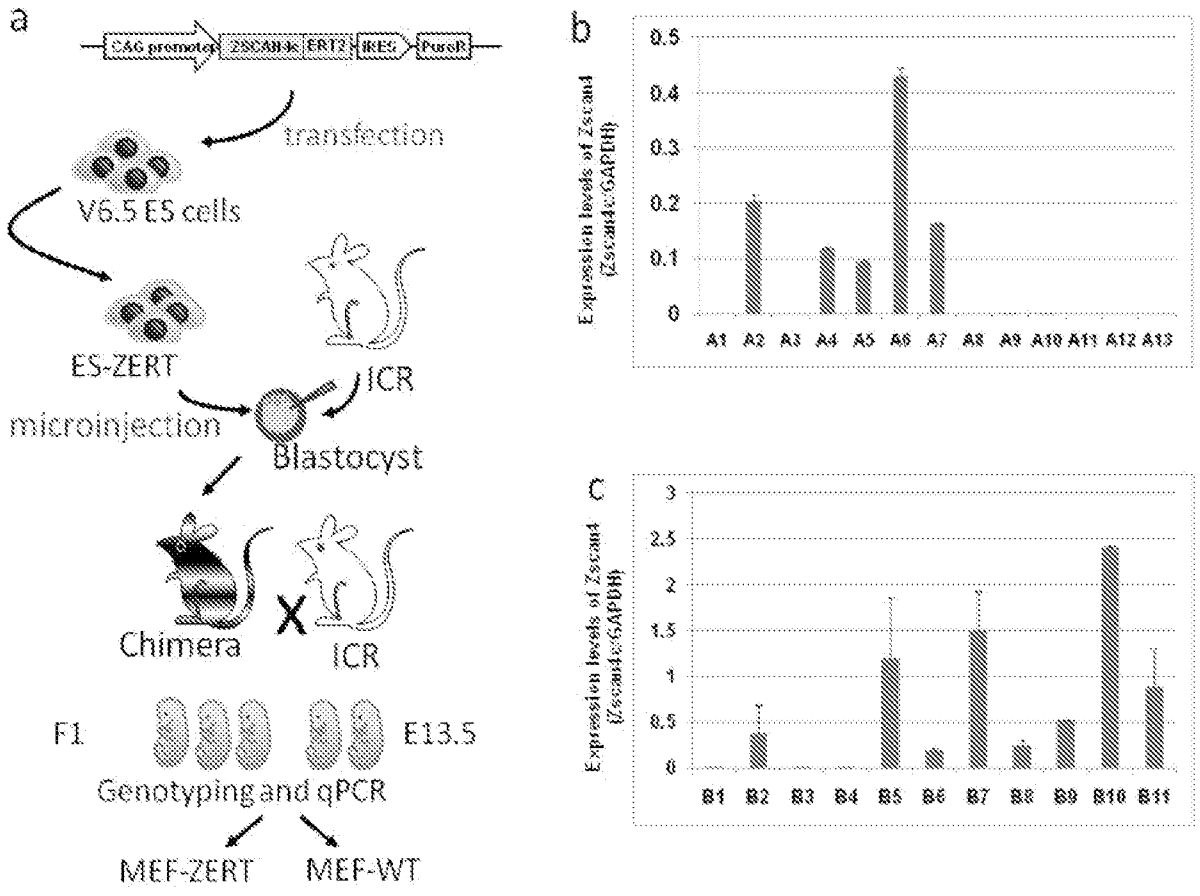

a
CAG promoter
ZSCAN4c
ERT2
IRES
PuroR
transfection
V6.5 ES cells
ES-ZERT
ICR
microinjection
Blastocyst
Chimera
X
ICR
F1
E13.5
Genotyping and qPCR
MEF-ZERT
MEF-WT
b
Expression levels of Zscan4
(Zscan4c:GAPDH)
0.5
0.4
0.3
0.2
0.1
0
A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13
c
Expression levels of Zscan4
(Zscan4c:GAPDH)
3
2.5
2
1.5
1
0.5
0
B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11

FIGURE 6

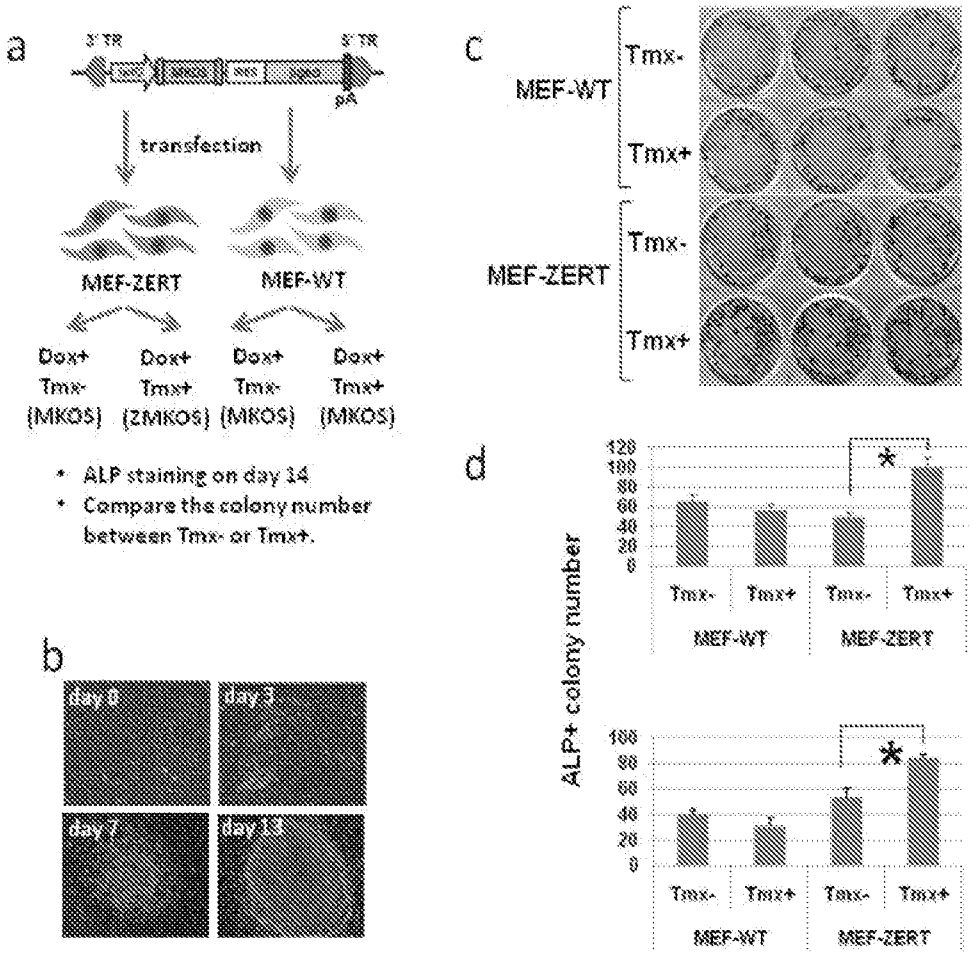

a
5' TR
3' TR
pA
transfection
MEF-ZERT
MEF-WT
Dox+ Tmx- (MKOS)
Dox+ Tmx+ (ZMKOS)
Dox+ Tmx- (MKOS)
Dox+ Tmx+ (MKOS)
• ALP staining on day 14
• Compare the colony number between Tmx- or Tmx+.
b
day 0
day 3
day 7
day 14
c
MEF-WT
Tmx-
Tmx+
MEF-ZERT
Tmx-
Tmx+
d
ALP+ colony number
120
100
80
60
40
20
0
Tmx-
Tmx+
Tmx-
Tmx+
MEF-WT
MEF-ZERT
*
160
80
60
40
20
0
Tmx-
Tmx+
Tmx-
Tmx+
MEF-WT
MEF-ZERT
*

FIGURE 7

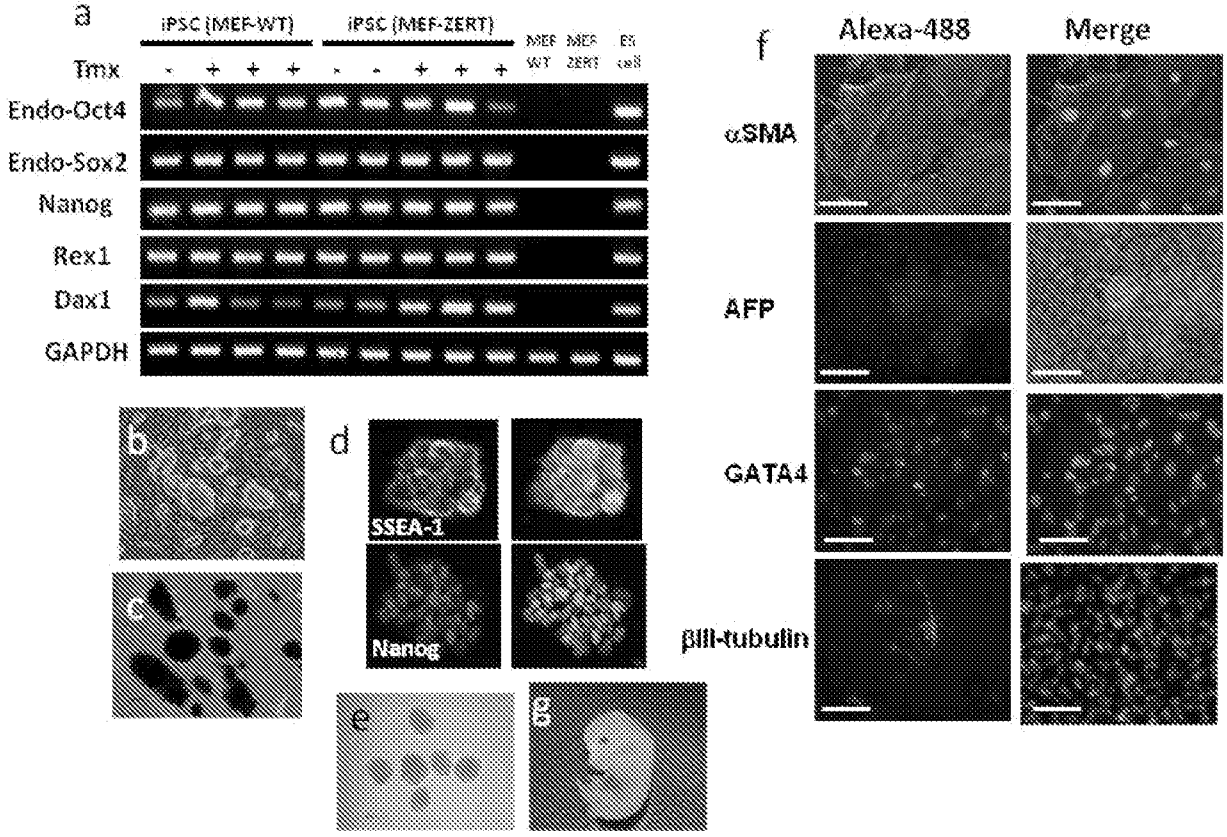

a
iPSC (MEF-WT)
iPSC (MEF-ZERT)
Tmx
Endo-Oct4
Endo-Sox2
Nanog
Rex1
Dax1
GAPDH
b
c
d
SSEA-1
Nanog
e
g
f
Alexa-488
Merge
αSMA
AFP
GATA4
βIII-tubulin

FIGURE 8

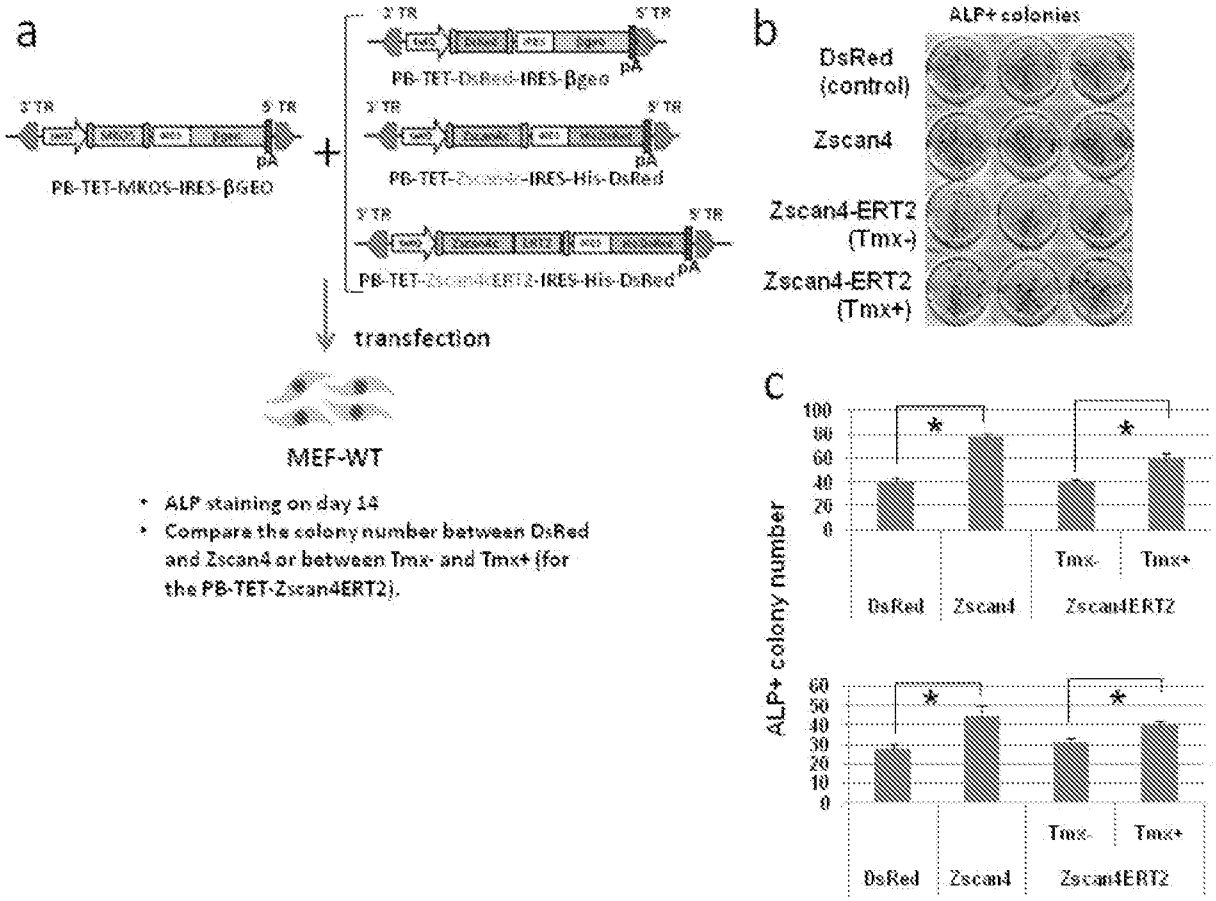

a
5' TR
3' TR
PB-TET-MKOS-IRES-βGEO
PB-TET-DsRed-IRES-βgeo
PB-TET-Zscan4-IRES-His-DsRed
PB-TET-Zscan4-ERT2-IRES-His-DsRed
pA
transfection
MEF-WT
ALP staining on day 14
Compare the colony number between DsRed and Zscan4 or between Tmx- and Tmx+ (for the PB-TET-Zscan4ERT2).
b
ALP+ colonies
DsRed (control)
Zscan4
Zscan4-ERT2 (Tmx-)
Zscan4-ERT2 (Tmx+)
c
ALP+ colony number
100
80
60
40
20
0
DsRed
Zscan4
Tmx-
Tmx+
Zscan4ERT2
60
50
40
30
20
10
0

FIGURE 9

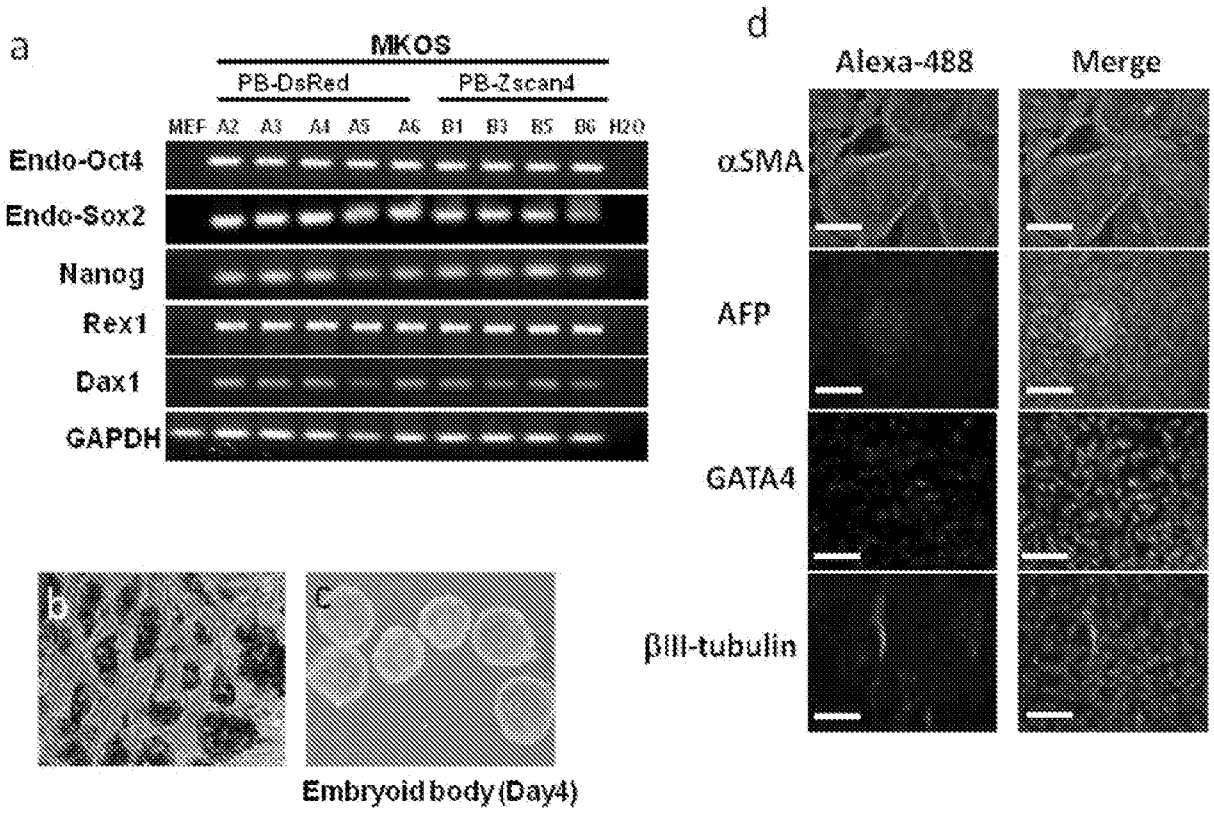

a
MKOS
PB-DsRed
PB-Zscan4
MEF A2 A3 A4 A5 A6 B1 B3 B5 B6 H2O
Endo-Oct4
Endo-Sox2
Nanog
Rex1
Dax1
GAPDH
b
c
Embryoid body (Day4)
d
Alexa-488
Merge
αSMA
AFP
GATA4
βIII-tubulin

FIGURE 10

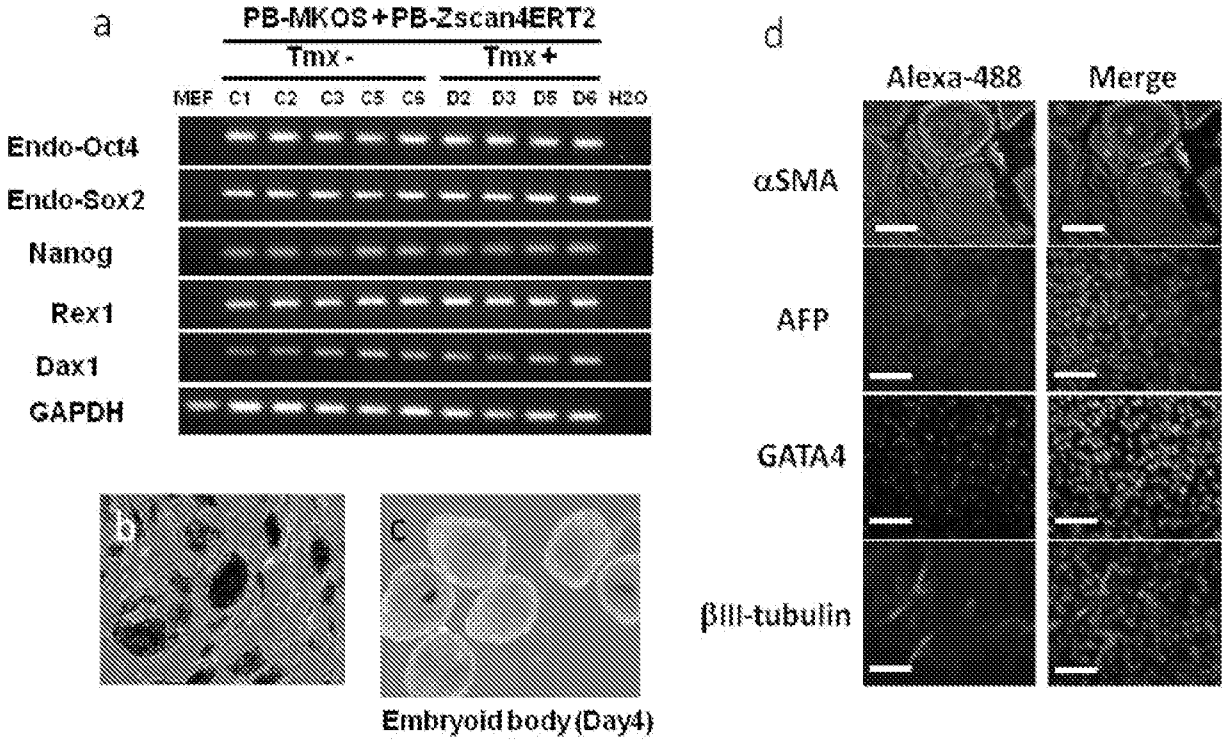

a
PB-MKOS+PB-Zscan4ERT2
Tmx -
Tmx +
MEF C1 C2 C3 C5 C6 D2 D3 D5 D6 H2O
Endo-Oct4
Endo-Sox2
Nanog
Rex1
Dax1
GAPDH
b
c
Embryoid body (Day4)
d
Alexa-488
Merge
αSMA
AFP
GATA4
βIII-tubulin

FIGURE 11

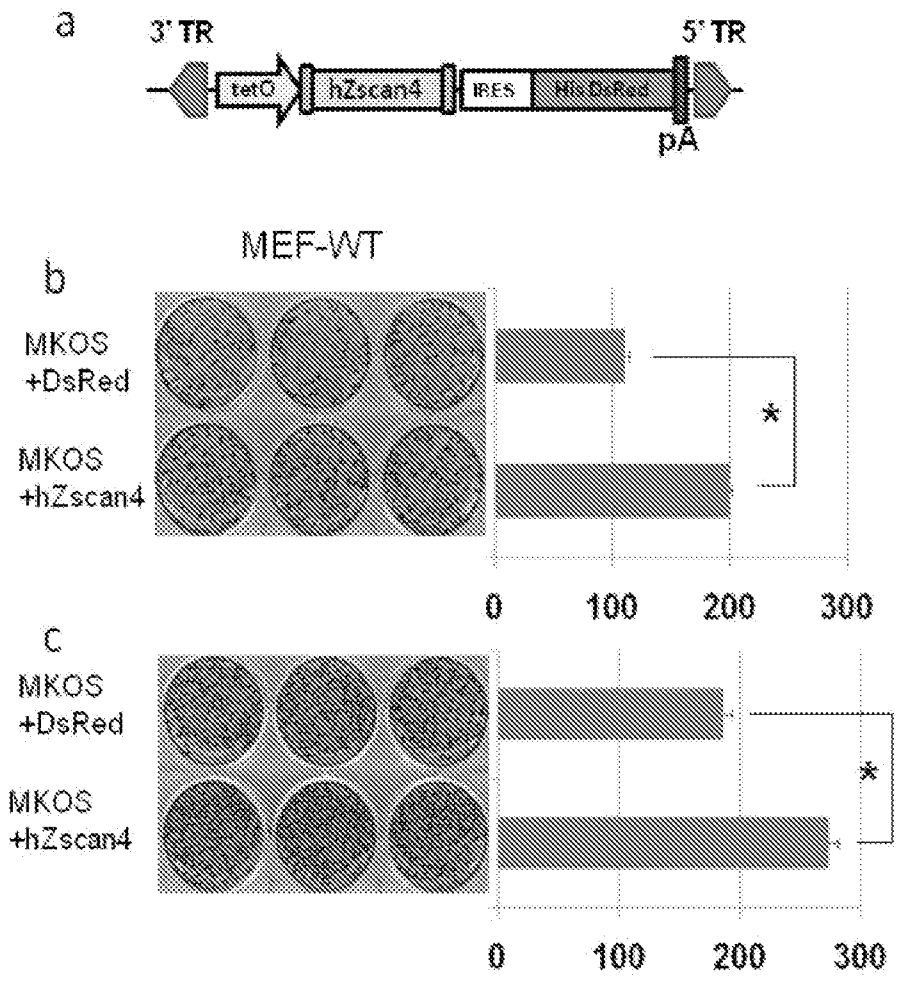

a
3' TR
5' TR
tetO
hZscan4
IRES
pA
MEF-WT
b
MKOS
+DsRed
MKOS
+hZscan4
*
0
100
200
300
c
MKOS
+DsRed
MKOS
+hZscan4
*
0
100
200
300

FIGURE 12

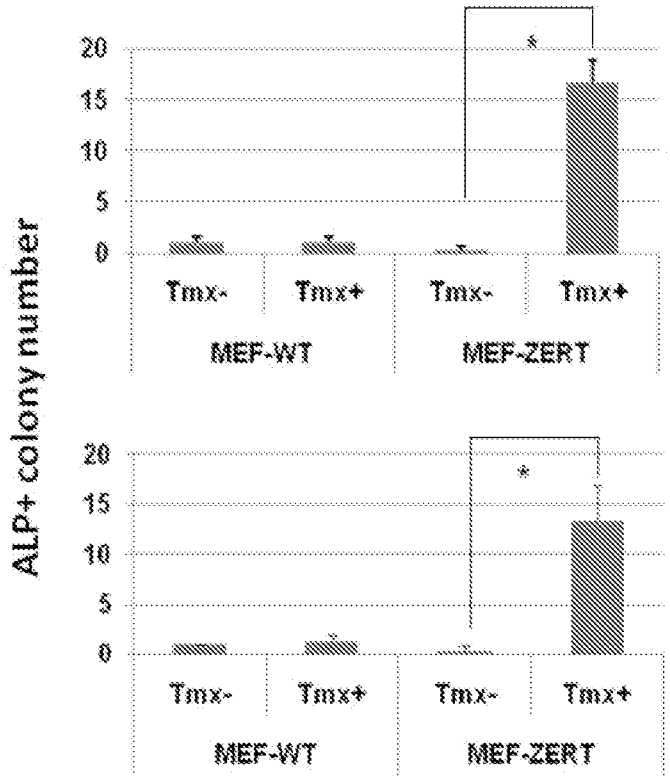

ALP+ colony number
20
15
10
5
0
*
Tmx-
Tmx+
Tmx-
Tmx+
MEF-WT
MEF-ZERT
20
15
10
5
0
*
Tmx-
Tmx+
Tmx-
Tmx+
MEF-WT
MEF-ZERT

FIGURE 13

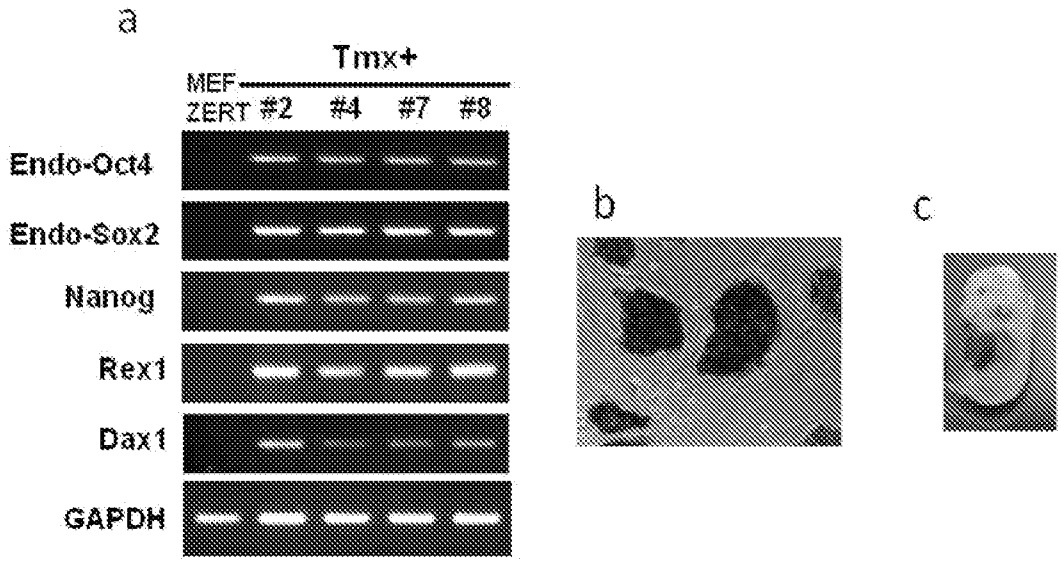

a
Tmx+
MEF
ZERT
#2
#4
#7
#8
Endo-Oct4
Endo-Sox2
Nanog
Rex1
Dax1
GAPDH
b
c

FIGURE 14

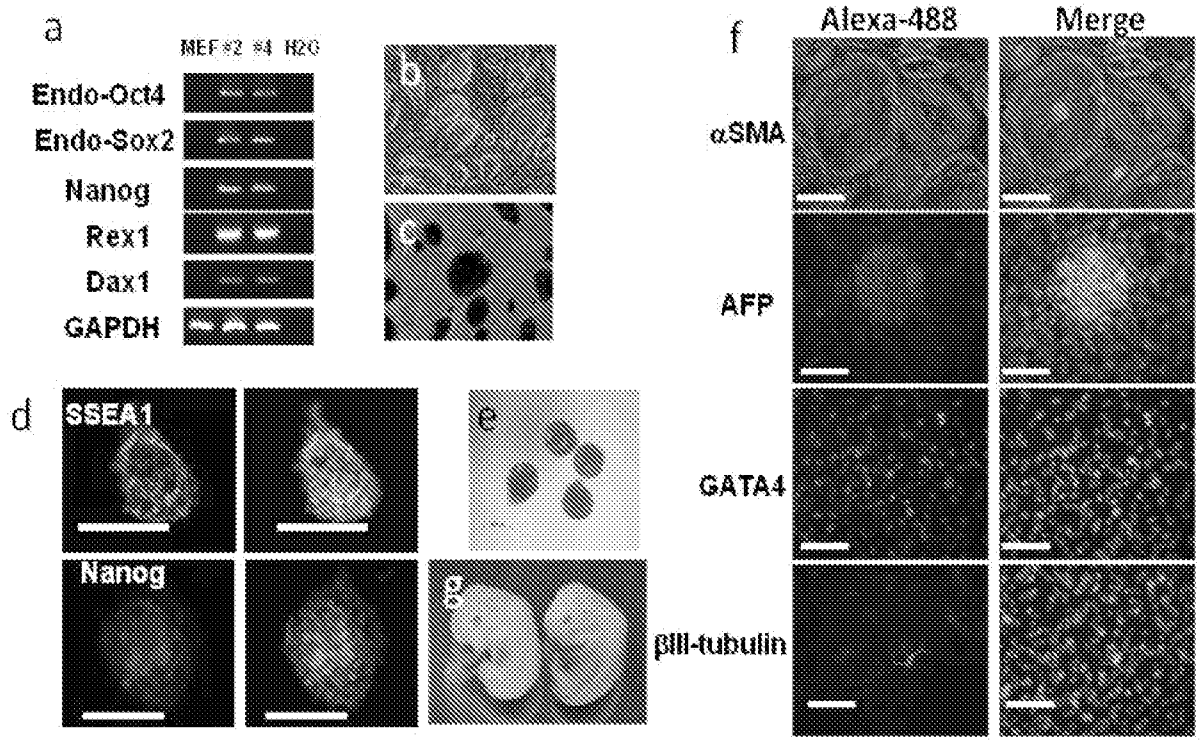

a
MEF #2 #4 H2O
Endo-Oct4
Endo-Sox2
Nanog
Rex1
Dax1
GAPDH
b
c
d
SSEA1
Nanog
e
g
f
Alexa-488
Merge
αSMA
AFP
GATA4
βIII-tubulin

FIGURE 15

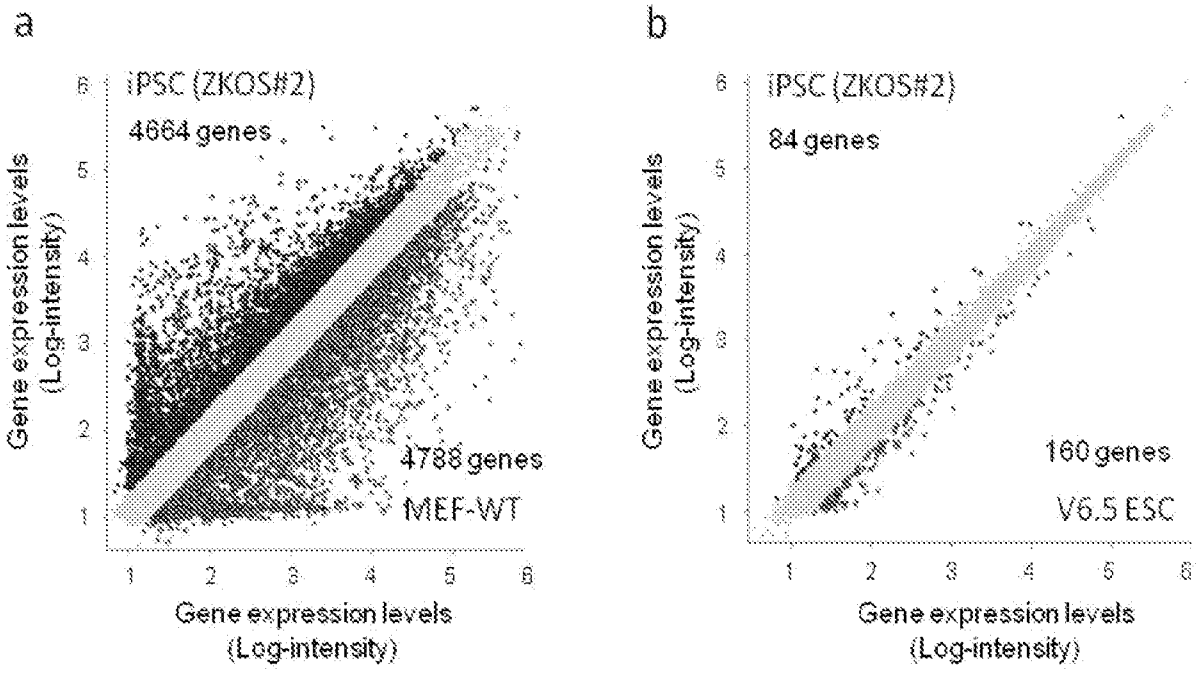

a
iPSC (ZKOS#2)
4664 genes
4788 genes
MEF-WT
Gene expression levels
(Log-intensity)
Gene expression levels
(Log-intensity)
b
iPSC (ZKOS#2)
84 genes
160 genes
V6.5 ESC
Gene expression levels
(Log-intensity)
Gene expression levels
(Log-intensity)

FIGURE 16

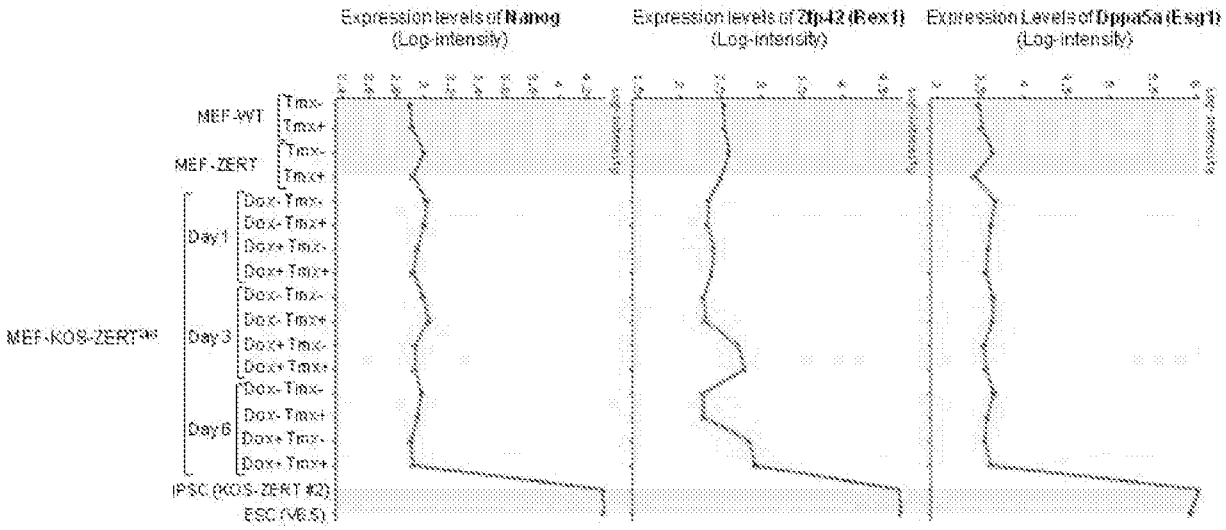

Expression levels of Nanog (Log-intensity)
Expression levels of Zfp42 (Rex1) (Log-intensity)
Expression Levels of Dppa5a (Esg1) (Log-intensity)
MEF-WT
MEF-ZERT
MEF-KOS-ZERT2R
Day 1
Day 3
Day 6
Dox- Tmx-
Dox- Tmx+
Dox+ Tmx-
Dox+ Tmx+
iPSC (KOS-ZERT #2)
ESC (V6.5)

FIGURE 17

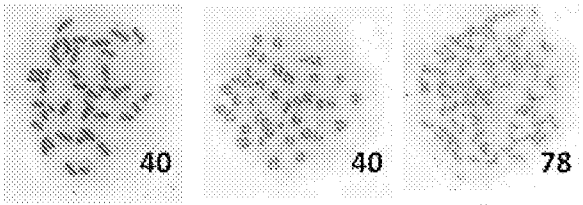
Purpose	*Primer Name*	*Primer Sequence*	*SEQ ID NO:*
For construction	attB1-koz-Zscan4c forward primer	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACCatggctcacagcaggcac	15
	attB2-Zscan4c reverse primer	GGGGACCACTTTGTACAAGAAAGCTGGGTtcagtcagaatctgtgtaat	16
	attB2-Zscan4c-ERT2 reverse primer	GGGGACCACTTTGTACAAGAAAGCTGGGTtcaagctgtggcagggaaac	17
	attB1-koz-KLF4 forward primer	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACCatgaggcagccaccgtgc	18
	attB2-SOX2 reverse primer	GGGGACCACTTTGTACAAGAAAGCTGGGTtcacatgtgcgacaggggca	19
	attB1-kozak-human Zscan4 forward primer	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACCatggcttag actaagaa	20
	attB2-human Zscan4 reverse primer	GGGGACCACTTTGTACAAGAAAGCTGGGTttaggaagttctgggtgtgg	21
For RT-PCR	endo-oct4 F	TCTTTCCACCAGGCCCCCGGCTC	22
	endo-oct4 R	TGCGGGCGGACATGGGGAGATCC	23
	endo-SOX2 F	TAGAGCTAGACTCCGGGCGATGA	24
	endo-SOX2 R	TTGCCTTAAACAAGACCACGAAA	25
	Nanog F	CACCCACCCATGCTAGTCTT	26
	Nanog R	ACCCTCAAACTCCTGGTCCT	27
	Dax1 F	TGCTGCGGTCCAGGCCATCAAGAG	28
	Dax1 R	GGGCACTGTTCAGTTCAGCGGATC	29
	Rex1 F	ACGAGTGGCAGTTTCTTCTTGGGA	30
	Rex1 R	TATGACTCACTTCCAGGGGGCACT	31
	GAPDH ON 025	CGGAGTCAACGGATTTGGTCGTAT	32
	GAPDH ON 032	GAAGATGGTGATGGGCTTCC	33
For RT-qPCR	Zscan4c_ON995	AGTCTGACTGATGAGTGCTTGAAGCC	34
	Zscan4c_ON996	GGCCTTGTTGCAGATTGCTGTTG	35
For methylation analysis	Me-Oct3/4-S	GGTTTTTTAGAGGATGGTTGAGTG	36
	Me-Oct3/4-AS	TCCAACCCTACTAACCCATCACC	37

FIGURE 18

A

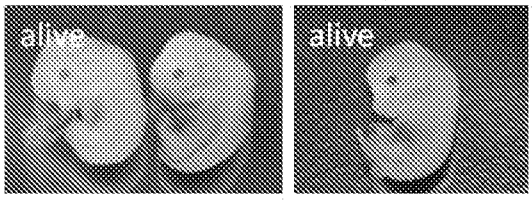
iPSC cell line	Zscan4 (Z)	Factors	Euploid (%)	No. injected 4N blastocysts	No. embryos at E13.5 (%)	No. live embryos at E13.5 (%)
MEF-ZERT-KOS #2	Z	KOS	80	45	9 (20%)	2 (4.4%)
B5+1	Z	MKOS	58	46	1 (2%)	1 (2%)
B5+2	Z	MKOS	76	25	0 (0%)	0 (0%)
B5+4	Z	MKOS	81	N.D.	N.D.	N.D.
B5-1	-	MKOS	47	N.D.	N.D.	N.D.
B5-3	-	MKOS	0	N.D.	N.D.	N.D.

B



KOS+ZERT #2 (Zscan4 +) B5+2 (Zscan4+) B5-3 (Zscan4-)

C



KOS+ZERT #2 B5+1

FIGURE 19

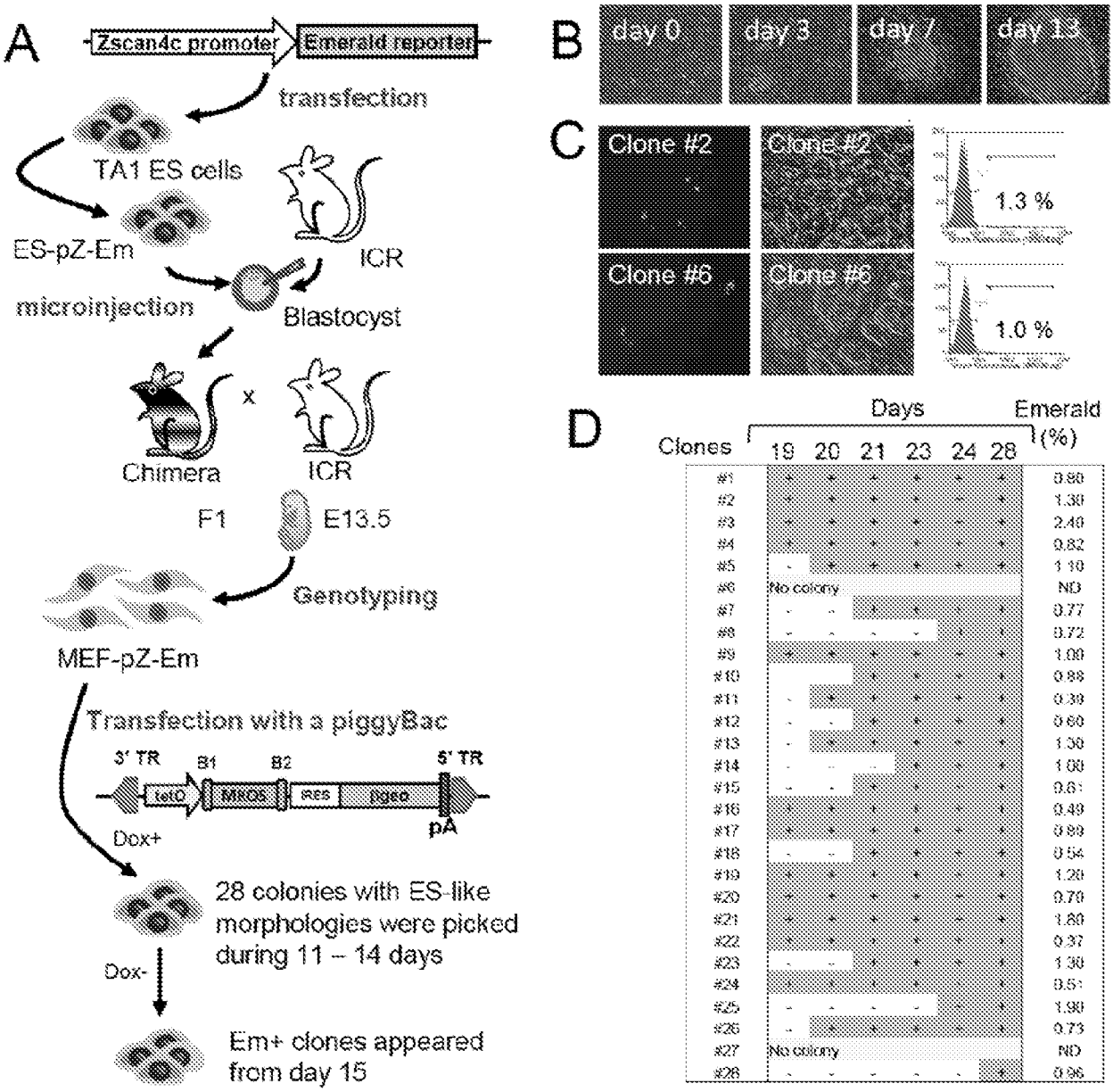

A
Zscan4c promoter
Emerald reporter
transfection
TA1 ES cells
ES-pZ-Em
ICR
microinjection
Blastocyst
Chimera
x
ICR
F1
E13.5
Genotyping
MEF-pZ-Em
Transfection with a piggyBac
5' TR
B1
B2
5' TR
tetO
MKOs
IRES
pgeo
pA
Dox+
28 colonies with ES-like morphologies were picked during 11 – 14 days
Dox-
Em+ clones appeared from day 15
B
day 0
day 3
day 7
day 13
C
Clone #2
Clone #2
1.3 %
Clone #6
Clone #6
1.0 %
D
Days
Emerald (%)
Clones
19 20 21 23 24 28
#1 + + + + + + 0.80
#2 + + + + + + 1.30
#3 + + + + + + 2.40
#4 + + + + + + 0.82
#5 - + + + + + 1.10
#6 No colony ND
#7 - - + + + + 0.77
#8 - - - - + + 0.72
#9 + + + + + + 1.00
#10 + + + + 0.88
#11 - + + + + + 0.30
#12 - - + + + + 0.60
#13 - + + + + + 1.30
#14 - - - + + + 1.00
#15 - - + + + + 0.61
#16 + + + + + + 0.49
#17 + + + + + + 0.89
#18 - - + + + + 0.54
#19 + + + + + + 1.20
#20 + + + + + + 0.70
#21 + + + + + + 1.80
#22 + + + + + + 0.37
#23 - - + + + + 1.30
#24 + + + + + + 0.51
#25 - - - - + + 1.90
#26 - + + + + + 0.73
#27 No colony ND
#28 - - - - - + 0.96

FIGURE 20

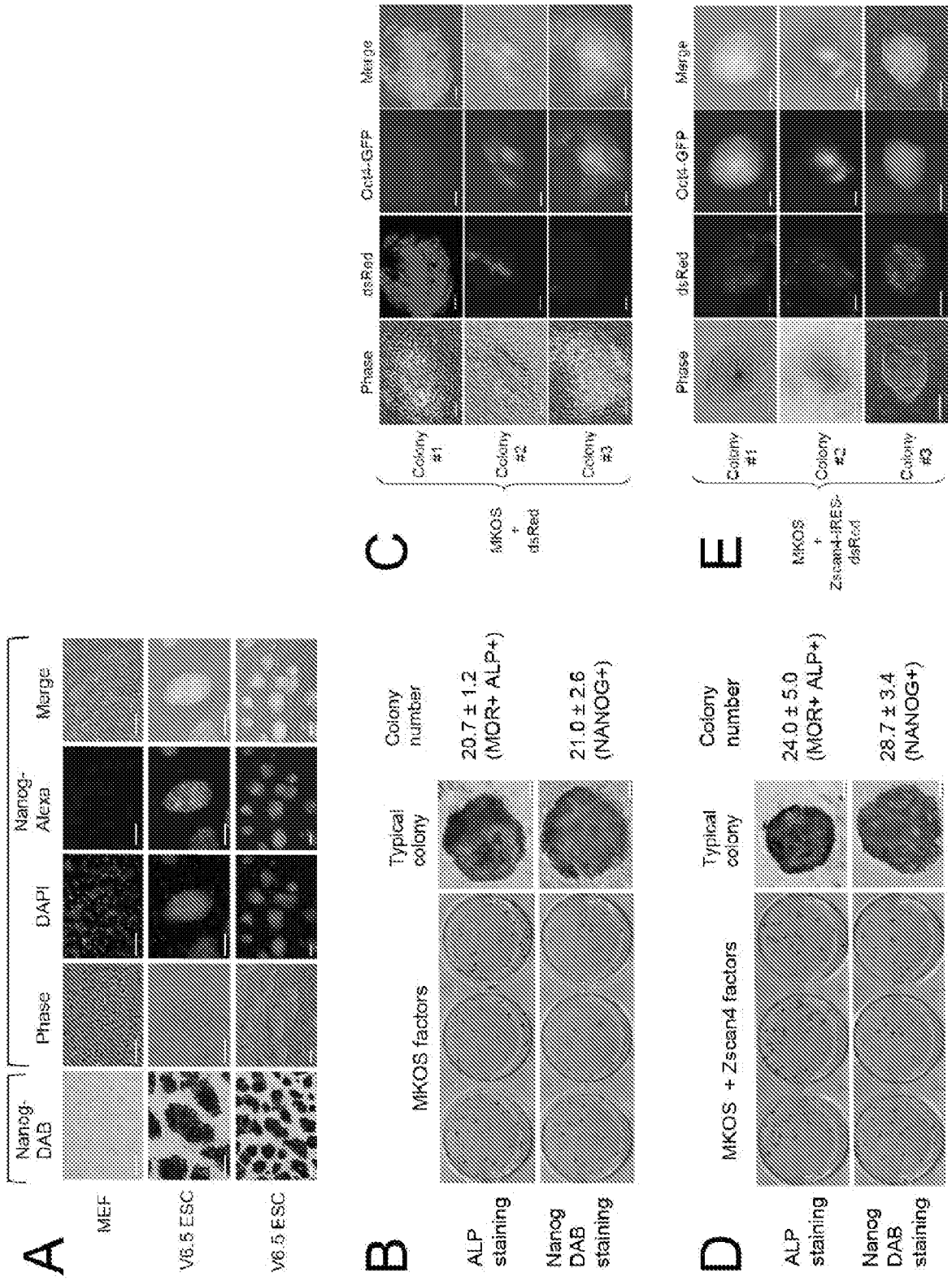

A
Nanog-DAB
Phase
DAPI
Nanog-Alexa
Merge
MEF
V6.5 ESC
V6.5 ESC
B
MKOS factors
Typical colony
Colony number
20.7 ± 1.2 (MOR+ ALP+)
21.0 ± 2.6 (NANOG+)
ALP staining
Nanog DAB staining
C
Phase
dsRed
Oct4-GFP
Merge
Colony #1
Colony #2
Colony #3
MKOS + dsRed
D
MKOS + Zscan4 factors
Typical colony
Colony number
24.0 ± 5.0 (MOR+ ALP+)
28.7 ± 3.4 (NANOG+)
ALP staining
Nanog DAB staining
E
Phase
dsRed
Oct4-GFP
Merge
Colony #1
Colony #2
Colony #3
MKOS + Zscan4-IRES-dsRed

FIGURE 21

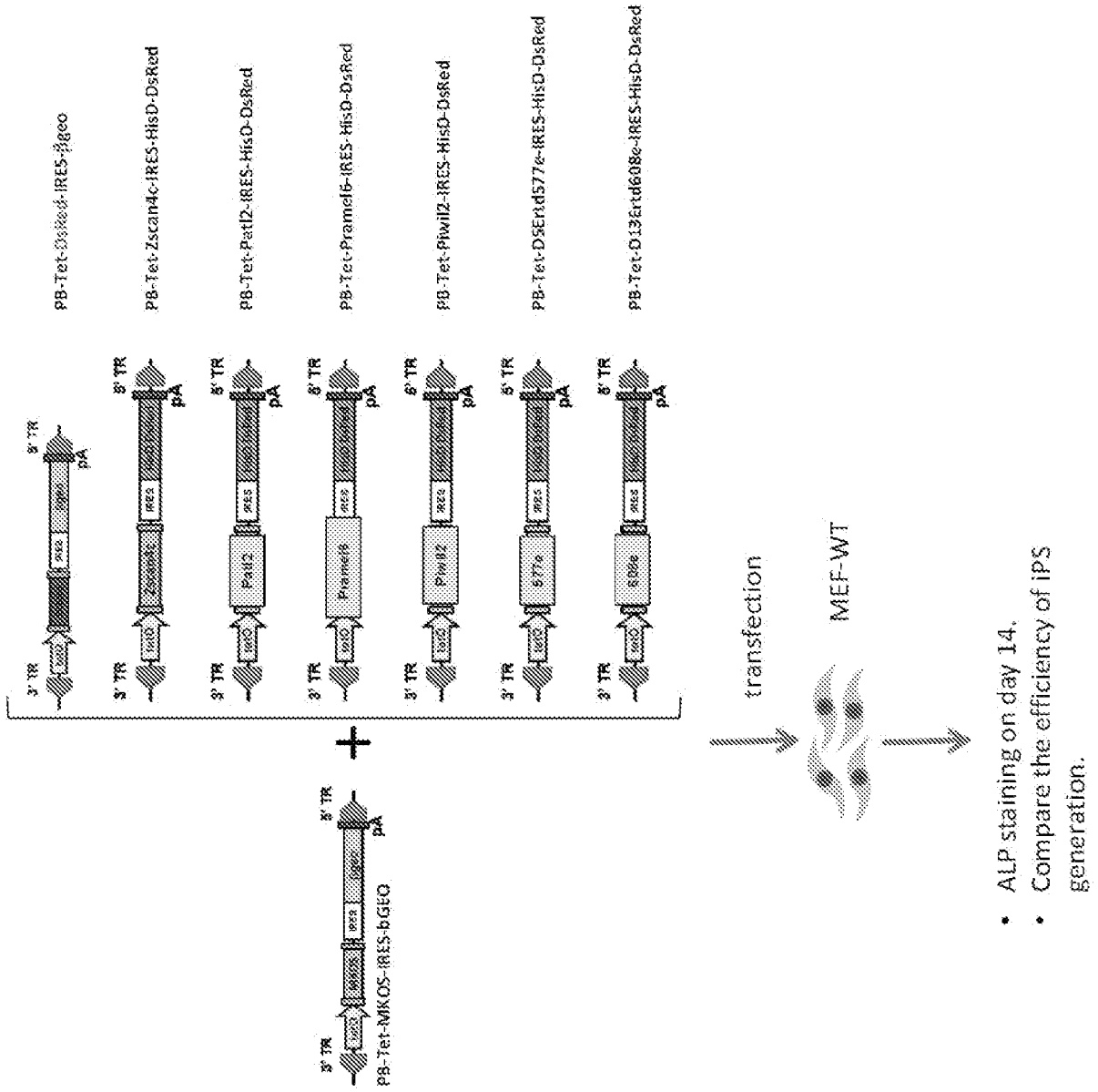
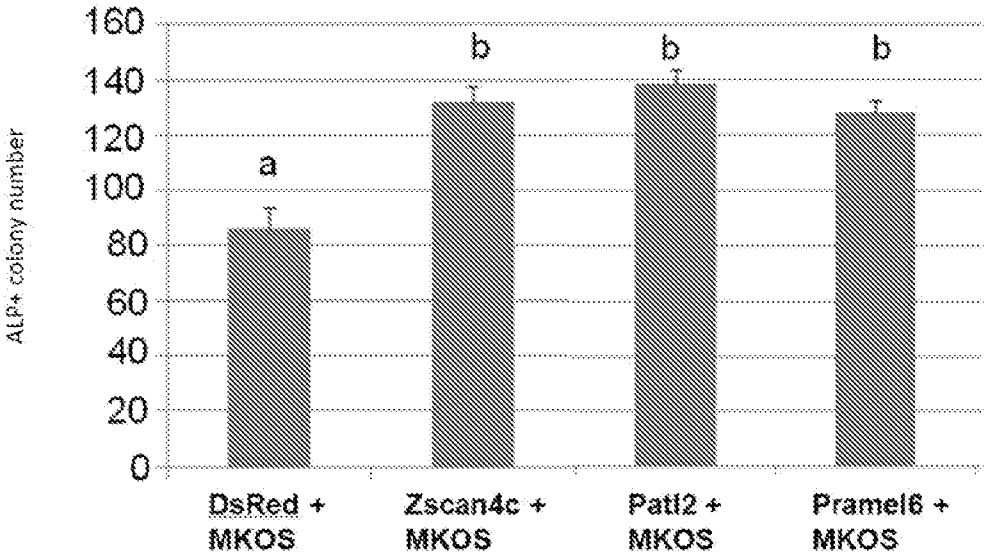
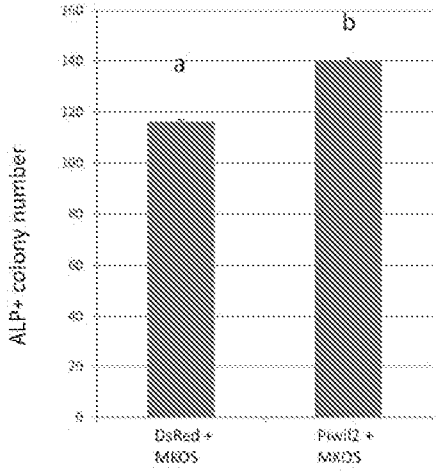

PB-Tet-Dppa4-IRES-βgeo
PB-Tet-Zscan4c-IRES-HisD-DsRed
PB-Tet-Patl2-IRES-HisD-DsRed
PB-Tet-Prame16-IRES-HisD-DsRed
PB-Tet-PiwiI2-IRES-HisD-DsRed
PB-Tet-DSErtd577e-IRES-HisD-DsRed
PB-Tet-D13Ertd608e-IRES-HisD-DsRed
PB-Tet-MKOS-IRES-bGEO
transfection
MEF-WT
• ALP staining on day 14.
• Compare the efficiency of iPS generation.

FIGURE 22

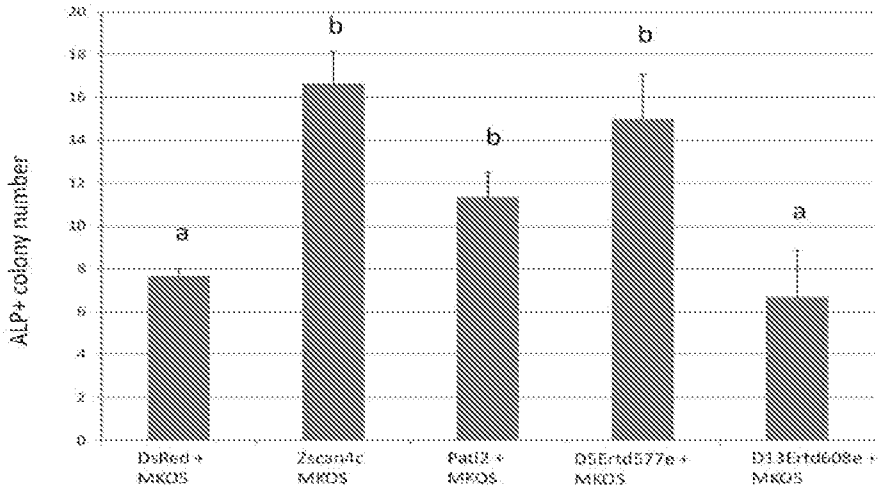
A



B



C

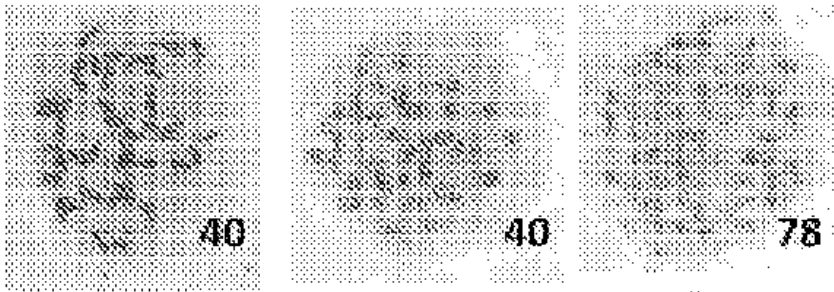


A

iPSC cell line	Zscan4 (Z)	Factors	Euploid (%)	No. injected 4N blastocysts	No. embryos at E13.5 (%)	No. live embryos at E13.5 (%)
MEF ZERT-KOS #2	Z	KOS	80	45	9 (20%)	2 (4.4%)
B5+1	Z	MKOS	58	46	1 (2%)	1 (2%)
B5+2	Z	MKOS	76	25	0 (0%)	0 (0%)
B5+4	Z	MKOS	81	N.D.	N.D.	N.D.
B5-1	-	MKOS	47	N.D.	N.D.	N.D.
B5-3	-	MKOS	0	N.D.	N.D.	N.D.

B

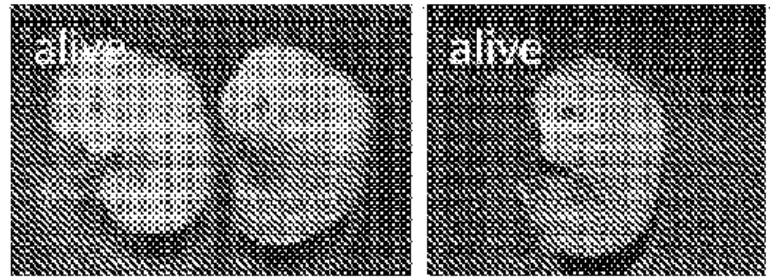
Karyotype examples



KOS+ZERT #2 (Zscan4 +) B5+2 (Zscan4+) B5-3 (Zscan4-)

C

Live embryos derived entirely from iPSC



KOS+ZERT #2 B5+1