(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 2012/158564 A1

(51) International Patent Classification:
C12N 5/0797 (2010.01)

(21) International Application Number:
PCT/US2012/037650

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

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/486,017 13 May 2011 (13.05.2011) US

(71) Applicant *(for all designated States except US)*: **THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH & 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,

[Continued on next page]

(54) Title: COMPOSITIONS AND METHODS FOR INCREASING THE FREQUENCY OF HOMOLOGOUS RECOMBINATION

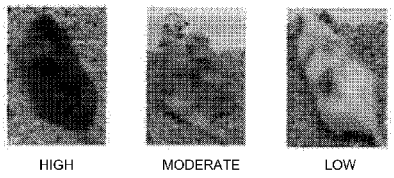

FIG. 3A

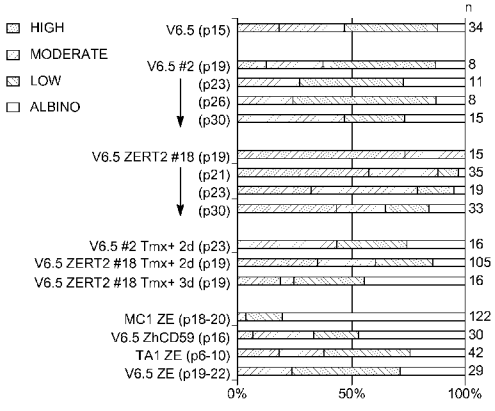

FIG. 3B

(57) Abstract: Described herein is the finding that increasing the frequency of Zscan4 activation in mouse ES cells not only enhances, but also maintains their developmental potency in long-term cell culture. As the potency increases, even a whole animal can be produced from a single ES cell injected into a 4N blastocyst at an unexpectedly high success rate. The studies disclosed herein indicate that ES cells acquire higher potency by going through the transient Zscan4 activation state more frequently than the regular state. Particularly disclosed herein is the finding that the constitutive presence of Zscan4-ERT2, even in the absence of its usual activator tamoxifen, can increase the frequency of endogenous Zscan4 activation in ES cells, resulting in the increase of developmental potency of the ES cells. Accordingly, provided herein are Zscan4-ERT2 fusion proteins and nucleic acid molecules and vectors encoding Zscan4-ERT2 fusion proteins.

WO 2012/158564 A1

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

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

COMPOSITIONS AND METHODS FOR INCREASING THE FREQUENCY OF HOMOLOGOUS RECOMBINATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0171] This application claims the benefit of U.S. Provisional Application No. 61/486,017, filed May 13, 2011, the disclosure of which is hereby incorporated by reference in its entirety.

SUBMISSION OF SEQUENCE LISTING ON ASCII TEXT FILE

[0172] The content of the following submission on ASCII text file is incorporated herein by reference in its entirety: a computer readable form (CRF) of the Sequence Listing (file name: 699442000640SeqList.txt, date recorded: May 2, 2012, size: 95 KB).

FIELD

[0173] This disclosure concerns compositions and methods for enhancing or prolonging the pluripotency of a stem cell, and the use of such pluripotent stem cells. This disclosure further concerns compositions and methods for increasing the frequency of homologous recombination in mammalian cells.

BACKGROUND

[0174] Mouse embryonic stem (ES) cells are prototypical pluripotent cells, which are derived from the inner cell mass of blastocysts (Martin, *Proc Natl Acad Sci USA* 78:7634-7638, 1981; Evans and Kaufman, *Nature* 292:154-156, 1981). ES cells have an unusual capacity of proliferating for a long time without losing their genome integrity and karyotype (Suda *et al.*, *J Cell Physiol* 133:197-201, 1987), and are capable of contributing to all the cell types in animals upon injection into mouse blastocysts (Niwa, *Development* 134:635-646, 2007). The most striking evidence of their potency has been demonstrated by injecting ES cells into tetraploid (4N) blastocysts, which produces healthy pups entirely from ES cells (Nagy *et al.*, *Proc Natl Acad Sci USA* 90:8424-8428, 1993). The ultimate test was to see if a single ES cell can form an entire healthy pup, though the success rate was extremely low (0.5%) (Wang and Jaenisch, *Dev Biol* 275:192-201, 2004).

[0175] One of the key technologies for the future of medicine, especially for gene therapy and cell transplantation therapy, is the ability to efficiently make precise modifications of genomic DNA. Gene targeting techniques, *e.g.*, replacing an endogenous DNA with an exogenous DNA fragment by homologous recombination (HR), provide an ideal means to achieve this goal (Nakayama, *Drug Discov Today* 15(5-6):198-202, 2010). For example, it would be beneficial if disease-causing mutations in a patient's DNA could be corrected by HR in the patient's cells. However, an extremely low rate of gene targeting in mammalian somatic cells ($< 10^{-6}$) hampers progress in this field (Pruett-Miller *et al.*, *Mol Ther* 16(4):707-717, 2008). Mouse ES cells are an exception, as they exhibit a high efficiency of HR, which is widely used to perform gene targeting in mouse ES cells. The mechanism by which HR is facilitated in mouse ES cells is not well understood. In higher eukaryotes, a chicken DT40 cell line derived from B lymphocytes is another notable exception (Buerstedde and Takeda, *Cell* 67(1):179-188, 1991). By contrast, the efficiency of HR in human ES cells is similar to that in somatic cells and much lower than that in mouse ES cells (Tenzen *et al.*, *J Cell Physiol* 222(2):278-281, 2009; Zou *et al.*, *Cell Stem Cell* 5(1):97-110, 2009).

[0176] Recently, zinc finger nuclease (ZFN)-mediated gene targeting technologies have been successfully used to alter a specific sequence of the genome with 1-100% efficiency (Meyer *et al.*, *Proc Natl Acad Sci USA* 107(34):15022-15026, 2010; Pruett-Miller *et al.*, *Mol Ther* 16(4):707-717, 2008; Zou *et al.*, *Cell Stem Cell* 5(1):97-110, 2009; Reyon *et al.*, *BMC Genomics* 12:83, 2011). The method utilized the ability of zinc finger proteins to recognize specific sequences in the genome, which brings in the nuclease to the site and causes double-strand DNA breaks (DSBs) in DNA and facilitates HR events. Although this technique is being investigated in clinical trials (Kaiser, *Science* 307(5715):1544-1545, 2005), the possibility of creating DSBs in many other sites in the genome is a major concern for therapeutic use of this technique.

[0177] It has recently been shown that Zscan4 (Zinc finger and scan domain-containing protein 4), which is expressed specifically in 2-cell stage embryos and ES cells (Falco *et al.*, *Dev Biol* 307:539-550, 2007), is required for the maintenance of genome stability and normal karyotype in ES cells (Zalzman *et al.*, *Nature* 464:858-863, 2010). Although only a small fraction (~5%) of undifferentiated ES cells express Zscan4 at a given time (Falco *et al.*, *Dev Biol* 307:539-550, 2007), essentially all of the ES cells in culture undergo the transient Zscan4$^+$ state within 9 passages (Zalzman *et al.*, *Nature* 464:858-863, 2010). Upon short hairpin RNA (shRNA)-mediated repression of Zscan4, after about 8 passages ES cells undergo massive karyotype deterioration. Prior studies have also shown that the Zscan4$^+$ state of ES cells is associated with

telomere extension (Zalzman *et al.*, *Nature* 464:858-863, 2010). Although ES cells have the best capacity to maintain their genome integrity in culture, it is also widely recognized that even ES cells, in long-term culture, gradually lose their developmental potency (*i.e.*, ability to contribute to tissues in chimeric mice).

SUMMARY

[0178] Disclosed herein is the finding that increasing the frequency of Zscan4 activation in mouse ES cells not only enhances, but maintains their developmental potency in long-term cell culture. In particular, disclosed herein is the finding that particular Zscan4 protein truncations and fusion proteins increase the number of Zscan4$^+$ cells and/or promote recurrent activation of Zscan4 in stem cells. Further disclosed herein is the finding that expression of Zscan4 enhances homologous recombination in mammalian cells.

[0179] Provided herein are nucleic acid molecules, including vectors, encoding a Zscan4-ERT2 fusion protein. Recombinant Zscan4-ERT2 fusion proteins are also provided. Compositions and cells (such as ES cell or iPS cells) comprising the Zscan4-ERT2 nucleic acid molecules and fusion proteins are also provided herein.

[0180] Further provided are nucleic acid molecules, including vectors, encoding a Zscan4 protein with a C-terminal truncation of at least one zinc finger domain, referred herein to as Zscan4-ΔC. Recombinant Zscan4-ΔC proteins are also provided. Compositions and cells (such as ES cell or iPS cells) comprising the Zscan4-ΔC nucleic acid molecules and proteins are also provided herein.

[0181] Further provided are methods of enhancing or prolonging the pluripotency of a stem cell or a stem cell population; methods of increasing the frequency of Zscan4-positive cells in a stem cell population; and methods of promoting genome stability or increasing telomere length, or both, in a stem cell or a stem cell population, by increasing the frequency of Zscan4 activation in the stem cell or stem cell population. In some embodiments, the methods include contacting the stem cell or stem cell population with a Zscan4-ERT2 nucleic acid molecule, fusion protein or composition as disclosed herein. In other embodiments, the methods include contacting the stem cell or stem cell population with a Zscan4-ΔC nucleic acid molecule, protein or composition as disclosed herein.

[0182] Also provided herein is a method of enhancing the efficiency of homologous recombination in a mammalian cell. In some embodiments, the method includes contacting the cell with a Zscan4-ERT2 nucleic acid molecule, vector or fusion protein, or composition thereof, as disclosed herein. The disclosed method can be carried out *in vitro* or *in vivo*.

[0183] 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

[0184] **FIG. 1** depicts that the constitutive expression of a Zscan4c-ERT2 fusion protein increases the number of Zscan4$^+$ ES cells. **FIG. 1A** is a schematic of the structure of a Zscan4c-ERT2 fusion protein. Zscan4c contains one SCAN domain and four C2H2 zinc finger domains. **FIG. 1B** are fluorescence microscopy images of MC1-ZE3 cells, in which a Zscan4 promoter drives the expression of Emerald marker (left), MC1-ZE3-ZERT2 clone #15 cells, in which the Zscan4c-ERT2 fusion protein is constitutively expressed, cultured in the absence of Tmx (middle), and MC1-ZE3-ZERT2 clone #15 cells cultured in the presence of Tmx for 3 days (right). **FIG. 1C** is a graph showing flow-cytometry analysis of MC1-ZE3 ES cells (left, control) and MC1-ZE3-ZERT2 #15 ES cells (right) in the absence or presence of 1 μM Tmx. Em fluorescence levels (average ± S.E.M.; n=6) are shown. Note 3-fold increase of Em$^+$ cells by the constitute expression of a Zscan4c-ERT2 fusion protein even without Tmx. **FIG. 1D** is a graph showing the results of quantitative RT-PCR analysis of endogenous Zscan4 expression measured by using PCR primer pairs specific for 3'-UTR of Zscan4 in MC1-ZE3 ES cells (left, control) and MC1-ZE3-ZERT2 #15 ES cells (right) in the absence or presence of 1 μM Tmx. The fold-induction of endogenous Zscan4 expression levels (average ± S.E.M.; n=6) compared to that of control MC1-ZE3 is shown. Note the 6-fold increase of endogenous Zscan4 at the RNA level by the constitute expression of a Zscan4c-ERT2 fusion protein even without Tmx. **FIG. 1E** is a series of images of V6.5 parental ES cells (passage number 14), V6.5 ZERT2 #2 (p.20), V6.5 ZERT2 #7 (p.21), V6.5 ZERT2 #10 (p.20), V6.5 ZERT2 #18 (p.22) ES cell colonies after whole-mount RNA *in situ* hybridization of a Zscan4 full-length probe, which detects both endogenous and exogenous Zscan4 RNAs (upper panel) or a Zscan4 3'-UTR probe, which detects only endogenous Zscan4 RNAs (lower panel). **FIG. 1F** is a schematic showing comparisons of global expression profiles between V6.5 ZERT2 #18 ES cells and Em$^+$ ES cells (upper panel), and between Tmx$^-$ and Tmx$^+$ conditions of V6.5 ZERT2 #18 ES cells (lower panel). Note that Zscan4-

related genes (Zscan4c, BC061212, Tmeme92, and Tcstv1/3) are already upregulated in the V6.5 ZERT2 #18 ES cells even without Tmx.

[0185] **FIG.** 2 depicts Zscan4 lacking the C-terminus increases the number of Zscan4$^+$ cells. **FIG. 2A** is a schematic showing the structure of Zscan4c, Zscan4c-ERT2, Zscan4c-ΔC and Zscan4c-ΔN proteins. Zscan4c-ΔC was made by deleting four Zinc finger domains at the C-terminus of Zscan4c protein. Zscan4c-ΔN was made by deleting the SCAN domain at the N-terminus. These mutated genes were placed under the strong and constitutive CAG promoter. Each vector was transfected into MC1-ZE16 ES cells (sister clones of MC1-ZE3). Multiple independent clones were isolated: ZDC-MC1-ZE16 #3, #4, #20 for Zscan4c-ΔC; ZDN-MC1-ZE16 #5, #8, #15 for Zscan4c-ΔN. **FIGS. 2B-2G** are fluorescence microscopic images of ZDC-MC1-ZE16 #3, #4, #20 for Zscan4c-ΔC and ZDN-MC1-ZE16 #5, #8, #15 for Zscan4c-ΔN. The results demonstrate that the expression of Zscan4c-ΔC increases the number of Zscan4$^+$ cells, whereas the expression of Zscan4c-ΔN does not change the number of Zscan4$^+$ cells.

[0186] **FIG. 3** depicts that the constitutive expression of a Zscan4c-ERT2 fusion protein increases and prolongs developmental potency of ES cells. **FIG. 3A** shows representative coat colors of chimeric mice generated by injecting various ES cells into blastocysts. The higher chimerism represents the higher contribution of injected ES cells to mice, indicating the higher developmental potency of ES cells. **FIG. 3B** is a graph showing the percent distribution of chimerism levels among "n" number of mice born from various ES cell lines.

[0187] **FIG. 4** depicts that tetraploid (4N) complementation assays confirm the higher potency of ES cells expressing a Zscan4c-ERT2 fusion protein. **FIG. 4A** is a table showing development of 4N blastocysts injected with multiple ES cells (10-15 ES cells): V6.5 parental ES cells (passage 18), V6.5 ZERT2 #7 (passage 22), V6.5 ZERT2 #10 (passage 22), V6.5 ZERT2 #18 (passage 19), and freshly isolated TA1 ES cells (passage 3). **FIG. 4B** is a table showing development of 4N blastocysts injected with single ES cells: V6.5 parental ES cells (passage 18), V6.5 ZERT2 #2 (passage 21), V6.5 ZERT2 #18, and freshly isolated TA1 ES cells (passage 4). **FIG. 4C** is an image of the embryos examined: only properly developed embryos were counted (the group on the right). **FIG. 4D** is a pair of images of two live embryos derived from single V6.5 ZERT2 #18 ES cells shown in FIG. 4A. **FIG. 4E** shows a proposed model of ES cell potency.

[0188] **FIGS. 5A-5C** depict tables providing lists of genes upregulated in MC1-ZE7 Em$^+$ cells compared to Em$^-$ cells.

[0189] **FIG. 6** depicts the generation and characterization of V6.5 ZERT2 ES cell clones. **FIG. 6A** is a graph showing results of qRT-PCR analysis of Zscan4 expression levels by a primer pair detecting RNA from both endogenous Zscan4 and exogenous Zscan4 (transcripts from a pCGA-Zscan4-ERT2). The primer sequences are 5'-AGTCTGACTGATGAGTGCTTGAAGCC-3' (SEQ ID NO: 15) and 5'-GGCCTTGTTGCAGATTGCTGTTG-3' (SEQ ID NO: 16). Data were normalized by the expression of H2A, using primers 5'-TTGCAGCTTGCTATACGTGGAGATG-3' (SEQ ID NO: 17) and 5'-TGTTGTCCTTTCTTCCCGATCAGC-3' (SEQ ID NO: 18). The expression levels are shown as a fold change relative to the Zscan4 expression levels of a parental V6.5 ES cells. **FIG. 6B** is a graph of growth curves of V6.5 ZERT2 #18 ES cells in the presence (Tmx+) or absence of Tamoxifen (Tmx-). The presence of Tmx dramatically reduced the proliferation of ES cells, which was restored by removing the Tmx from the media even after long-term culture with Tmx (Tmx +>-). **FIG. 6C** is a series of images showing morphologies of cells in each culture condition.

[0190] **FIG. 7A** depicts a scatter plot showing genes expressed differentially between V6.5 ZERT2 #18 ES cells and control V6.5 #2 ES cells. **FIG. 7B** depicts a scatter plot showing genes expressed differentially between V6.5 ZERT2 #18 ES cells cultured for 2 days in the presence of Tmx and control V6.5 #2 ES cells.

[0191] **FIG. 8** is a table listing the top 50 genes upregulated in V6.5 ZERT2 #18 ES cells compared to V6.5 #2 ES cells.

[0192] **FIG. 9** is a table listing the top 50 genes upregulated in V6.5 ZERT2 #18 ES cells cultured in the presence of Tmx for 2 days compared to V6.5 #2 ES cells.

[0193] **FIG. 10** depicts the derivation of new F1 (C57BL/6J X 129S6/SvEvTac) hybrid ES cell lines. **FIG. 10A** is a table showing blastocysts obtained by mating C57BL/6J females with 129S6/SvEvTac males. Blastocysts were cultured *in vitro* on the mouse embryo fibroblasts (MEFs) feeders in 15% KSR medium (Invitrogen) supplemented with 50 nM PD98059 (MEK1 inhibitor). *Outgrowths showed undifferentiated (U), differentiated (D), and mixed (U/D) cellular phenotypes. **FIG. 10B** is table providing a summary of ES derivation results.

[0194] **FIG. 11** depicts the results of testing developmental potency of newly derived F1 hybrid ES cell lines by tetraploid complementation assays. **FIG. 11A** is a table of six ES cell lines that showed undifferentiated cellular phenotypes when injected into tetraploid (4N) mouse

blastocysts. Success rates of obtaining live embryos at E13.5 varied among ES cell lines, ranging from 15% to 60%. Clone #10 was selected for its highest success rate (named TA1 ES cell line) and was used for subsequent studies. **FIG. 11B** is a series of representative images of 4N placentas and E13.5 embryos derived from the TA1 ES cell line. Normal appearance of female and male gonads dissected from these embryos indicates their germline competence.

[0195] **FIG. 12A** is a schematic presentation of random integration events. Because a drug selectable marker is under the control of a strong and constitutive promoter (CAG), after the drug selection, cells with a vector integrated into any random genome locations can proliferate and form colonies. **FIG. 12B** is a schematic presentation of homologous recombination events. Because a vector lacks a promoter sequence that drives the expression of drug selectable marker, after the drug selection, only cells with the proper integration of a vector into a unique site in the genome by homologous recombination can proliferate and form colonies. **FIG. 12C** is a schematic representation of the detailed structure of a vector for homologous recombination. **FIG. 12D** is a graph showing the rate of random integration events with a pCAG-EGFP-IRES-Zeo-pA vector and the rate of homologous recombination events with a pOct4-IRES-HygtkpA vector in V6.5 wild type ES cells. Rates were calculated by the colony number after the drug selection by the number of survived cell colonies after electroporation without the drug selection. Bars indicate the standard errors among the replicate.

[0196] **FIG. 13A** is a schematic showing CGZ3 ES cells were established by transfecting V6.5 ES cells sequentially with a pCAG-CreGR-IRES-His-pA plasmid and a pfloxedCAG-Zscan4ERT2-IRES-puro-pA plasmid. In GCZ3 cells without dexamethasone (Dex-), a Zscan4-ERT2 fusion protein is constitutively expressed. When cells are exposed to dexamethasone (Dex+), the Zscan4-ERT2 expression unit is excised and removed from the cells. **FIG. 13B** is a schematic of experimental procedures: CGZ3 ES cells (4×10^7 cells) were split into four 10cm dishes with or without Dex, cultured for 2 days, passaged into six 10cm dishes, and cultured for 3 days. Cells were harvested from each culture condition and transfected with 100 μg of a pBRCAG-EGFP-IRES-hyg-pA vector for the random integration or a pOct4-IRES-HygtkpA vector for the homologous recombination. After electroporation, cells were cultured in Dex-conditions for 7 days with the drug selection. **FIG. 13C** is a set of images showing cell colonies visualized by Leishman's staining. **FIG. 13D** is a graph showing the number of colonies by random integration and homologous recombination. Bars indicate the standard error among replications. **FIG. 13E** is a pair of images showing surviving colonies after electroporation. To normalize the plasmid integration efficiency by the electroporation conditions, cells were plated

after electroporation in a series of ½ dilutions and cultured for 7 days without drug selection. Colonies were visualized by Leishman's staining and scored. **FIG. 13F** is a graph showing the number of colonies that survived after electroporation. Bars indicate the standard error among replications. **FIG. 13G** is a graph showing the frequency of gene integration per survived cell. Rates of gene integration were calculated by dividing the colony numbers after the drug selection (FIG. 13D) by the number of survived colonies without the drug selection (FIG. 13F). Bars indicate the standard errors among replications. **FIG. 13H** is a summary table showing the comparison of the gene integration efficiency.

SEQUENCE LISTING

[0197] 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. In the accompanying sequence listing:

[0198] **SEQ ID NOs: 1 and 2** are nucleotide and amino acid sequences of human ZSCAN4.

[0199] **SEQ ID NOs: 3 and 4** are nucleotide and amino acid sequences of mouse Zscan4a.

[0200] **SEQ ID NOs: 5 and 6** are nucleotide and amino acid sequences of mouse Zscan4b.

[0201] **SEQ ID NOs: 7 and 8** are nucleotide and amino acid sequences of mouse Zscan4c.

[0202] **SEQ ID NOs: 9 and 10** are nucleotide and amino acid sequences of mouse Zscan4d.

[0203] **SEQ ID NOs: 11 and 12** are nucleotide and amino acid sequences of mouse Zscan4e.

[0204] **SEQ ID NOs: 13 and 14** are nucleotide and amino acid sequences of mouse Zscan4f.

[0205] **SEQ ID NOs: 15-18** are nucleotide sequences of primers used for qRT-PCR analysis of Zscan4 and H2A.

[0206] **SEQ ID NO: 19** is the nucleotide acid sequence of plasmid pPyCAG-mZscan4c-ERT2.

[0207] **SEQ ID NO: 20** is the nucleotide sequence of plasmid pPyCAG-hZscan4-ERT2.

[0208] **SEQ ID NO: 21** is the amino acid sequence of human ERT2.

[0209] **SEQ ID NO: 22** is the amino acid sequence of a mouse Zscan4c-ERT2 fusion protein.

[0210] **SEQ ID NO: 23** is the amino acid sequence of a human ZSCAN4-ERT2 fusion protein.

[0211] **SEQ ID NO: 24** is the nucleotide sequence of plasmid pCAG-Zscan4-ΔC.

[0212] **SEQ ID NO: 25** is the amino acid sequence of mouse Zscan4c-ΔC (lacking all four zinc finger domains).

DETAILED DESCRIPTION

I. Abbreviations

a.a.	amino acid
cDNA	complementary deoxyribonucleic acid
Em	Emerald
ES	embryonic stem
hCG	human chorionic gonadotropin
HR	homologous recombination
ICM	inner cell mass
LIF	leukemia inhibitory factor
MEF	murine embryonic fibroblast
ORF	open reading frame
PFA	paraformaldehyde
qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative reverse transcriptase polymerase chain

	reaction
shRNA	short hairpin ribonucleic acid
Tmx	Tamoxifen

II. Terms and Methods

[0213] 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).

[0214] In order to facilitate review of the various embodiments of the disclosure, the following explanations of specific terms are provided:

[0215] **Administration:** To provide or give a subject an agent, such as an ES cell or population of ES cells, by any effective route. An exemplary route of administration includes, but is not limited to, injection (such as subcutaneous, intramuscular, intradermal, intraperitoneal, intravenous or intra-arterial).

[0216] **Agent**: Any protein, nucleic acid molecule, compound, cell, small molecule, organic compound, inorganic compound, or other molecule of interest.

[0217] **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.

[0218] **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.

[0219] **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 embryonic 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.

[0220] **Embryonic stem (ES) cells:** Pluripotent cells isolated from the inner cell mass of a developing blastocyst. ES cells can be derived from any organism, such as a mammal. In one embodiment, ES cells are produced from mice, rats, rabbits, guinea pigs, goats, pigs, cows, non-human primates or humans. Human and murine derived ES cells are exemplary. ES cells are pluripotent cells, meaning that they can generate all of the cells present in the body (bone, muscle, brain cells, etc.). Methods for producing murine ES cells can be found, for example, in U.S. Patent No. 5,670,372. Methods for producing human ES cells can be found, for example, in U.S. Patent No. 6,090,622, PCT Publication No. WO 00/70021 and PCT Publication No. WO 00/27995. A number of human ES cell lines are known in the art and are publically available. For example, the National Institutes of Health (NIH) Human Embryonic Stem Cell Registry provides a list of a number of human ES cell lines that have been developed (a list can be found online at the NIH Office of Extramural Research website at http://grants.nih.gov/stem_cells/registry/current.htm).

[0221] **Encapsulated:** As used herein, a molecule "encapsulated" in a nanoparticle refers to a molecule (such as Zscan4-ERT2 fusion protein) that is either contained within the nanoparticle or attached to the surface of the nanoparticle, or a combination thereof.

[0222] **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). An exemplary amino acid sequence for ERT2 is set forth herein as SEQ ID NO: 21, and the corresponding coding sequence is set forth herein as nucleotides 3989-4936 of SEQ ID NO: 19.

[0223] **ES cell therapy:** A treatment that includes administration of ES cells to a subject. In particular examples, the ES cells are $Zscan4^+$ ES cells.

[0224] **Functional fragment or variant (of Zscan4):** The disclosed Zscan4 polynucleotides and polypeptides (such as those set forth as SEQ ID NOs: 1-14) include functional fragments and variants of Zscan4 that retain Zscan4 biological activity. Functional fragments and/or variants of

Zscan4 generally comprise at least about 80%, at least about 85%, at least about 90%, at least about 95% or at least about 99% sequence identity with one of the Zscan4 sequences set forth as SEQ ID NOs 1-14. When less than the entire sequence is being compared for sequence identity, fragments will typically possess at least 80% sequence identity over the length of the fragment, and can possess, for example, sequence identities of at least 85%, 90%, 95% or 99%.

[0225] **Fusion protein:** A protein 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. In some embodiments herein, the fusion protein is a Zscan4-ERT2 fusion protein. In some examples, the fusion protein comprises a linker between the two different proteins.

[0226] **Genome stability:** The ability of a cell to faithfully replicate DNA and maintain integrity of the DNA replication machinery. An ES cell with a stable genome generally defies cellular senescence, can proliferate more than 250 doublings without undergoing crisis or transformation, has a low mutation frequency and a low frequency of chromosomal abnormalities (*e.g.*, relative to embryonal carcinoma cells), and maintains genomic integrity. Long telomeres are thought to provide a buffer against cellular senescence and be generally indicative of genome stability and overall cell health. Chromosome stability (*e.g.*, few mutations, no chromosomal rearrangements or change in number) is also associated with genome stability. A loss of genome stability is associated with cancer, neurological disorders and premature aging. Signs of genome instability include elevated mutation rates, gross chromosomal rearrangements, alterations in chromosome number, and shortening of telomeres.

[0227] **Heterologous:** A heterologous polypeptide or polynucleotide refers to a polypeptide or polynucleotide derived from a different source or species. For example, a mouse Zscan4 peptide expressed in a human ES cell is heterologous to that ES cell.

[0228] **Homologous recombination:** A type of genetic recombination in which nucleotide sequences are exchanged between two similar or identical molecules of DNA. For example, homologous recombination occurs during meiosis when paired chromosomes from two parents align so that similar DNA sequences from the paired chromosomes cross over each other. Crossing over results in an exchange of genetic material between the chromosomes. Cells also utilize homologous recombination to repair harmful double-strand DNA breaks. Homologous

recombination is also used in horizontal gene transfer to exchange genetic material between different strains and species of bacteria and viruses, and to generate transgenic (knockout) mice.

[0229] **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.

[0230] **Induced pluripotent stem (iPS) cells**: A type of pluripotent stem cell artificially derived from a non-pluripotent cell, typically an adult somatic cell, by inducing a "forced" expression of certain genes. iPS cells can be derived from any organism, such as a mammal. In one embodiment, 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.

[0231] 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. For example, iPS cells are typically derived by transfection of certain stem cell-associated genes (such as Oct-3/4 (Pouf51) and Sox2) into non-pluripotent cells, such as adult fibroblasts. Transfection can be achieved through viral vectors, such as retroviruses, lentiviruses, or adenoviruses. 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. In one example, iPS from adult human cells are generated by the method of Yu *et al.* (*Science* 318(5854):1224, 2007) or Takahashi *et al.* (Cell 131(5):861-72, 2007).

[0232] **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.

[0233] **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.

[0234] **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.

[0235] **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 transgenic animal, such as a transgenic mouse, cow, sheep, or goat. In one specific, non-limiting example, the transgenic non-human animal is a mouse.

[0236] **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.

[0237] **Pharmaceutically acceptable carriers:** The pharmaceutically acceptable carriers of use are conventional. *Remington's Pharmaceutical Sciences*, by E. W. Martin, Mack Publishing Co., Easton, PA, 15th Edition (1975), describes compositions and formulations suitable for pharmaceutical delivery of the Zscan4 proteins (including fusion proteins), Zscan4 nucleic acid molecules, or cells herein disclosed.

[0238] In general, the nature of the carrier will depend on the particular mode of administration being employed. For instance, parenteral formulations usually comprise injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, glycerol or the like as a vehicle. For solid compositions (e.g., powder, pill, tablet, or capsule forms), conventional non-toxic solid carriers can include, for example, pharmaceutical grades of mannitol, lactose, starch, or magnesium stearate. In addition to biologically-neutral carriers, pharmaceutical compositions to be administered can contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example, sodium acetate or sorbitan monolaurate.

[0239] **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, but cannot form an entire organisms autonomously. As used herein, enhancing or prolonging pluripotency refers to increasing the pluripotent capacity or duration of pluripotency of a stem cell.

[0240] **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.

[0241] 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.

[0242] 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.

[0243] 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

[0244] 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 (or Zscan4 fusion protein, such as Zscan4-ERT2), or other polypeptides disclosed herein, 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. Variant amino acid sequences may be,

for example, at least 80%, 90% or even 95% or 98% identical to the native amino acid sequence (such as a native Zscan4 sequence or a Zscan4-ERT2 sequence, such as SEQ ID NO: 22 or 23).

[0245] **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).

[0246] **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.

[0247] **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).

[0248] 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.* 16:1088190, 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.

[0249] 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.

[0250] **Stem cell:** A cell having the unique capacity to produce unaltered daughter cells (self-renewal; cell division produces at least one daughter cell that is identical to the parent cell) and to give rise to specialized cell types (potency). Stem cells include, but are not limited to, ES cells, EG cells, GS cells, MAPCs, maGSCs, USSCs, adult stem cells and induced pluripotent stem cells. In one embodiment, stem cells can generate a fully differentiated functional cell of more than one given cell type. The role of stem cells *in vivo* is to replace cells that are destroyed during the normal life of an animal. Generally, stem cells can divide without limit. After division, the stem cell may remain as a stem cell, become a precursor cell, or proceed to terminal differentiation. A precursor cell is a cell that can generate a fully differentiated functional cell of at least one given cell type. Generally, precursor cells can divide. After division, a precursor cell can remain a precursor cell, or may proceed to terminal differentiation.

[0251] **Subject:** Living multi-cellular vertebrate organisms, a category that includes human and non-human mammals.

[0252] **Subpopulation:** An identifiable portion of a population. As used herein, a "subpopulation" of ES cells expressing Zscan4 is the portion of ES cells in a given population that has been identified as expressing Zscan4.

[0253] **Telomere:** Refers to the end of a eukaryotic chromosome, a specialized structure involved in the replication and stability of the chromosome. Telomeres consist of many repeats of a short DNA sequence in a specific orientation. Telomere functions include protecting the ends of the chromosome so that chromosomes do not end up joined together, and allowing replication of the extreme ends of the chromosomes (by telomerase). The number of repeats of telomeric DNA at the end of a chromosome decreases with age.

[0254] **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.

[0255] **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.

[0256] **Zscan4:** A group of genes that have previously 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). 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 sequences are provided herein (see SEQ ID NOs: 1-14).

[0257] **Zscan4-ΔC:** In the context of the present disclosure, "Zscan4-ΔC" includes any mouse or human Zscan4 protein having a C-terminal deletion of at least one zinc finger domain. In some embodiments, the Zscan4-ΔC protein includes a deletion of at least two, such as three or all four zinc finger domains. SEQ ID NO: 2 and SEQ ID NO: 8 provide the amino acid sequences of human ZSCAN4 and mouse Zscan4c, respectively, and delineate the N-terminal SCAN domain and C-terminal zinc finger (C2H2-type) domains. In addition, the nucleotide and amino acid regions of each domain of human ZSCAN4 is listed in Table 1, and the nucleotide and amino acid regions of each domain of mouse Zscan4c is listed in Table 2.

Table 1: Human ZSCAN4

Domain	Nucleotides (SEQ ID NO: 1)	Amino Acids (SEQ ID NO: 2)
SCAN	826-1074	44-126
C2H2-type 1	1630-1698	312-334
C2H2-type 2	1714-1782	340-362
C2H2-type 3	1798-1866	368-390
C2H2-type 4	1882-1950	396-418

Table 2: Mouse Zscan4c

Domain	Nucleotides (SEQ ID NO: 7)	Amino Acids (SEQ ID NO: 8)
SCAN	309-557	37-119
C2H2-type 1	1383-1451	395-417
C2H2-type 2	1470-1538	424-446
C2H2-type 3	1554-1622	452-474
C2H2-type 4	1638-1709	480-503

[0258] **Zscan4-ERT2:** A fusion protein made up of a Zscan4 amino acid sequence and an ERT2 amino acid sequence. "Zscan4-ERT2" can also refer to a nucleic acid sequence encoding a Zscan4-ERT2 fusion protein. Exemplary amino acid sequences for Zscan4 (including SEQ ID NO: 2, 8, 10 and 14) and ERT2 (SEQ ID NO: 21) are set forth herein. In some embodiments, the Zscan4 sequence is a functional fragment or variant of a known Zscan4 sequence (such as SEQ ID NO: 2, 8, 10 or 14) and/or the ERT2 sequence is a functional fragment or variant of a known ERT2 sequence (such as SEQ ID NO: 21). Any fragment or variant of Zscan4 or ERT2 is contemplated as long as the fragment or variant retains activity. In some examples, the Zscan4-ERT2 fusion protein comprises a linker (or spacer) between Zscan4 and ERT2. Linkers are well known in the art and an appropriate linker can be selected by one of ordinary skill in the art. In particular examples, the linker is encoded by the nucleotide sequence GCTAGC.

[0259] 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. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. 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. Enhancing the Pluripotency of Stem Cells

[0260] The gold standard for examining the pluripotency of stem cells is to see whether cells can contribute to the entire body of an animal. It is disclosed herein that increasing the frequency of Zscan4 activation in mouse ES cells not only enhances, but also maintains their developmental potency in long-term cell culture. As the potency increases, even a whole animal can be produced from a single ES cell injected into a 4N blastocyst at an unexpectedly high success rate. Although Zscan4-activated cells express genes that are also expressed in 2-cell stage mouse embryos, the transiently Zscan4-activated state of ES cells is not associated with the high potency of ES cells. While not wishing to be bound by theory, these findings indicate that ES cells acquire higher potency by going through the transient Zscan4 activation state more frequently than the regular state. Taken together, these results demonstrate that the frequent activation of Zscan4 can rejuvenate pluripotent stem cells.

[0261] Particularly disclosed herein is the finding that the constitutive presence of Zscan4-ERT2, even in the absence of its usual activator tamoxifen, can increase the frequency of endogenous Zscan4 activation in ES cells, resulting in the increase of developmental potency of the ES cells. ES cells cultured in the accelerated Zscan4 activation cycle exhibited improved chimerism and potency, which are evidenced by a high contribution to chimeric mice and efficient production of a whole mouse from a single ES cell. Further disclosed herein is the finding that expression of C-terminally truncated Zscan4 (lacking the zinc finger domains) increases the number of Zscan4$^+$ cells, thus having an effect similar to Zscan4-ERT2.

[0262] Accordingly, provided herein are isolated nucleic acid molecules encoding a Zscan4-ERT2 fusion protein. In particular examples, the Zscan4 is mouse Zscan4c or human ZSCAN4. Further provided are vectors comprising a Zscan4-ERT2 coding sequence, cells comprising such vectors (such as ES cells, iPS cells or other stem cells), and compositions that include the Zscan4-ERT2 encoding nucleic acid molecules or vectors. Further provided are recombinant Zscan4-ERT2 fusion proteins, cells comprising Zscan4-ERT2 fusion proteins and compositions that include the Zscan4-ERT2 fusion proteins.

[0263] Further provided herein are isolated nucleic acid molecules encoding a Zscan4ΔC protein (a Zscan4 protein having a deletion of at least one zinc finger domain). In particular examples, the Zscan4 is mouse Zscan4c or human ZSCAN4. Further provided are vectors comprising a Zscan4-ΔC coding sequence, cells comprising such vectors (such as ES cells, iPS cells or other stem cells), and compositions that include the Zscan4-ΔC encoding nucleic acid

molecules or vectors. Further provided are recombinant Zscan4-ΔC proteins, cells comprising Zscan4-ΔC proteins and compositions that include the Zscan4-ΔC proteins.

[0264] Also provided herein are methods of using the Zscan4-ERT2 or Zscan4-ΔC nucleic acid molecules and proteins. For example, methods of enhancing or prolonging the pluripotency of a stem cell or a stem cell population by contacting the stem cell or stem cell population with a Zscan4-ERT2 nucleic acid molecule or fusion protein are disclosed herein. In other examples, methods of enhancing or prolonging the pluripotency of a stem cell or a stem cell population by contacting the stem cell or stem cell population with a Zscan4-ΔC nucleic acid molecule or protein are provided. Similarly, methods of increasing the frequency of Zscan4-positive cells in a stem cell population, as well as methods of promoting genome stability and/or increasing telomere length in a stem cell or a stem cell population, are provided.

A. Compositions, Vectors and Cells Comprising Zscan4-ERT2

[0265] Provided herein are isolated nucleic acid molecules encoding a fusion protein, wherein the fusion protein includes a Zscan4 protein fused to an ERT2 protein. ERT2 is a mutated version of the ligand binding domain of human estrogen receptor. ERT2 does not bind its natural ligand (17β-estradiol) at physiological concentrations, but is highly sensitive to nanomolar concentrations of tamoxifen or its metabolite 4-hydroxytamoxifen (4OHT).

[0266] In some embodiments, the nucleic acid molecule encoding the Zscan4-ERT2 fusion protein includes human ZSCAN4, mouse Zscan4c, mouse Zscan4d or mouse Zscan4f, or a functional fragment or variant thereof. Functional fragments and variants of Zscan4 include, for example, any Zscan4 fragment or variant that retains one or more biological activities of Zscan4, such as the capacity to increase pluripotency of a stem cell, promote genomic stability or increase telomere length.

[0267] Exemplary nucleic acid sequences for a variety of Zscan4 polynucleotides are known in the art (see, for example, PCT Publication No. WO 2008/118957) and are set forth herein, such as SEQ ID NO: 1 (human ZSCAN4), SEQ ID NO: 7 (mouse Zscan4c), SEQ ID NO: 9 (mouse Zscan4d) and SEQ ID NO: 13 (mouse Zscan4f). One skilled in the art will appreciate that sequences having 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% sequence identity to these sequences and retain Zscan4 activity are contemplated and can be used in the compositions and methods provided herein.

[0268] Zscan4 nucleic acid sequences from other species are also publically available, including dog ZSCAN4 (GenBank Accession Nos. XM_541370 and XM_848557); cow ZSCAN4 (GenBank Accession No. XM_001789250); and horse ZSCAN4 (GenBank Accession No. XM_001493944). Each of the above-listed GenBank Accession numbers is herein incorporated by references as it appears in the GenBank database on February 22, 2011.

[0269] Fragments and variants of Zscan4 polynucleotides can readily be prepared by one of skill in the art using molecular techniques. In one embodiment, a fragment of a Zscan4 nucleic acid sequences includes at least 250, at least 500, at least 750, at least 1000, at least 1500, or at least 2000 consecutive nucleic acids of the Zscan4 polynucleotide. In a further embodiment, a fragment of Zscan4 is a fragment that confers a function of Zscan4 when expressed in a cell of interest, such as, but not limited to, promoting pluripotency, enhancing genome stability and/or increasing telomere length. The Zscan4 nucleic acid sequences contemplated herein include sequences that are 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 Zscan4 polypeptide encoded by the nucleotide sequence is functionally unchanged.

[0100] In some embodiments, the Zscan4 nucleic acid sequence portion of the nucleic acid molecule encoding the Zscan4-ERT2 fusion protein 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: 1, 7, 9 or 13. In some embodiments, the Zscan4 nucleic acid sequence comprises the nucleic acid sequence set forth in SEQ ID NO: 1, 7, 9 or 13. In some embodiments, the Zscan4 nucleic acid sequence consists of the nucleic acid sequence set forth in SEQ ID NO: 1, 7, 9 or 13.

[0101] In some examples, the Zscan4 portion of the Zscan4-ERT2 fusion protein comprises mouse Zscan4c. Thus, in particular examples, the Zscan4 nucleic 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: 7. In other examples, the Zscan4 comprises human ZSCAN4. In particular examples, the Zscan4 nucleic 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: 1.

[0102] In some embodiments, the nucleic acid sequence encoding the ERT2 portion of the Zscan4-ERT2 fusion protein 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 nucleotides 3989-4936 of SEQ ID NO: 19.

In some examples, the nucleic acid sequence encoding ERT2 comprises or consists of nucleotides 3989-4936 of SEQ ID NO: 19.

[0103] In some embodiments, the nucleic acid molecule encoding the Zscan4-ERT2 fusion protein includes a linker sequence between the Zscan4 and ERT2 coding sequences. Linkers are well known in the art and selection of an appropriate linker is well within the capabilities of one of ordinary skill in the art. In some examples, the linker is at least 2 amino acids (aa), at least 3, at least 5, at least 10, at least 20, at least 50 or at least 100 aa, such as 2 to 50 or 2 to 10 aa. In particular examples disclosed herein, the linker includes the nucleic acid sequence GCTAGC (nucleotides 3983-3988 of SEQ ID NO: 19).

[0104] In some embodiments in which the Zscan4-ERT2 nucleic acid molecule encodes mouse Zscan4c, the nucleic acid molecule comprises a 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 nucleotides 2465-4936 of SEQ ID NO: 19. In particular examples, the nucleic acid molecule comprises, or consists of, the sequence of nucleotides 2465-4936 of SEQ ID NO: 19.

[0105] In other embodiments in which the Zscan4-ERT2 nucleic acid molecule encodes human ZSCAN4, the nucleic acid molecule comprises a 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 nucleotides 2479-4731 of SEQ ID NO: 20. In particular examples, the nucleic acid molecule comprises, or consists of, the sequence of nucleotides 2479-4731 of SEQ ID NO: 20.

[0106] Also provided are vectors that include a Zscan4-ERT2 encoding nucleic acid molecule disclosed herein. Any suitable expression vector, such as an expression (plasmid) vector (e.g., pPyCAG-BstXI-IP), or viral vector (e.g., an adenovirus, adenoassociated virus, lentivirus or retrovirus vector), is contemplated. Numerous expression vectors and viral vectors are known in the art and the selection of an appropriate vector is well within the capabilities of one of ordinary skill in the art.

[0107] In some embodiments, the vector comprises a nucleotide sequence that 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: 19 or SEQ ID NO: 20. In some examples, the vector comprises a nucleic acid sequence that is at least 95% identical to SEQ ID NO: 19 or SEQ ID NO: 20. In specific non-limiting embodiments, the nucleic acid sequence of the vector comprises, or consists of, SEQ ID NO: 19 or SEQ ID NO: 20.

[0108] Further provided herein are isolated cells containing a Zscan4-ERT2 nucleic acid molecule or vector as described herein. In some embodiments, the cell is a stem cell. In particular examples, the stem cell is an ES cell or an iPS cell. The origin of the stem cell can be from any suitable species. In some examples, the stem cell is a mouse, rat, human or non-human primate stem cell.

[0109] Compositions comprising a nucleic acid molecule or vector encoding a Zscan4-ERT2 fusion protein are also provided herein. The compositions may further include a carrier or diluent, such as a pharmaceutically acceptable carrier or diluent.

[0110] Zscan4-ERT2 fusion proteins encoded by the nucleic acid molecules and vectors described herein are further provided.

[0111] Also provided herein are recombinant Zscan4-ERT2 fusion proteins. In some embodiments, the recombinant Zscan4-ERT2 fusion protein includes human ZSCAN4, mouse Zscan4c, mouse Zscan4d or mouse Zscan4f, or a functional fragment or variant thereof. Functional fragments and variants of Zscan4 include, for example, any Zscan4 fragment or variant that retains one or more biological activities of Zscan4, such as the capacity to increase pluripotency of a stem cell, promote genomic stability or increase telomere length.

[0112] Exemplary amino acid sequences for a variety of Zscan4 proteins are known in the art (see, for example, PCT Publication No. WO 2008/118957) and are set forth herein, such as SEQ ID NO: 2 (human ZSCAN4), SEQ ID NO: 8 (mouse Zscan4c), SEQ ID NO: 10 (mouse Zscan4d) and SEQ ID NO: 14 (mouse Zscan4f). One skilled in the art will appreciate that sequences having 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% sequence identity to these sequences and retain Zscan4 activity are contemplated and can be used in the methods provided herein.

[0113] Zscan4 amino acid sequences from other species are publically available, including dog ZSCAN4 (GenBank Accession Nos. XP_541370 and XP_853650); cow ZSCAN4 (GenBank Accession No. XP_001789302); and horse ZSCAN4 (GenBank Accession No. XP_001493994). Each of the above-listed GenBank Accession numbers is herein incorporated by references as it appears in the GenBank database on February 22, 2011.

[0114] Fragments and variants of a Zscan4 protein can readily be prepared by one of skill in the art using molecular techniques. In one embodiment, a fragment of a Zscan4 protein includes at

least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 450 or at least 500 consecutive amino acids of the Zscan4 polypeptide. In a further embodiment, a fragment of Zscan4 is a fragment that confers a function of Zscan4, such as, but not limited to, increasing pluripotency, enhancing genome stability or increasing telomere length.

[0115] In some embodiments, the Zscan4 protein portion of the Zscan4-ERT2 fusion protein includes an amino acid 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 the amino acid sequence set forth in SEQ ID NO: 2, 8, 10 or 14. In a further embodiment, the Zscan4 protein is a conservative variant of SEQ ID NO: 2, 8, 10 or 14, such that it includes no more than fifty conservative amino acid substitutions, such as no more than two, no more than five, no more than ten, no more than twenty, or no more than fifty conservative amino acid substitutions in SEQ ID NO: 2, 8, 10 or 14. In another embodiment, the Zscan4 peptide portion of the Zscan4-ERT2 fusion protein has an amino acid sequence comprising or consisting of the amino acid sequence set forth in SEQ ID NO: 2, 8, 10 or 14.

[0116] In some embodiments of the Zscan4-ERT2 fusion proteins, the Zscan4 comprises mouse Zscan4c. In some examples, the Zscan4c 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 the amino acid sequence of SEQ ID NO: 8. In other embodiments of the Zscan4-ERT2 fusion proteins, the Zscan4 portion comprises 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 the amino acid sequence of SEQ ID NO: 2.

[0117] In some embodiments, the amino acid sequence of the ERT2 portion of the Zscan4-ERT2 fusion protein 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: 21. In some examples, the amino acid sequence of ERT2 comprises or consists of SEQ ID NO: 21.

[0118] In some embodiments, the Zscan4-ERT2 fusion protein includes a linker between the Zscan4 and ERT2 sequences. Linkers are well known in the art and selection of an appropriate linker is well within the capabilities of one of ordinary skill in the art. In particular examples disclosed herein, the linker includes the amino acid sequence Ala-Ser.

[0119] In some embodiments in which the Zscan4-ERT2 fusion protein includes mouse Zscan4c, the amino acid sequence of the fusion protein 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: 22. In particular examples, the amino acid sequence of the Zscan4-ERT2 fusion protein comprises, or consists of, the amino acid sequence of SEQ ID NO: 22

[0120] In other embodiments in which the Zscan4-ERT2 fusion protein includes human ZSCAN4, the amino acid sequence of the fusion protein 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: 23. In particular examples, the amino acid sequence of the Zscan4-25 ERT2 fusion protein comprises, or consists of, the amino acid sequence of SEQ ID NO: 23.

[0121] Further provided herein are isolated cells comprising a Zscan4-ERT2 fusion protein disclosed herein. In some embodiments, the cells are stem cells. In particular examples, the stem cells are ES cells or iPS cells. The origin of the stem cell can be from any suitable species. In some examples, the stem cell is a mouse, rat, human or non-human primate stem cell.

[0122] Compositions comprising a Zscan4-ERT2 fusion protein are also provided herein. The compositions may further include a carrier or diluent, such as a pharmaceutically acceptable carrier or diluent, for example saline.

B. Compositions, Vectors and Cells Comprising Zscan4-ΔC

[0123] Also provided herein are isolated nucleic acid molecules encoding a Zscan4 protein with a C-terminal truncation (referred to herein as Zscan4-ΔC). The C-terminally truncated Zscan4 comprises a deletion of at least one zinc finger domain. Thus, in some embodiments, the Zscan4-ΔC protein has a deletion of one, two, three or four zinc finger domains.

[0124] In some embodiments, the nucleic acid molecule encoding the Zscan4-ΔC protein includes C-terminally truncated human ZSCAN4, mouse Zscan4c, mouse Zscan4d or mouse Zscan4f. In particular embodiments, the Zscan4-ΔC protein is either human ZSCAN4 or mouse Zscan4c with a deletion of all four zinc finger domains. In one non-limiting example, the Zscan4-ΔC protein comprises the amino acid sequence of SEQ ID NO: 25 and/or is encoded by nucleotides 2465-3649 of SEQ ID NO: 24.

[0125] The Zscan4-ΔC nucleic acid sequences contemplated herein include sequences that are 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 Zscan4-ΔC polypeptide encoded by the nucleotide sequence is functionally unchanged.

[0126] In some embodiments, the Zscan4-ΔC nucleic 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 nucleotides 2465-3649 of SEQ ID NO: 24. In some embodiments, the Zscan4-ΔC nucleic acid sequence comprises the nucleic acid sequence set forth as nucleotides 2465-3649 of SEQ ID NO: 24. In some embodiments, the Zscan4-ΔC nucleic acid sequence consists of the nucleic acid sequence set forth as nucleotides 2465-3649 of SEQ ID NO: 24.

[0127] In some embodiments, the Zscan4-ΔC nucleic acid molecule is a human Zscan4ΔC nucleic acid molecule. In particular examples, the human Zscan4-ΔC nucleic acid molecule comprises a deletion of at least nucleotides 1630-1950, nucleotides 1714-1950, nucleotides 1798-1950 or nucleotides 1882-1950 of SEQ ID NO: 1. In some embodiments, the human Zscan4-ΔC nucleic acid molecule 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 nucleotides 1-1629, nucleotides 1-1713, nucleotides 1-1797 or nucleotides 1-1881 of SEQ ID NO: 1. In some examples, the human Zscan4-ΔC nucleic acid molecule comprises or consists of nucleotides 1-1629, nucleotides 1-1713, nucleotides 1-1797 or nucleotides 1-1881 of SEQ ID NO: 1.

[0128] In some embodiments, the Zscan4-ΔC nucleic acid molecule is a mouse Zscan4ΔC nucleic acid molecule. In particular examples, the mouse Zscan4-ΔC nucleic acid molecule comprises a deletion of at least nucleotides 1383-1709, nucleotides 1470-1709, nucleotides 1554-1709 or nucleotides 1638-1709 of SEQ ID NO: 7. In some embodiments, the mouse Zscan4-ΔC nucleic acid molecule 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 nucleotides 1-1382, nucleotides 1-1469, nucleotides 1-1553 or nucleotides 1-1637 of SEQ ID NO: 7. In some examples, the mouse Zscan4-ΔC protein comprises or consists of nucleotides 1-1382, nucleotides 1-1469, nucleotides 1-1553 or nucleotides 1-1637 of SEQ ID NO: 7.

[0129] Also provided are vectors that include a Zscan4-ΔC encoding nucleic acid molecule disclosed herein. Any suitable expression vector, such as an expression (plasmid) vector (e.g., pPyCAG-BstXI-IP), or viral vector (e.g., an adenovirus, adenoassociated virus, lentivirus or retrovirus vector), is contemplated. Numerous expression vectors and viral vectors are known in the art and the selection of an appropriate vector is well within the capabilities of one of ordinary skill in the art.

[0130] In some embodiments, the vector comprises a nucleotide sequence that 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: 24. In some examples, the vector comprises a nucleic acid sequence that is at least 95% identical to SEQ ID NO: 24. In specific non-limiting embodiments, the nucleic acid sequence of the vector comprises, or consists of, SEQ ID NO: 24.

[0131] Further provided herein are isolated cells containing a Zscan4-ΔC nucleic acid molecule or vector as described herein. In some embodiments, the cell is a stem cell. In particular examples, the stem cell is an ES cell or an iPS cell. The origin of the stem cell can be from any suitable species. In some examples, the stem cell is a mouse, rat, human or non-human primate stem cell.

[0132] Compositions comprising a nucleic acid molecule or vector encoding a Zscan4ΔC protein are also provided herein. The compositions may further include a carrier or diluent, such as a pharmaceutically acceptable carrier or diluent.

[0133] Zscan4-ΔC proteins encoded by the nucleic acid molecules and vectors described herein are further provided.

[0134] Also provided herein are recombinant Zscan4-ΔC proteins. In some embodiments, the recombinant Zscan4-ΔC protein includes C-terminally truncated human ZSCAN4, mouse Zscan4c, mouse Zscan4d or mouse Zscan4f.

[0135] In some embodiments, the Zscan4-ΔC protein includes an amino acid 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 the amino acid sequence set forth in SEQ ID NO: 25. In a further embodiment, the Zscan4-ΔC protein is a conservative variant of SEQ ID NO: 25, such that it includes no more than fifty conservative amino acid substitutions, such as no more than two, no more than five, no more than ten, no more than twenty, or no more than fifty conservative amino acid substitutions in SEQ ID NO: 25. In another embodiment, the Zscan4-ΔC protein has an amino acid sequence comprising or consisting of the amino acid sequence set forth in SEQ ID NO: 25.

[0136] In some embodiments, the Zscan4-ΔC protein is a human Zscan4-ΔC protein. In particular examples, the human Zscan4-ΔC protein comprises a deletion of at least amino acids 312-418, amino acids 340-418, amino acids 368-390 or amino acids 396-418 of SEQ ID NO: 2. In some embodiments, the human Zscan4-ΔC protein 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 amino acids 1-311, amino acids 1-339, amino acids 1-367 or amino acids 1-395 of SEQ ID NO: 2. In some examples, the human Zscan4-ΔC protein comprises or consists of amino acids 1-311, amino acids 1-339, amino acids 1-367 or amino acids 1-395 of SEQ ID NO: 2.

[0137] In some embodiments, the Zscan4-ΔC protein is a mouse Zscan4-ΔC protein. In particular examples, the mouse Zscan4-ΔC protein comprises a deletion of at least amino acids 395-503, amino acids 424-503, amino acids 452-503 or amino acids 480-503 of SEQ ID NO: 8. In some embodiments, the mouse Zscan4-ΔC protein 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 amino acids 1-394, amino acids 1-423, amino acids 1-451 or amino acids 1-479 of SEQ ID NO: 8. In some examples, the mouse Zscan4-ΔC protein comprises or consists of amino acids 1-394, amino acids 1-423, amino acids 1-451 or amino acids 1-479 of SEQ ID NO: 8.

[0138] Further provided herein are isolated cells comprising a Zscan4-ΔC protein disclosed herein. In some embodiments, the cells are stem cells. In particular examples, the stem cells are ES cells or iPS cells. The origin of the stem cell can be from any suitable species. In some examples, the stem cell is a mouse, rat, human or non-human primate stem cell.

[0139] Compositions comprising a Zscan4-ΔC protein are also provided herein. The compositions may further include a carrier or diluent, such as a pharmaceutically acceptable carrier or diluent, for example saline.

C. Recurrent Activation of Zscan4 in Stem Cells and Methods of Use

[0140] Disclosed herein is the finding that recurrent activation of Zscan4 enhances the pluripotency of stem cells. In particular, it is disclosed herein that increasing the frequency of Zscan4 activation in ES cells enhances and maintains developmental potency in long-term culture. The results described in the Examples below indicate that ES cells acquire higher potency by going through the transient Zscan4 activation state more frequently than the regular state.

[0141] Thus, provided herein are methods of enhancing or prolonging the pluripotency of a stem cell or a stem cell population by inducing frequent activation of Zscan4 in the stem cell or stem cell population. Methods of increasing the frequency of Zscan4-positive cells in a stem cell population by inducing frequent activation of Zscan4 are also provided. Further provided are methods of promoting genome stability or increasing telomere length, or both, in a stem cell or a

stem cell population by promoting recurrent activation of Zscan4 in the stem cell or stem cell population.

[0142] In some embodiments of the methods disclosed herein, the methods include contacting the stem cell or stem cell population with (i) a nucleic acid molecule encoding a Zscan4-ERT2 fusion protein or a composition thereof, (ii) a vector encoding a Zscan4-ERT2 fusion protein or a composition thereof, or (iii) a Zscan4-ERT2 fusion protein or a composition thereof.

[0143] In other embodiments of the methods disclosed herein, the methods include contacting the stem cell or stem cell population with (i) a nucleic acid molecule encoding a Zscan4-ΔC protein or a composition thereof, (ii) a vector encoding a Zscan4-ΔC protein or a composition thereof, or (iii) a Zscan4-ΔC protein or a composition thereof.

[0144] In other embodiments, a stem cell or stem cell population is contacted with an agent that promotes frequent activation of Zscan4. The agent can be, for example, any nucleic acid molecule, polypeptide, small molecule or other compound that results in recurrent activation of Zscan4 in a cell.

[0145] In some examples, the stem cell is an ES cell or an iPS. The methods can be applied to stem cells of any species, for example, mouse, rat, human or non-human primate stem cells.

1. Enhancing or prolonging pluripotency of stem cells

[0146] Provided herein is a method of enhancing or prolonging the pluripotency of a stem cell or a stem cell population. In some embodiments, the method includes contacting the stem cell or stem cell population with a nucleic acid molecule or vector encoding a Zscan4-ERT2 fusion protein as disclosed herein. In other embodiments, the method includes contacting the stem cell or stem cell population with a Zscan4-ERT fusion protein disclosed herein.

[0147] In yet other embodiments, the method includes contacting the stem cell or stem cell population with a nucleic acid molecule or vector encoding a Zscan4-ΔC protein as disclosed herein. In other embodiments, the method includes contacting the stem cell or stem cell population with a Zscan4-ΔC protein disclosed herein.

[0148] Methods of delivering a nucleic acid molecule into a cell are well known in the art. In some examples, "contacting" the stem cell with a nucleic acid molecule or vector includes

transfection (such as liposomal-mediated transfection), electroporation, injection or any other suitable technique for introducing a nucleic acid molecule into a cell.

[0149] Methods for delivery of proteins to cells are also well known in the art. In some examples, the Zscan4-ERT2 fusion protein or Zscan4-ΔC 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.

[0150] 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.

[0151] 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).

[0152] 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).

[0153] Methods of assessing the pluripotency of a cell are known in the art. Example 2 below describes exemplary methods that can be used to evaluate the potency of an ES cell. In one example, ES cells are injected into mouse blastocysts, transferred to uteri and the extent of ES cell potency is determined by the percent chimerism of the pups based on coat color. In another example, a 4N complementation assay is performed. In this assay, ES cells are injected into a tetraploid (4N) blastocyst. Potency of the ES cells is determined by the ability of the ES cells to produce live embryos.

[0154] In some examples, the pluripotency of a stem cell or a stem cell population is increased by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or at least 100%, as compared to the pluripotency of a stem cell or a stem cell population in the absence of increased Zscan4 activation frequency (such as in the absence of expression of an Zscan4-ERT2 fusion protein).

2. Increasing the frequency of Zscan4$^+$ cells in a population

[0155] Also provided herein is a method of increasing the frequency of Zscan4-positive cells in a stem cell population. In some embodiments, the method includes contacting the stem cell population with a nucleic acid molecule or vector encoding a Zscan4-ERT2 fusion protein disclosed herein. In other embodiments, the method includes contacting the stem cell population with a Zscan4-ERT fusion protein disclosed herein.

[0156] In yet other embodiments, the method includes contacting the stem cell population with a nucleic acid molecule or vector encoding a Zscan4-ΔC protein disclosed herein. In other embodiments, the method includes contacting the stem cell population with a Zscan4-ΔC protein disclosed herein.

[0157] Methods of delivering nucleic acid molecules encoding Zscan4-ERT2 or Zscan4-ΔC, and Zscan4-ERT2 or Zscan4-ΔC proteins to stem cells are known in the art and are described above.

[0158] Methods of detecting Zscan4$^+$ cells in a cell population are routine and have been previously described (see for example, PCT Publication No. WO 2008/118957, herein incorporated by reference). For example, antibodies specific for Zscan4 (which are commercially available or can be produced according to standard procedures) can be used in immunological based assays to detect Zscan4$^+$ cells. For instance, fluorescence-activated cell sorting can be used to detect and quantify Zscan4+ cells in a population. As another example, a Zscan4 reporter construct can be used to detect expression of Zscan4 (such as the pZscan4-Emerald vector as described in PCT Publication No. WO 2008/118957).

[0159] In particular examples, the increase in frequency of Zscan4$^+$ cells in the population is an increase of at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 50%, at least 75%, at least 90% or at least 100%. The increase is relative to, for example, a population of cells that has not been contacted with a Zscan4-ERT2

nucleic acid or fusion protein, or a Zscan4-ΔC nucleic acid or protein (and thus has not undergone frequent activation of Zscan4).

3. Promoting genome stability and increasing telomere length

[0160] Methods of promoting genome stability or increasing telomere length, or both, in a stem cell or a stem cell population are further provided. In some embodiments, the method includes contacting the stem cell or stem cell population with a nucleic acid molecule or vector encoding a Zscan4-ERT2 fusion protein disclosed herein. In other embodiments, the method includes contacting the stem cell or stem cell population with a Zscan4-ERT fusion protein disclosed herein.

[0161] In yet other embodiments, the method includes contacting the stem cell or stem cell population with a nucleic acid molecule or vector encoding a Zscan4-ΔC protein disclosed herein. In other embodiments, the method includes contacting the stem cell or stem cell population with a Zscan4-ΔC protein disclosed herein.

[0162] Methods of delivering nucleic acid molecules encoding Zscan4-ERT2 or Zscan4-ΔC, and Zscan4-ERT2 or Zscan4-ΔC proteins to stem cells are known in the art and are described above.

[0163] In particular examples, genome stability is increased in a stem cell by at least 20%, at least 40%, at least 50%, at least 60%, at least 75%, at least 80%, at least 90%, at least 95%, or at least 98%, for example relative to stem cell that has not been contacted with a Zscan4-ERT2 or Zscan4-ΔC protein or a nucleic acid encoding a Zscan4-ERT2 or Zscan4-ΔC protein (or compared to a value or range of values expected in a stem cell that has not undergone frequent activation of Zscan4). Methods of measuring genome stability and telomere length are routine in the art, and the disclosure is not limited to particular methods. The particular examples provided herein are exemplary.

[0164] In some examples, genome stability in a stem cell is measured by detecting cell proliferation. Genome stability is increased if cell proliferation is increased, for example relative to a control cell (for example, a stem cell that has not been contacted with a Zscan4-ERT or Zscan4-ΔC protein or nucleic acid). For example, ES cell proliferation can be detected by growing ES cells in culture and measuring the doubling time of the cells after each passage. In

one example, genome stability is increased if crisis (*e.g.*, cell death) does not occur at passage 8 or earlier.

[0165] In some examples, genome stability in a stem cell, such as an ES cell or iPS cells, is measured by performing karyotype analysis. Genome stability is increased if the presence of karyotype abnormalities (such as chromosome fusions and fragmentations) is decreased or even absent, for example relative to a cell that has not undergone frequent activation of Zscan4. For example, karyotype analysis can be performed in stem cells by inducing metaphase arrests, then preparing metaphase chromosome spreads.

[0166] In some examples, genome stability in stem cell is measured by measuring telomere sister chromatid exchange (T-SCE). Genome stability is increased if the presence of T-SCE is increased relative to a control (such as a stem cell that has not undergone frequent activation of Zscan4). For example, T-SCE can be measured in an stem cell by using telomere chromosome-orientation FISH (CO-FISH).

[0167] In some examples, genome stability in stem cell is measured by measuring sister chromatid exchange (SCE). Genome stability is increased if the presence of SCE is decreased relative to a control, such as a stem cell that has not undergone frequent activation of Zscan4. For example, SCE can be measured in a stem cell by detecting SCE in a metaphase spread.

[0168] In some examples, telomere length is measured in stem cell. Telomere length is increased in a stem cell if the length of the telomeres is greater, for example relative to telomere length in a control cell that has not undergone frequent activation of Zscan4 (such as a cell that has not been contacted with a Zscan4-ERT2 or Zscan4-ΔC protein or nucleic acid). For example, telomere length can be detected in a stem cell by fluorescence *in situ* hybridization (FISH), quantitative FISH (Q-FISH), or telomere qPCR.

IV. Increasing the Frequency of Homologous Recombination

[0169] One of the unique features of mouse ES cells is the high efficiency of homologous recombination (HR), which has been widely used to carry out gene targeting in mouse ES cells. In higher eukaryotes, except for a chicken DT40 cell line derived from B lymphocytes, HR occurs relatively infrequently. The mechanism by which HR is facilitated in mouse ES cells is not well understood. It is disclosed herein that HR events occur more frequently in ES cells carrying a

Zscan4-expressing transgene. These data indicate a means of enhancing HR frequency by manipulating Zscan4 expression. These finding have several applications, as discussed below.

A. Manipulation of mouse genome by gene targeting technology in ES, iPS, and other cell types

[0170] CGZ3 ES cells described above or any other mouse ES cells carrying the similar plasmid construct can be immediately used to increase the efficiency of making gene-manipulated ("knockout") mice. For example, CGZ3 ES cells can be transfected with any kind of a gene-targeting vector, which will be integrated into the desired region of the mouse genome with high efficiency. Previously described methods require isolating and testing hundreds of ES cell colonies due to the extremely low efficiency of the homologous recombination-targeting event. This is a labor intensive step, which typically takes at least several months to complete. By dramatically increasing the homologous recombination efficiency, the method disclosed herein will reduce the cost and time for this effort. The same strategy can be applied to any type of mouse cell, which are difficult to use for gene manipulation by homologous recombination.

B. Manipulation of the human genome by gene targeting technology in ES, iPS, and other cell types

[0171] The strategy described above can also be applied to human ES, iPS, and other cell types. There is increasing interest in homologous gene targeting, particularly in human ES and iPS cells. The current paradigm for the therapeutic application of human ES and iPS cells requires the generation of patient-specific iPS cells from patient fibroblast cells, correcting the genetic defects such as disease-causing DNA mutations in iPS cells, differentiating the corrected iPS cells into desired cell types (such as neurons or cardiomyocytes) and transplanting these cells/tissues back into the patient. As human ES/iPS cells possess a much lower efficiency of homologous recombination, correcting the mutations in iPS cells is currently considered a great technical challenge. The method disclosed herein will solve this problem by dramatically increasing the efficiency of homologous recombination in human iPS cells.

C. Gene therapy

[0172] Gene therapy (i.e., correcting DNA mutations in patients' cells/tissues/organs) has been one of the major goals in medicine for many years. To this end, homologous recombination-mediated correction of DNA has been extensively studied. However, currently there is no safe method of efficiently correcting a patient's DNA mutations. As disclosed herein, Zscan4-mediated enhancement of homologous recombination can be applied to the efficient DNA

correction for gene therapy. For example, a Zscan4-ERT2 -expressing vector or Zscan4-ERT2 protein (e.g., in the nanoparticle-encapsulated form) with a gene-targeting vector can be transiently transferred to the patients' tissue/organs to correct DNA mutations.

D. Methods of enhancing the efficiency of homologous recombination

[0173] Provided herein is a method of enhancing the efficiency of homologous recombination in a mammalian cell. In some examples, HR is increased by at least 25-fold, at least 50-fold, at least 75-fold, at least 90-fold, or even at least 100-fold, relative to the absence of Zscan4 (such as a Zscan-4 nucleic acid molecule, protein, or activator), such as Zscan4-ERT2. In some embodiments, the method includes contacting the cell with a Zscan4-ERT2 nucleic acid molecule or fusion protein disclosed herein. In some embodiments, the method includes delivery of a vector encoding the Zscan4-ERT2 fusion protein. In particular examples, the cell is a mouse cell or a human cell. In some examples, the cell is an ES cell or an iPS cell. In one non-limiting example, the nucleic acid molecule or vector encoding the Zscan4-ERT2 fusion protein, or the Zscan4-ERT2 fusion protein, is encapsulated in a nanoparticle. The disclosed methods can be carried out *in vitro* or *in vivo*.

[0174] Further provided is an *in vitro* or *in vivo* method of enhancing the efficiency of homologous recombination in a mammalian cell. In some embodiments, the method includes contacting the cell with a recombinant Zscan4 protein or a recombinant nucleic acid molecule encoding a Zscan4 protein, thereby enhancing efficiency of homologous recombination in the cell. In some examples, the Zscan4 protein comprises a Zscan4-ERT2 fusion protein.

[0175] Also provided is a method of correcting a genetic mutation in a cell, tissue or organ of a subject by contacting the cell, tissue or organ with (1) a Zscan4-ERT2 fusion protein or a nucleic acid molecule encoding a Zscan4-ERT2 fusion protein; and (2) a gene-targeting vector, thereby correcting the genetic mutation in the cell, tissue or organ. In some embodiments, the Zscan4-ERT2 fusion protein is encapsulated in a nanoparticle. In some embodiments, the cell is isolated from the subject and the method is performed *in vitro*. In other embodiments, the method is an *in vivo* method. The gene targeting vector can include any gene of interest containing a mutation in a cell, tissue or organ of a subject.

[0176] 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: Materials and Methods

[0177] This example describes the experimental procedures used for the studies described in Example 2.

ES cell culture

[0178] MC1 ES cells derived from 129S6/SvEvTac and MC2 ES cells derived from C57BL/6J (Olson *et al., Cancer Res* 63:6602-6606, 2003) were purchased from the Transgenic Core Laboratory of the Johns Hopkins University School of Medicine (Baltimore, MD). V6.5 ES cells (Eggan *et al., Proc Natl Acad Sci USA* 98:6209-6214, 2001) derived from an F1 hybrid strain (C57BL/6 x 129/Sv) were purchased from Thermo Scientific Open Biosystem. All ES cell lines, except for TA1 ES cell line (see below), were cultured at 37°C in 5% CO^2 in the complete ES medium as previously described (Zalzman *et al., Nature* 464:858-863, 2010): DMEM (Gibco), 15% FBS (Atlanta Biologicals), 1000 U/ml leukemia inhibitory factor (LIF) (ESGRO, Chemicon), 1 mM sodium pyruvate, 0.1 mM non-essential amino acids (NEAA), 2 mM GlutaMAX™, 0.1 mM beta-mercaptoethanol, and penicillin/streptomycin (50 U/50 μg/ml). TA1 ES cell lines were cultured as described above. For all cell lines, media was changed daily and cells were passaged every 2 to 3 days routinely.

Derivation of TA1 ES cell line

[0179] C57BL/6J females (The Jackson Laboratory, Bar Harbor, ME) and 129S6/SvEvTac males (Taconic) were naturally mated to collect 2-cell embryos, which were then cultured in KSOM medium for 3 days at 37°C in 5% CO^2. Resulting blastocysts were transferred onto mouse embryo fibroblast (MEF) feeder cells treated with mitomycin C (Sigma) and cultured for 7 days in the complete ES medium (described above) after replacing 15% FBS with 15% KSR (Invitrogen) and adding 50 nM PD98059 (MEK1 inhibitor). After picking inner cell mass (ICM) clumps and dissociating them by ACCUTASE™ (Millipore), they were seeded onto fresh feeder cells and cultured in the same condition for an additional 7 days. Newly derived ES cell lines were directly tested for their developmental potency by 4N-complementation (see below).

pCAG-Zscan4-ERT2 vector construction

[0180] Genes collectively called Zscan4 consist of 6 paralogous genes and 3 pseudogenes clustered on a ~850 kb region of chromosome 7 (Falco *et al., Dev Biol* 307:539-550, 2007). Among six paralogs named Zscan4a to Zscan4f, the open reading frames (ORFs) of Zscan4c, Zscan4d, and Zscan4f are very similar to each other and encode a SCAN domain and four zinc finger domains (Falco *et al., Dev Biol* 307:539550, 2007). To construct a pCAG-Zscan4-ERT2 plasmid, an entire ORF (506 a.a.) of mouse Zscan4c gene (Falco *et al., Dev Biol* 307:539-550, 2007) was fused with ERT2 (Feil *et al., Proc Natl Acad Sci USA* 93:10887-10890, 1996) (314 a.a.) and cloned into XhoI/NotI sites of pPyCAG-BstXI-IP (Niwa *et al., Gene* 108:193-199, 1991). The resultant plasmid vector expresses Zscan4c-ERT2 fusion protein-IRES-puromycin-resistant protein under a strong CAG promoter.

Generation of ZE and ZERT2 ES cell clones

[0181] ES cells were grown in 6-well plates. For ZE ES cell clones, 5 x 10^5 ES cells in suspension were transfected with 1 μg of a linearized pZscan4-Emerald vector (Zalzman *et al., Nature* 464:858-863, 2010) using EFFECTENE™ (QIAGEN) according to manufacturer's protocol, and plated in 100 mm dishes. After selecting with 5 μg/ml blasticidin for 8 days, resulting ES cell colonies were picked, expanded, and frozen for further analysis. For ZERT2 ES cell clones, 5 x 10^5 ES cells in suspension were cotransfected with 0.5 μg of a linearized pCAG-Zscan4-ERT2 vector and 0.5 μg of PL452 (PGK promoter-Neo) (Liu *et al., Genome Res* 13:476-484, 2003) using EFFECTENE™ (QIAGEN) according to manufacturer's protocol, and plated in 100 mm dishes. After selecting with G418 for 8 days, resulting ES cell colonies were picked, expanded, and frozen for further analysis.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

[0182] RNA was isolated from cells by TRIZOL™ (Invitrogen) in biological triplicate. One μg of total RNA was reverse transcribed by SuperScript™ III (Invitrogen) following the manufacturer's protocol. 100 ng of oligo dT primers (Promega) was used per reaction. For qPCR, SYBR™ green master mix (Applied Biosystems) was used following the manufacturer's protocol. 96-well optical plates with a 25 μl total reaction volume were used, 10 ng of cDNA was used per well. Plates were run on 7300 or 7500 system (Applied Biosystems). Fold induction was calculated by the ΔΔCt method (Livak *et al., Methods* 25:402-408, 2001) using H2A as normalizer.

RNA isolation, cDNA preparation and qPCR analysis in mouse preimplantation embryos

[0183] Four to six week-old B6D2F1 female mice were superovulated with 5 I.U. of PMSG (Sigma) and 5 I.U. of human chorionic gonadotropin (hCG) (Sigma). Eggs or embryos for qRT-PCR experiments were collected after 20, 23, 30, 43, 55, 66, 80 and 102 hours post hCG injection for MII (unfertilized oocytes), 1-cell, early and late-2 cell, 4-cell, 8-cell, morula and blastocyst embryos, respectively. Three sets of 10 synchronized eggs or embryos were stored in liquid nitrogen and mechanically ruptured by a freeze/thaw step for the cDNA preparation template. Oligo-dT primers and SuperScript™ III reverse transcriptase (Invitrogen) were used according to the manufacturer's instruction. Analysis was performed on the ABI 7300 Fast Real Time PCR system (Applied Biosystems). Data was normalized by Chuk (Falco *et al., Reprod Biomed Online* 13:394-403, 2006) with the ΔΔCt method (Livak *et al., Methods* 25:402-408, 2001).

RNA *in situ* hybridization

[0184] Whole mount *in situ* hybridization was performed as previously described (Carter *et al., Gene Expr Patterns* 8:181-198, 2008). Briefly, ES cells in triplicates, grown for 3 days, were fixed in 4% paraformaldehyde (PFA) at 4°C overnight. After digestion with proteinase K, cells were hybridized with 1 μg/ml digoxigenin-labeled riboprobe at 62°C overnight. Cells were then washed, blocked, incubated with alkaline phosphatase-conjugated anti-digoxigenin antibody, and incubated with NBT/BCIP detection buffer for 30 minutes or overnight.

Double-fluorescence RNA *in situ* hybridization

[0185] Digoxigenin (DIG)- and biotin (BIO)-labeled RNA probes were transcribed from the PCR product templates using RNA Labeling Mix (Roche). Ethanol-precipitated probes were resuspended in water and quantified by RNA 6000 Nano Assay on a 2100 Bioanalyzer (Agilent Technologies). 10^5 cells/well were seeded in glass chamber slides, cultured for three days, fixed with PFA, and permeabilized with 0.5% TritonX-100. Cells were washed and incubated with 1 μg/ml DIG and BIO probes for 12 hours at 60°C in hybridization solution. Probes were detected by mouse anti-DIG antibody and by sheep anti-BIO, and visualized by fluorophore-conjugated secondary antibodies. Nuclei were stained with DAPI (blue).

Microarray analysis

[0186] DNA microarray analyses were carried out as described (Aiba *et al., DNA Res* 16:73-80, 2009). Briefly, universal Mouse Reference RNA (Stratagene) were labeled with Cy5-dye,

mixed with Cy3-labeld samples, and used for hybridization on the NIA Mouse 44K Microarray v2.2 (Carter *et al., Genome Biol* 6:R61, 2005) (manufactured by Agilent Technologies #014117). The intensity of each gene feature was extracted from scanned microarray images using Feature Extraction 9.5.1.1 software (Agilent Technologies). Microarray data analyses were carried out by using an application developed in-house to perform ANOVA and other analyses (NIA Array Analysis software; online at lgsun.grc.nia.nih.gov/ANOVA/) (Sharov *et al., Bioinformatics* 21:2548-2549, 2005). All the DNA microarray data have been deposited to the NCBI Gene Expression Omnibus (GEO, online at www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number (GSE26278) and the NIA Array Analysis software website (online at lgsun.grc.nia.nih.gov/ANOVA/) (Sharov *et al., Bioinformatics* 21:2548-2549, 2005). For GEO reviewer link: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=fhaxtmiueykigvm&acc=GSE26278.

ES cell injection into 2N or tetraploid (4N) blastocysts

[0187] CD1 females (Charles River, 8-12 week old) were used for superovulation by PMSG (Sigma) followed 48 hours later by hCG (Sigma) administration. After hCG administration, females were mated with males of the same strain 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 0.3 M mannitol solution and aligned automatically by alternate current (AC) pulse in an electrofusion chamber. Then two direct current (DC) pulses with 140V/mm were applied for 40 μs using LF101 Electro Cell Fusion Generator. Fused embryos (4N) that had one blastomere were collected at 60 minutes cultivation and then culture continued in KSOM medium until they reached the blastocyst stage. A single ES cell or 10-15 ES cells were injected into 2N or 4N blastocysts to assess their developmental potency and then transferred to E2.5 recipient females. To study the effects of Tmx on ES cells, ES cells were cultured in the presence of 200 nM Tmx for 2-3 days before injection.

Example 2: Rejuvenation of pluripotent stem cells by frequent activation of Zscan4

[0188] This example describes the finding that increasing the frequency of Zscan4 activation in mouse ES cells not only enhances, but also maintains their developmental potency in long-term cell culture.

Commonality between transient Zscan4+ state and 2-cell stage embryos

[0189] As a first step to characterize the Zscan4$^+$ state of ES cells, global gene expression profiles were compared between Zscan4$^+$ and Zscan4$^-$ state of ES cells. In an earlier study, a reporter cell line, pZscan4-Emerald cells (hereafter called "MC1-ZE"), was established in which a Zscan4c-promoter-driven reporter green fluorescence protein GFP-Emerald (Em) recapitulates the expression of endogenous Zscan4 (Zalzman *et al., Nature* 464:858-863, 2010). DNA microarray analysis of FACS-sorted Em$^+$ and Em$^-$ cells was carried out. Em$^+$ cells showed a very similar gene expression profile to the Em$^-$ cells with only 161 differentially expressed genes (FIG. 5; see also PCT Publication No. WO 2008/118957 and Falco *et al., Dev Biol* 307:539-550, 2007). Pluripotency-related markers remained unchanged in Em$^+$ cells compared to Em -cells, but Tcstv1 and Tcstv3 (two cell-specific transcript variant 1 and 3) genes (Struwe and Solter, GenBank accession AF067057.1; Zhang *et al., Nucleic Acids Res* 34:4780-4790, 2006) were among the most highly upregulated genes (FIG. 5). RNA whole-mount *in situ* hybridization revealed "Zscan4-like" expression for 7 other genes in the list (Tcstv1/3, Eif1a, Pif1, AF067063, EG668777, RP23-149D11.5, BC061212, and EG627488; see PCT Publication No. WO 2008/118957, herein incorporated by reference).

[0190] Furthermore, double-label fluorescence RNA *in situ* hybridization confirmed coexpression of these genes with Zscan4. As Zscan4 is a 2-cell embryo marker (Falco *et al., Dev Biol* 307:539-550, 2007), 6 genes were selected from the list based on additional gene expression information in preimplantation embryos (Ko *et al., Development* 127:1737-1749, 2000; Sharov *et al., PLoS Bio*l 1:E74, 2003) and were examined for their expression profiles by qRT-PCR. All six genes tested showed a high expression peak in 2-cell embryos: 2 genes showed the highest peak at the late 2-cell stage as Zscan4, whereas 4 others showed their highest peak at the early 2-cell stage (see PCT Publication No. WO 2008/118957, herein incorporated by reference). Considering the fact that a large-scale screening of ~250 transcription factor genes by whole-mount *in situ* hybridization identified only two other genes (Rhox9 and Whsc2) with a "Zscan4-like" expression pattern (Carter *et al., Gene Expr Patterns* 8:181-198, 2008), the high incidence of finding 2-cell genes with a Zscan4-like expression pattern in ES cells suggests that some of the gene expression program in early-stage embryos are reactivated in the Zscan4$^+$ state of ES cells.

Transient Zscan4$^+$ state is not associated with higher developmental potential

[0191] ES cells are thought to be equivalent to cells in the inner cell mass (ICM) of blastocysts (Nichols and Smith, *Development* 138:3-8, 2011; Yoshikawa *et al., Gene Expr*

Patterns 6:213-224, 2006). Commonality between Zscan4$^+$ state and 2-cell embryos suggest that in standard cell culture conditions, ES cells are a mixed population of ~5% of 2-cell like cells and ~95% of ICM-like cells. As it has been shown that by nuclear transplantation (cloning) the 2-cell nucleus has a higher developmental potential than the ICM nucleus (Tsunoda *et al., Development* 107:407-411, 1989; Kono *et al., J Reprod Fertil* 93:165-172, 1991), the Zscan4$^+$ state may represent high-potential true stem cells among the regular ES cell population.

[0192] To test this notion, V6.5 ZE cells (clone #17) were generated and their developmental potency was assessed by transfecting a pZscan4-Emerald vector into V6.5 ES cells derived from an F1 hybrid strain (C57BL/6 x 129/Sv), which has been extensively used for testing developmental potency (Eggan *et al., Proc Natl Acad Sci USA* 98:6209-6214, 2001; Wang and Jaenisch, *Dev Biol* 275:192-201, 2004). To avoid cell damage caused by cell sorting or long UV exposure, Em$^+$ or Em$^-$ cells were separated manually by pipetting, single ES cells were injected into 2N blastocysts, and the subsequent embryo development was observed. Based on the coat colors, it was found that Em$^-$ ES cells were able to contribute to the tissues of chimeric mice at a relatively high rate (31%), whereas Em$^+$ ES cells were not (0%). The results indicate that, contrary to expectations, Zscan4$^+$ cells are not associated with high developmental potency compared to Zscan4 cells.

Zscan4-ERT2 increases the frequency of endogenous Zscan4$^+$ cells in the absence of Tmx

[0193] Intermittent and transient activation of Zscan4 is required for the long-term maintenance of ES cell cultures (Zalzman *et al., Nature* 464:858-863, 2010). It was therefore hypothesized that more frequent activation of Zscan4 further improves the quality of ES cells, including their developmental potency. A system to mimic the transient expression of Zscan4 was sought. To this end, ERT2, the tamoxifen (Tmx)inducible system was selected (Feil *et al., Proc Natl Acad Sci USA* 93:10887-10890, 1996). This system allows one to keep a transgene off in the absence of Tmx and turn it on in the presence of Tmx at will (Feil *et al., Proc Natl Acad Sci USA* 93:10887-10890, 1996). First, the plasmid construct pCAG-Zscan4-ERT2 was made in which Zscan4c open reading frame (ORF) fused with ERT2 domain can be driven by a strong ubiquitous promoter CAG (Niwa *et al., Gene* 108:193-199, 1991) (FIG. 1A).

[0194] When the pCAG-Zscan4-ERT2 plasmid was transfected into MC1-ZE3 cells, it was found that the constitutive expression of Zscan4-ERT2 in ES cells increased the fraction of Em$^+$ cells even in the Tmx$^-$ condition (FIG. 1B). Adding Tmx to the culture media further increased the fraction of Em$^+$ cells, but also made ES cells (both Em$^+$ and Em$^-$ cells) flatter, resulting in the

flattening of ES cell colonies – a deviation from the typical pluripotent ES colony morphologies (FIG. 1B). The results were further confirmed by quantitative assays for five independent clones: the constitutive expression of Zscan4-ERTs even in the absence of Tmx caused a 3-fold increase of Em^+ cells by the flow cytometry analysis (FIG. 1C) and 5-fold increase by the qRT-PCR analysis (FIG. 1D); and addition of Tmx to the medium caused further 2-fold and 1.2-fold increase, respectively (FIGS. 1C-1D).

[0195] To further investigate this unexpected result, the pCAG-Zscan4-ERT2 plasmid was transfected into V6.5 ES cells (Eggan *et al., Proc Natl Acad Sci USA* 98:62096214, 2001) and multiple cell clones named V6.5 ZERT2 were isolated. Based on the qRT-PCR analysis of Zscan4 ORF, clone #18 was selected for the highest Zscan4 expression levels, clones #7 and #10 were selected for the second and third highest Zscan4 levels, and clone #2 was selected with the background Zscan4 level (FIG. 6A). Based on genotyping by PCR, clone #2 did not have any copies of the pCAG-Zscan4-ERT2 plasmid, and was thus used as a control (V6.5 #2). As expected, Tmx^+ conditions slowed down the proliferation of ES cells (FIG. 6B) and made ES cells flatter (FIG. 6C). When the Tmx was removed from the medium after 10 passages in the Tmx+ conditions, the cell proliferation and morphology returned to normal (FIGS. 6B-6C), suggesting that effects of Tmx on the V6.5 ZERT2 cells were reversible.

[0196] To check if the frequency of Zscan4+ cells is increased even in the Tmx^- condition, whole mount *in situ* hybridization was carried out using a full-length Zscan4c probe to detect both endogenous and exogenous copies of Zscan4 as well as a 3'-UTR Zscan4c probe to detect only endogenous Zscan4. The results showed ~3-fold increase of the number of $Zscan4^+$ cells in V6.5 ZERT2 ES cell clones (#7, #10, and #18) in the absence of Tmx compared to the usual level of $Zscan4^+$ cells in the control cells (V6.5 and V6.5 #2) (FIG. 1E). Further comparison of global gene expression profiles by DNA microarrays confirmed that the expression of Zscan4 was upregulated by 3.6-fold in V6.5 ZERT2 #18 ES cells even in the Tmx -condition (FIGS. 7 and 8). Similarly, other key Zscan4-related genes identified in *Falco et al. (Dev Biol* 307:539-550, 2007), such as Tcstv1, Tcstv3, Tmem92, RP23-149D11.5, and BC061212, were also upregulated in V6.5 ZERT2 #18 ES cells in the Tmx -condition (FIG. 1F, FIG. 7, FIG. 8). Adding Tmx increased the expression of Zscan4 and other Zscan4-related genes only slightly, but increased that of Zscan4-unrelated genes significantly (FIG. 1G, FIG. 7 and FIG. 9).

[0197] Taken together, use of constitutively expressing Zscan4-ERT2 without Tmx became an unexpected, but attractive strategy to enhance the naturally occurring Zscan4 effects by increasing the number of endogenous Zscan4$^+$ cells.

Zscan4 protein lacking the C-terminus (Zscan4c-ΔC) increases the number of Zscan4$^+$ cells

[0198] Based on the results described above, it was hypothesized that the effect of ERT2 was due to blocking the function of the Zscan4 zinc finger domains at the C-terminus of the protein. Thus, to evaluate whether C-terminally truncated Zscan4 has the same effect as Zscan4-ERT2 of inducing recurrent activation of Zscan4, vectors encoding either C-terminal truncated (lacking all four zinc finger domains) or N-terminal truncated (lacking the SCAN domain) Zscan4 were constructed. FIG. 2A provides a schematic of the structure of Zscan4c, Zscan4c-ERT2, Zscan4c-ΔC and Zscan4c-ΔN proteins. The amino acid sequence of Zscan4c-ΔC is set forth herein as SEQ ID NO: 25.

[0199] The mutated Zscan4c genes were placed under the strong and constitutive CAG promoter. The sequence of the pCAG-Zscan4-ΔC vector is set forth herein as SEQ ID NO: 24. Each vector was transfected into MC1-ZE16 ES cells (sister clones of MC1ZE3). Multiple independent clones were isolated: ZDC-MC1-ZE16 #3, #4, #20 for Zscan4c-ΔC; ZDN-MC1-ZE16 #5, #8, #15 for Zscan4c-ΔN. Fluorescence microscopy was performed on each cell clone. The images of ZDC-MC1-ZE16 #3, #4, #20 and ZDN-MC1-ZE16 #5, #8, #15 are shown in FIGS. 2B-2G. The results clearly show that the expression of Zscan4c-ΔC increases the number of Zscan4$^+$ cells, whereas the expression of Zscan4c-ΔN does not change the number of Zscan4$^+$ cells. The results indicate that Zscan4c-ΔC functions in a manner similar to Zscan4-ERT2 (Tmx-condition).

Zscan4-ERT2 enhanced and prolonged developmental potency of ES cells in the absence of Tmx

[0200] To assess the effects of Zscan4-ERT2 on the developmental potency of ES cells, various ES cells were injected into mouse blastocysts, transferred to uteri, and their development was followed. The extent of ES cell potency was assessed by the percent chimerism in the pups based on coat colors: high (>70% chimerism), moderate (40%-70%), low (<40%), and albino (0%) (FIG. 3A).

[0201] A V6.5 parental ES cell line at its early passage (p15) showed 18% high, 29% moderate, and 41% low chimerism, which are within the standard range for F1 hybrid ES cell

lines. It is known that the developmental potency of ES cells generally becomes lower after multiple passages and/or plasmid transfection/drug selection. As expected, compared to a V6.5 parental ES cell line, a control V6.5 #2 ES cell line, which did not carry Zscan4-ERT2 but was generated after transfection and drug selection, showed a slightly lower overall potency, which was further reduced over multiple passages (p21, p23, and p30) (FIG. 3B). By contrast, V6.5 ZERT2 #18 ES cells showed much higher developmental potency than parental V6.5 and control V6.5 #2 ES cells: 73% high and 27% moderate chimerism at passage 19 (FIG. 3B). Even more surprising was that such a high level of potency was maintained for an extended period of time and passages: for example, even at passage 30, more than 40% of pups derived from V6.5 ZERT2 #18 ES cells showed "high" chimerism, whereas none of the pups derived from control V6.5 #2 ES cells showed "high" chimerism (FIG. 3B). Five other ES cell lines of different genetic backgrounds and transgenes were tested, including a very early passage line from freshly isolated ES cells (TA1). Potency-wise none of these ES cell lines could even come close to V6.5 ZERT2 #18 cell lines (FIG. 3B).

[0202] Interestingly, the exposure to Tmx for 2 to 3 days lowered the potency of both V6.5 #2 and V6.5 ZERT2 #18 ES cells relative to that in the Tmx -condition, although the V6.5 ZERT2 #18 ES cells still showed higher potency than V6.5 #2 ES cells (FIG. 3B). These results seem to be consistent with the observation made by the global expression profiling (FIG. 1F): Tmx+ conditions increased the expression of genes unrelated to naturally occurring Zscan4^{+} (i.e., Em^{+}) state in V6.5 ZERT2 #18 ES cells.

Testing developmental potency of ES cells by the 4N complementation assay

[0203] It is widely recognized that the ultimate test for developmental potency is to see if ES cells alone injected into tetraploid (4N) blastocysts become an entire mouse (Nagy *et al., Development* 110:815-821, 1990). Compared to early passage V6.5 ES cells reported previously, which has achieved 15-25 % pups alive at term (Eggan *et al., Proc Natl Acad Sci USA* 98:6209-6214, 2001), V6.5 ES cells at passage 18 only produced 2% live embryos (FIG. 4A). By contrast, V6.5 ZERT2 #18 ES cells even at passage 19 showed a much higher success rate – 43% live embryos (FIGS. 4A and 4C). Similarly, two other independent clones (V6.5 ZERT2 #7; V6.5 ZERT2#10) also showed a high success rate of producing live embryos when 10-15 ES cells were injected into 4N blastocysts (FIG. 4A).

[0204] To compare the high success rate of V6.5 ZERT2 #18 cells with those of the best ES cells possible, freshly isolated ES cells were established from blastocysts with the same genetic

background – F1 hybrid of C57BL/6J x 129S6/SvEvTac and were cultured in the best conditions currently available (Wong *et al., Methods Enzymol* 476:265-283, 2010) (FIG. 10 and FIG. 11). Of 20 blastocysts, 19 formed outgrowths *in vitro*, 13 of which continued to be cultured for an additional 7 days to form ES cell colonies, resulting in newly established ES cell lines (FIG. 10). Six clones out of 13 ES cell lines at the earliest passages (p3) were tested for their potency by injecting 10-15 ES cells into 4N blastocysts: one ES line, named "TA1", showed the highest efficiency (60%) of producing live embryo at E13.5 (FIG. 4A and FIG. 11). Overall, these results obtained by the 4N complementation assays indicate that the developmental potency of V6.5 ZERT2 #18 ES cells even at the higher passage number is comparable to that of freshly isolated early passage ES cells.

[0205] To exclude the possibility that Zscan4-ERT2 affects only F1 hybrid ES cell lines, MC2 ZERT2 #6 ES cells were generated by transfecting a Zscan4-ERT2 plasmid to an MC2 ES cell line (C57BL/6J) (Olson *et al., Cancer Res* 63:6602-6606, 2003). Consistent with the reported low potency of C57BL/6J-derived ES cells (Brook *et al., Proc Natl Acad Sci USA* 94:5709-5712, 1997; Eggan *et al., Proc Natl Acad Sci USA* 98:6209-6214, 2001), both MC2 ES cells at passage 17 and genetically modified MC2 ES cells at passage 12-13 did not produce any live embryos (FIG. 4A). By contrast, MC2 ZERT2 #6 ES cells, which were cultured for more than 10 passages with the constitutive expression of Zscan4-ERT2, successfully achieved the production of 6% live embryos (FIG. 4A). The results thus suggest that the Zscan4-ERT2 construct can be used as a universal tool to enhance the developmental potency of pluripotent stem cells.

[0206] The unusually high developmental potency of V6.5 ZERT2 #18 cells prompted the further examination of the potency of single ES cells. It has been shown once that even a single ES cell can form a live pup, although the success rate is extremely low (1 mouse/192 injected blastocyst: 0.5%) (Wang and Jaenisch, *Dev Biol* 275:192-201, 2004). As expected from the fact that the same cell line was used as for the earlier study (Wang and Jaenisch, *Dev Biol* 275:192-201, 2004), the injection of a single parental V6.5 ES cell at passage 18 into 4N blastocysts produced one live embryo (1%) (FIG. 4B). Furthermore, single control V6.5 #2 ES cells did not produce any live embryos after injecting them into 77 tetraploid blastocysts (FIG. 4B). By contrast, of 44 tetraploid blastocysts that received a single V6.5 ZERT2 #18 cell, 3 (7%) became complete embryos, 2 (5%) of which were alive at the time of dissection (FIGS. 4B and 4D). This unusually high level of potency for V6.5 ZERT2 #18 ES cells was indeed comparable to that of early passage TA1 ES cells with 4% live embryos (FIG. 4B).

Discussion

[0207] It is disclosed herein that the constitutive presence of Zscan4-ERT2, without its usual activator Tmx, can increase the frequency of endogenous Zscan4 activation, resulting in the increase of developmental potency of ES cells. ES cells cultured in the accelerated Zscan4 activation cycle show improved chimerism and potency, which are demonstrated by high contribution to chimeric mice and efficient production of a whole mouse from a single ES cell.

[0208] How does the frequent activation of Zscan4 enhance and prolong the developmental potency of ES cells? Previously, it was demonstrated that the immortality of ES cells is maintained by an intermittent activation of Zscan4 (Zalzman *et al., Nature* 464:858-863, 2010). The shRNA-mediated continuous repression of Zscan4 makes ES cells undergo culture crisis after multiple cell passages (Zalzman *et al., Nature* 464:858-863, 2010). It is thus conceivable that even in their regular proliferating condition ES cells gradually lose their potency, which is rapidly restored by the transient activation of Zscan4 (Zalzman *et al., Nature* 464:858-863, 2010). Consistent with the notion that drastic changes, including rapid telomere extension by telomere sister chromatid exchange (Zalzman *et al., Nature* 464:858-863, 2010), are occurring in ES cells in Zscan4$^+$ state, Zscan4$^+$ cells (Em$^+$ cells in the experiments described herein) did not produce chimeric animals when injected into blastocysts. In standard ES cells, the interval of transient Zscan4 activation may be longer than ideal; thus, ES cells steadily lose their average potency, irrespective of the occasional activation of Zscan4 (FIG. 4E, upper panel). More frequent activation of Zscan4 by the presence of Zscan4-ERT2 may maintain or even increase ES cell potency (FIG. 4E, lower panel).

[0209] Activation of endogenous Zscan4 by Zscan4-ERT2 without Tmx was unexpected, because ERT2-fusion proteins usually require Tmx for their activation. It is speculated that this may be related to a partial blocking of Zscan4 function, because the ERT2 domain is fused to the C-terminus of Zscan4, near four zinc-finger (C2H2) domains, whereas a SCAN domain is located at the N-terminus (Falco *et al., Dev Biol* 307:539-550, 2007). Considering the fact that Zscan4 should not be constitutively active in ES cells, the unexpected finding of Zscan4-ERT2 function provides an ideal means to increase the intermittent activation of endogenous Zscan4 expression. Irrespective of the mechanism, the presence of Zscan4-ERT2 in ES cells has beneficial effects on the potency of ES cells in long-term culture.

Example 3: Efficient homologous recombination in Zscan4-expressing cells

[0210] This example describes studies to investigate HR events in Zscan4-positive cells. It is demonstrated herein that HR occurs more frequently in ES cells carrying a Zscan4-expressing transgene. These data suggest that enhancing HR frequency can be achieved by manipulating Zscan4 expression.

Methods, Results and Discussion

[0211] FIG. 12 shows experimental procedures for random integration of a plasmid vector into the mouse genome (FIG. 12A) and homologous recombination/integration of a plasmid vector into the mouse genome (FIG. 12B). The random integration vector (pBRCAG-EGFP-IRES-Hyg-pA) carries a CAG promoter, and thus can produce hygromycin-resistant cell colonies when integrated anywhere in the genome. By contrast, the homologous integration vector (pOct4-IRES-HygtkpA) does not carry a promoter sequence, and thus can produce hygromycin-resistant cell colonies only when integrated into a unique site in the mouse genome, which shares homology with the vector sequence (Oct4/Pou5f1 locus; see FIG. 12C for the details of the vector design). In general, the efficiency of homologous integration/recombination is much lower than that of random integration. For example, the efficiency for the random integration was approximately 400-fold higher than that for the homologous integration in mouse ES cells (FIG. 12D).

[0212] Constitutive expression of Zscan4-ERT2 fusion protein in mouse ES cells increased the efficiency of HR. Therefore, the following experiments were designed (1) to confirm this result; (2) to create an ES cell line that can be used for mouse gene targeting (*i.e.*, knockout mice). For the latter, it is desirable to remove the Zscan4-ERT2 expression unit from the ES cells before generating mice from gene-targeted ES cells. To this end, an ES cell line, named CGZ3 ES cell, was made by sequentially transfecting the following two plasmids into the V6.5 ES cell line (FIG. 13A): (a) pfloxedCAG-Zscan4ERT2 -IRES-puro-pA; (b) pCAG-CreGR-IRES-His-pA. The expression unit (a) constitutively produces Zscan4-ERT2 fusion protein. When the cells are cultured in the presence of dexamethasone (Dex), the Cre-GR fusion protein becomes active and excises the expression unit (a), which is flanked by LoxP sites. In this way, the Zscan4-ERT2 expression unit (a) can be completely removed from the ES cells, and thus, it will not interfere with the production of gene-targeted mice. 100% removal of transgene was confirmed in puromycin containing medium with 4-day Dex treatment.

[0213] First, CGZ3 ES cells were split into four 10 cm dishes with or without Dex, cultured for 2 days, passaged into six 10 cm dishes, and cultured for 3 days. As described above, culturing CGZ3 ES cells with Dex removed the Zscan4-ERT2 expression unit entirely. Therefore, Dex- conditions represent the presence of Zscan4-ERT2 expression, whereas Dex+ conditions represent the absence of Zscan4-ERT2 expression. Cells were harvested from each culture condition and transfected by electroporation with 100 μg of pBRCAG-EGFP-IRES-hyg-pA for random integration or pOct4-IRES-HygtkpA. After electroporation, cells were cultured in Dex-conditions for 7 days with hygromycin selection. Cell colonies were visualized by Leishman's stain (FIG. 13C) and the number of colonies was scored (FIG. 13D). To normalize the colony numbers between Dex- and Dex+ conditions, we also measured the number of surviving cells by plating cells after Dex-or Dex+ treatment and mock electroporation in dilution series, culturing them for 7 days, and scoring them after Leishman's staining (FIGS. 13E and 13F). The efficiency of homologous recombination and random integration was then calculated by dividing the number of colonies (FIG. 13D) by the number of surviving cells (FIG. 13F). The results shown in FIG. 13G indicate the following:

[0214] (1) In the presence of Zscan4-ERT2 (Dex-condition), the efficiency of homologous recombination was significantly increased and only 3-fold lower than that of random integration. Considering approximately 400-fold lower efficiency of homologous recombination compared to random integration in wild-type regular ES cells (FIG. 12D), this result indicates an approximately 100-fold increase of homologous recombination efficiency.

[0215] (2) In the presence of Zscan4-ERT2 (Dex-condition), the efficiency of random integration is also increased by about 7-fold (compare FIG. 12D and FIG. 12G).

[0216] (3) The removal of Zscan4-ERT2 from the CGZ3 ES cells decreased both random integration efficiency and homologous recombination efficiency by 3-fold. This indicates that the enhancement of integration efficiency is caused by the presence of Zscan4-ERT2. However, the efficiency did not return to the level of that in wild-type ES cells, suggesting that at least some effects of Zscan4-ERT2 last longer than 5 days even after removing the Zscan4-ERT2.

[0217] (4) In summary, these data demonstrate that the expression of Zscan4-ERT2 in ES cells enhances the efficiency of homologous recombination by approximately 100-fold.

[0218] In view of the many possible embodiments to which the principles of the disclosure 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 all that comes within the scope and spirit of these claims.

CLAIMS

1. An isolated nucleic acid molecule encoding a Zscan4-ERT2 fusion protein.

2. The nucleic acid molecule of claim 1, wherein the Zscan4 comprises mouse Zscan4c, mouse Zscan4d, mouse Zscan4f or human ZSCAN4, or a functional fragment or variant thereof.

3. The nucleic acid molecule of claim 1 or claim 2, wherein the Zscan4 comprises mouse Zscan4c or human ZSCAN4.

4. The nucleic acid molecule of any one of claims 1-3, comprising (i) a nucleic acid sequence at least 95% identical to nucleotides 2465-4936 of SEQ ID NO: 19; or (ii) a nucleic acid sequence at least 95% identical to nucleotides 2479-4731 of SEQ ID NO: 20.

5. The nucleic acid molecule of any one of claims 1-3, comprising (i) the nucleic acid sequence of nucleotides 2465-4936 of SEQ ID NO: 19 or (ii) the nucleic acid sequence of nucleotides 2479-4731 of SEQ ID NO: 20.

6. A vector comprising the nucleic acid molecule of any one of claims 1-5.

7. The vector of claim 6, comprising a nucleic acid sequence at least 95% identical to SEQ ID NO: 19 or SEQ ID NO: 20.

8. The vector of claim 6, comprising the nucleic acid sequence of SEQ ID NO: 19 or SEQ ID NO: 20.

9. An isolated cell comprising the nucleic acid molecule or vector of any one of claims 1-8.

10. The isolated cell of claim 9, which is a stem cell.

11. The isolated cell of claim 10, which is an embryonic stem (ES) cell.

12. The isolated cell of claim 10, which is an induced pluripotent stem (iPS) cell.

13. The isolated cell of any one of claims 10-12, wherein the stem cell is a mouse, rat, human or non-human primate stem cell.

14. A composition comprising the nucleic acid molecule or vector of any one of claims 1-8 and a pharmaceutically acceptable carrier.

15. A Zscan4-ERT2 fusion protein encoded by the nucleic acid molecule or vector of any one of claims 1-8.

16. A recombinant Zscan4-ERT2 fusion protein.

17. The fusion protein of claim 16, wherein the Zscan4 comprises mouse Zscan4c, mouse Zscan4d, mouse Zscan4f or human ZSCAN4, or a functional fragment or variant thereof.

18. The fusion protein of claim 17, wherein the Zscan4 comprises mouse Zscan4c or human ZSCAN4.

19. The fusion protein of any one of claims 16-18, comprising an amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO: 22 or SEQ ID NO: 23.

20. The fusion protein of claim 19, comprising the amino acid sequence of SEQ ID NO: 22 or SEQ ID NO: 23.

21. A composition comprising the protein of any one of claims 15-20 and a pharmaceutically acceptable carrier.

22. A method of enhancing the efficiency of homologous recombination in a mammalian cell, comprising contacting the cell with the nucleic acid molecule or vector of any one of claims 1-8, the composition of claim 14 or claim 21, or the protein of any one of claims 15-20.

23. The method of claim 22, wherein the mammalian cell is a mouse cell.

24. The method of claim 22, wherein the mammalian cell is a human cell.

25. The method of any one of claims 22-24, wherein the mammalian cell is an ES cell.

26. The method of any one of claims 22-24, wherein the mammalian cell is an iPS cell.

27. The method of any one of claims 22-26, wherein the nucleic acid molecule, vector or protein is encapsulated in a nanoparticle.

28. A method of enhancing the efficiency of homologous recombination in a mammalian cell, comprising contacting the cell with a recombinant Zscan4 protein or a recombinant nucleic acid molecule encoding a Zscan4 protein, thereby enhancing the efficiency of homologous recombination in the cell.

29. The method of claim 27, wherein the Zscan4 protein comprises a Zscan4-ERT2 fusion protein.

30. The method of any one of claims 22-28, which is an *in vitro* method.

31. The method of any one of claims 22-28, which is an *in vivo* method.

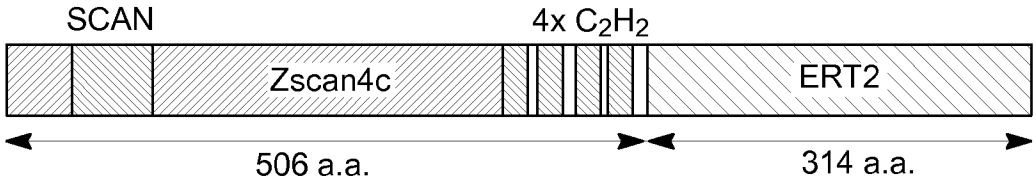

FIG. 1A

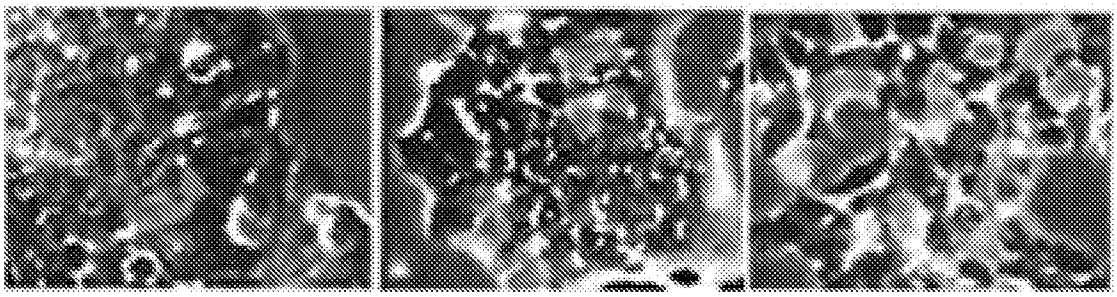

FIG. 1B

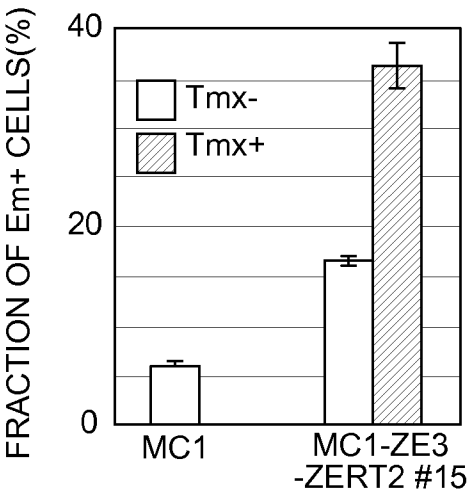

FIG. 1C

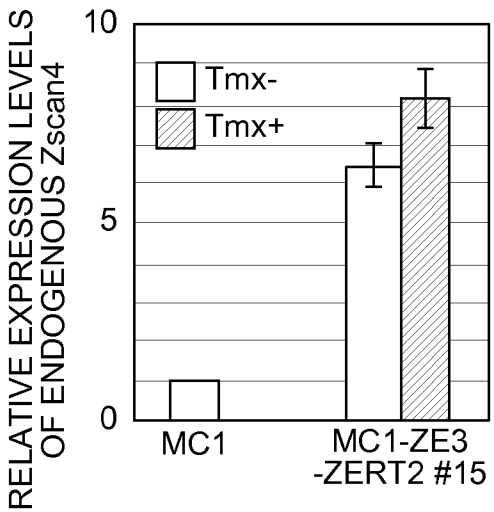

FIG. 1D

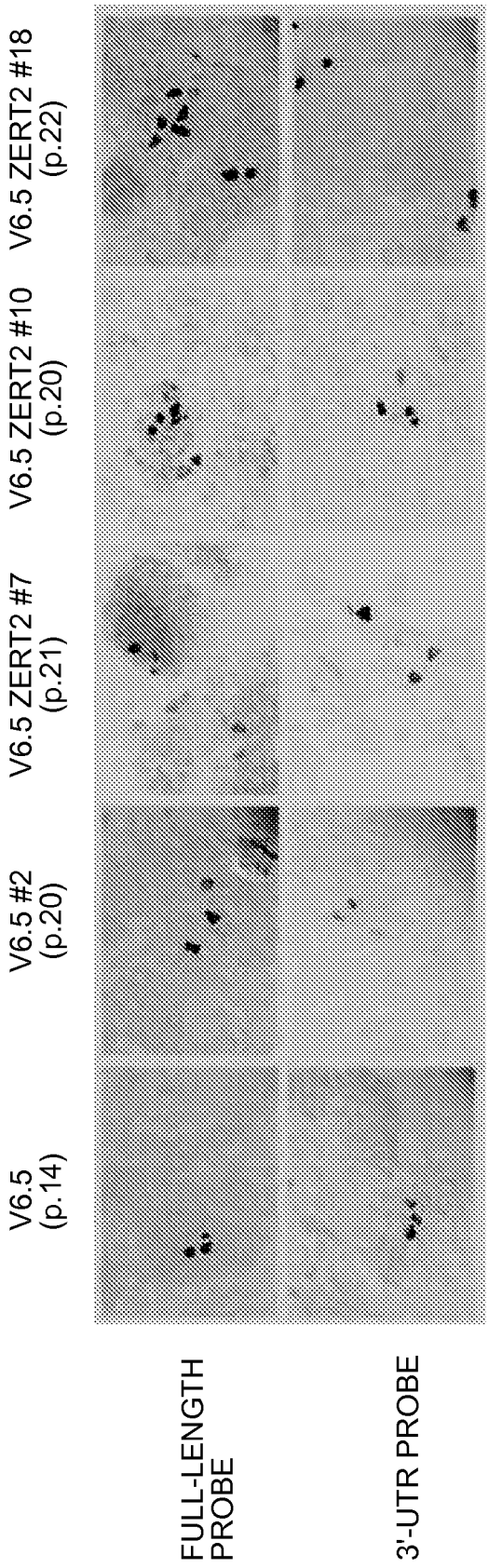
V6.5 (p.14)
V6.5 #2 (p.20)
V6.5 ZERT2 #7 (p.21)
V6.5 ZERT2 #10 (p.20)
V6.5 ZERT2 #18 (p.22)
FULL-LENGTH PROBE
3'-UTR PROBE

FIG. 1E

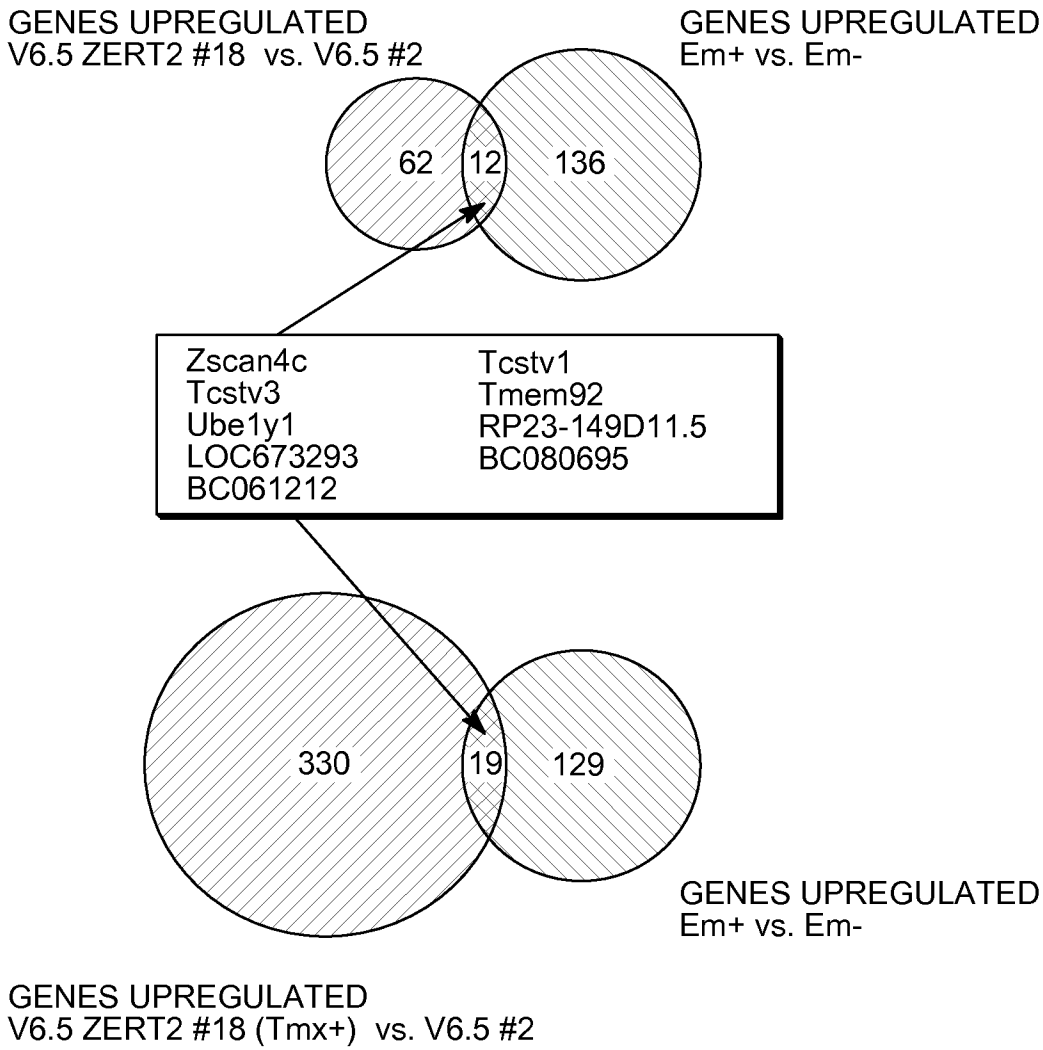
GENES UPREGULATED
V6.5 ZERT2 #18 vs. V6.5 #2
GENES UPREGULATED
Em+ vs. Em-
62
12
136
Zscan4c
Tcstv3
Ube1y1
LOC673293
BC061212
Tcstv1
Tmem92
RP23-149D11.5
BC080695
330
19
129
GENES UPREGULATED
Em+ vs. Em-
GENES UPREGULATED
V6.5 ZERT2 #18 (Tmx+) vs. V6.5 #2

FIG. 1F

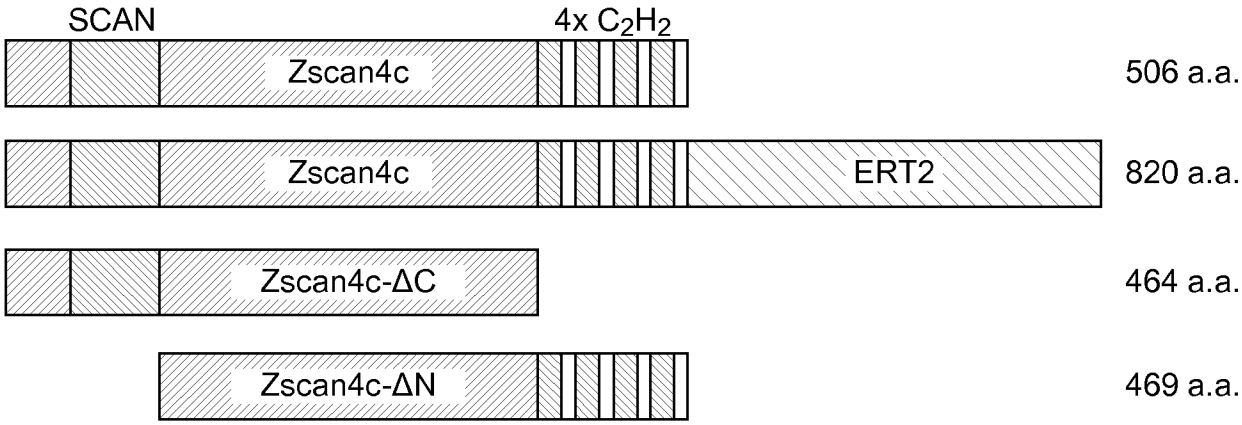

FIG. 2A

ZDC-MC1-ZE16 CLONES, DAY3, PASSAGE 3 (Zscan4c-ΔC)

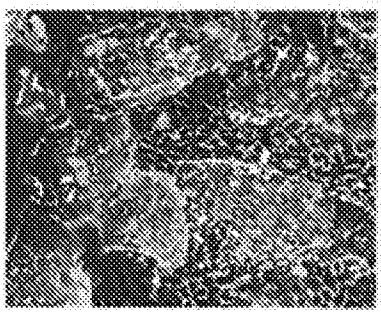

FIG. 2B

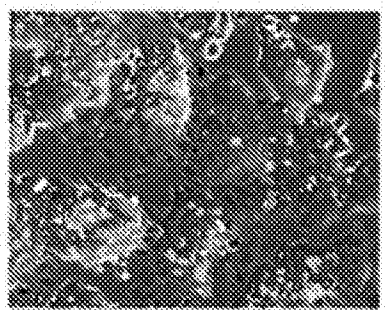

FIG. 2C

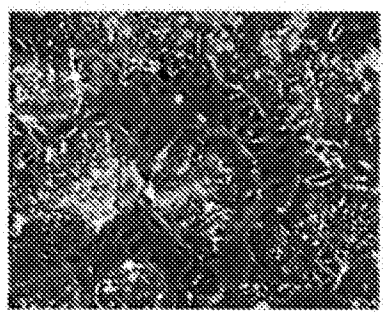

FIG. 2D

ZDN-MC1-ZE16 CLONES, DAY3, PASSAGE 3 (Zscan4c-ΔN)

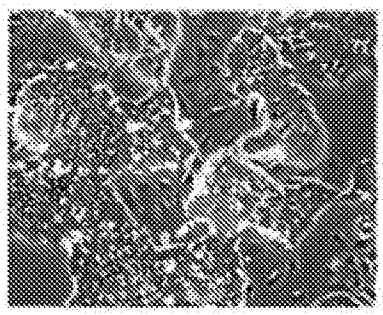

FIG. 2E

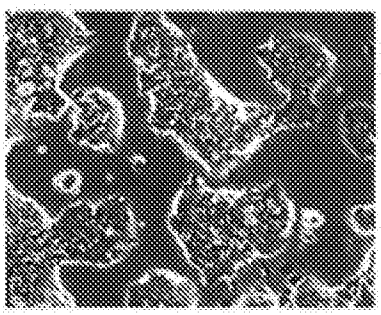

FIG. 2F

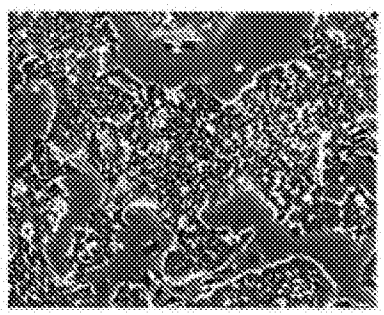

FIG. 2G

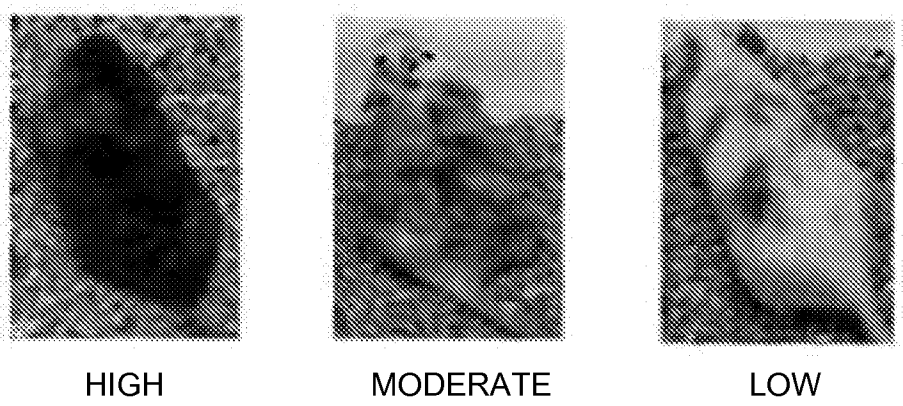
HIGH
MODERATE
LOW

FIG. 3A

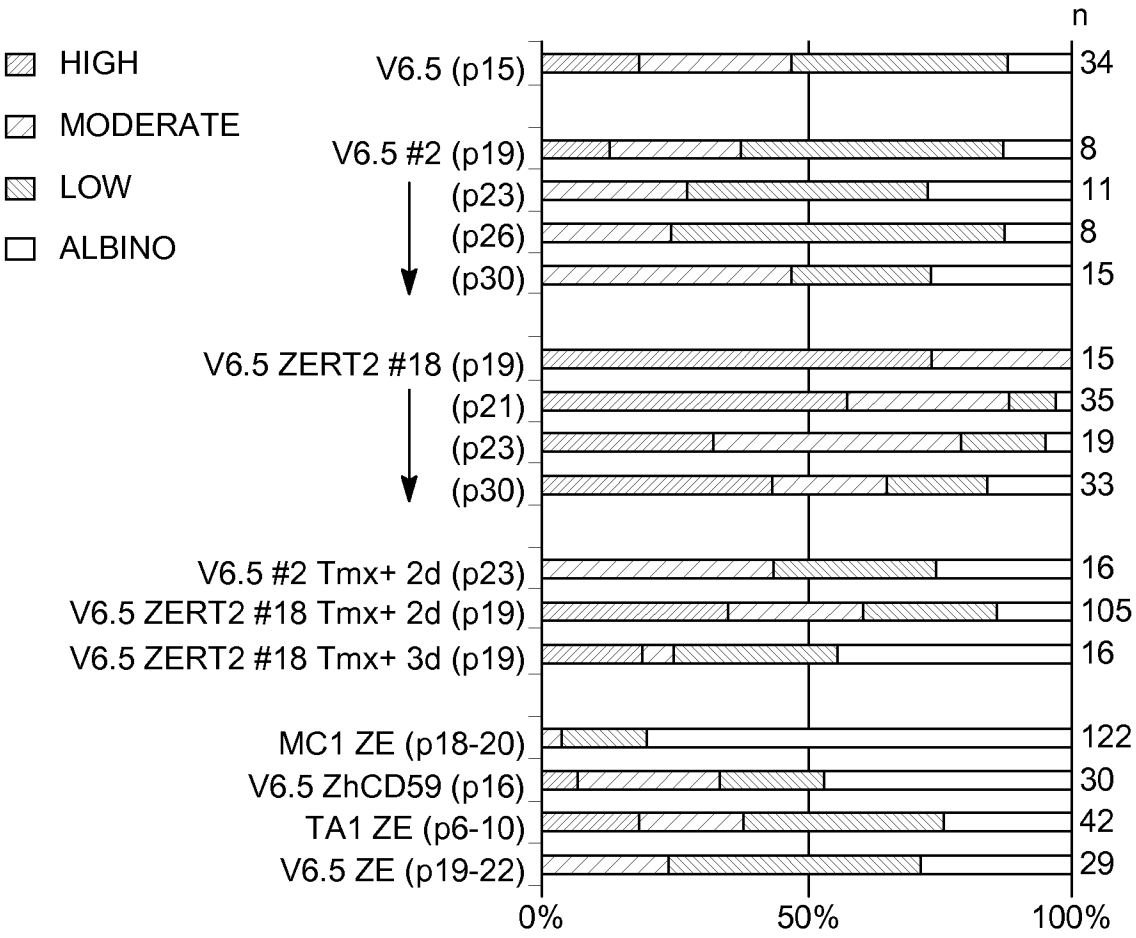
n
HIGH
MODERATE
LOW
ALBINO
V6.5 (p15)
34
V6.5 #2 (p19)
8
(p23)
11
(p26)
8
(p30)
15
V6.5 ZERT2 #18 (p19)
15
(p21)
35
(p23)
19
(p30)
33
V6.5 #2 Tmx+ 2d (p23)
16
V6.5 ZERT2 #18 Tmx+ 2d (p19)
105
V6.5 ZERT2 #18 Tmx+ 3d (p19)
16
MC1 ZE (p18-20)
122
V6.5 ZhCD59 (p16)
30
TA1 ZE (p6-10)
42
V6.5 ZE (p19-22)
29
0%
50%
100%

FIG. 3B

CELL LINE	PASSAGE NO	NO. INJECTED ES CELLS PER BLASTOCYST	NO. TRANSFERRED BLASTOCYSTS	NO. EMBRYOS (%)		NO. LIVE EMBRYOS (%)
				E13.5-14.5	E18.5	
V6.5	18	10-15	41	2 (5%)		1 (2%)
V6.5 ZERT2 #7	22	10-15	40	2 (5%)		1 (3%)
V6.5 ZERT2 #10	22	10-15	41	9 (22%)		9 (22%)
V6.5 ZERT2 #18	19	10-15	40	22 (55%)		17 (43%)
TA1	3	10-15	20	13 (65%)		12 (60%)
MC2	17	10-15	45	0 (0%)		0 (0%)
MC2 ZE #18	12-13	10-15	123	0 (0%)		0 (0%)
MC2 ZERT2 #6	18	10-15	48	3 (6%)		3 (6%)

FIG. 4A

CELL LINE	PASSAGE NO	NO. INJECTED ES CELLS PER BLASTOCYST	NO. TRANSFERRED BLASTOCYSTS	NO. EMBRYOS (%)		NO. LIVE EMBRYOS (%)
				E13.5-14.5	E18.5	
V6.5	18	1	71		1 (1%)	1 (1%)
V6.5 #2	21	1	77		0 (0%)	0 (0%)
V6.5 ZERT2 #18	21	1	44		3 (7%)	2 (5%)
TA1	4	1	79	3(4%)		3 (4%)

FIG. 4B

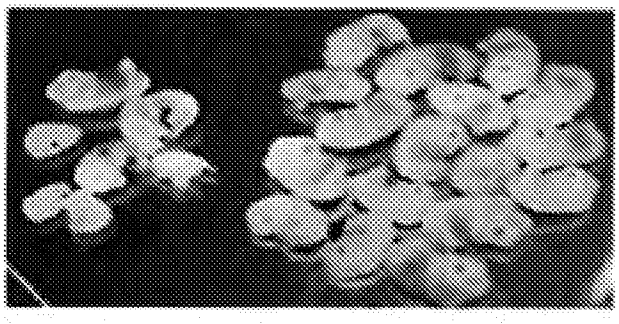

FIG. 4C

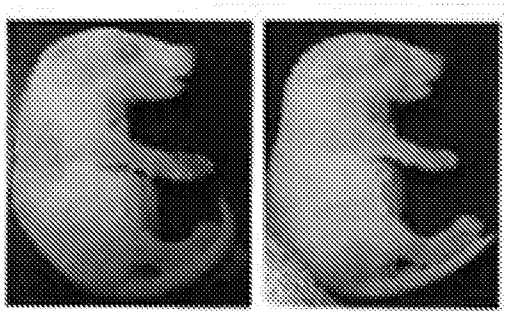

FIG. 4D

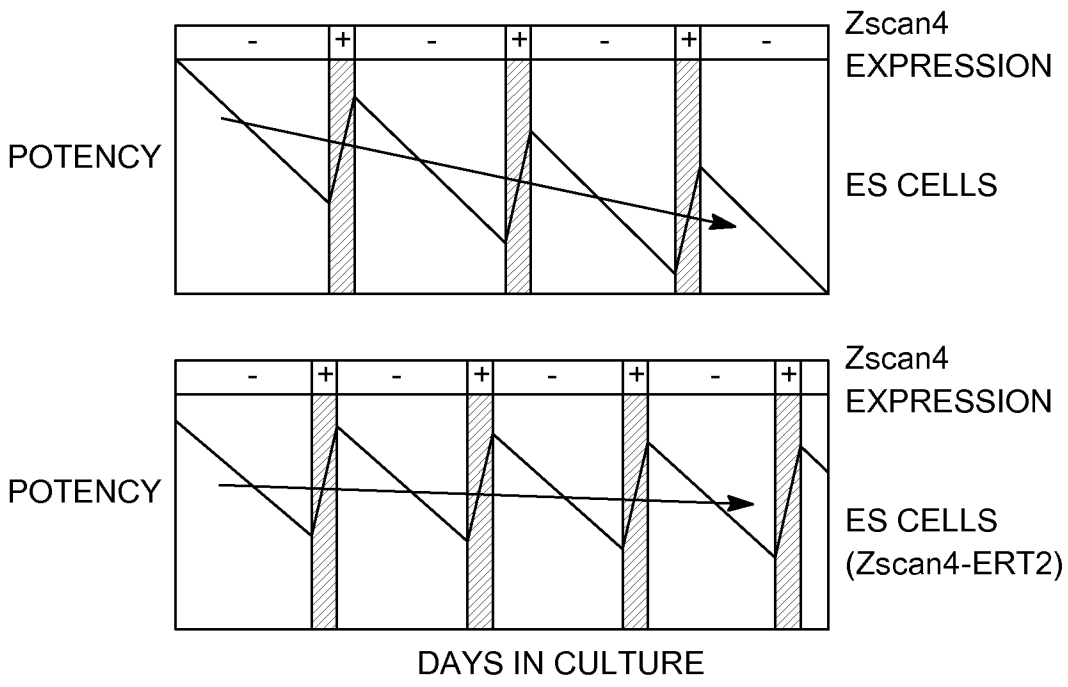

FIG. 4E

FIG. 5A | FIG. 5AA | FIG. 5AB

GENE SYMBOL	ANNOTATION	FOLD CHANGE
Tcstv3	2-CELL-STAGE, VARIABLE GROUP, MEMBER 3	107.8
Zscan4c	ZINC FINGER AND SCAN DOMAIN CONTAINING 4C	68.7
Tmem92	TRANSMEMBRANE PROTEIN 92	40.7
A530040E14Rik	RIKEN cDNA A530040E14 GENE	39.7
RP23-149D11.5	NOVEL PROTEIN SIMILAR TO PRAME PROTEINS	27.6
LOC432715	SIMILAR TO GLE1 RNA EXPORT MEDIATOR-LIKE (YEAST	26.2
Arg2	ARGINASE TYPE II	24.9
BC061212	cDNA SEQUENCE BC061212	21.9
Tcstv1	2-CELL-STAGE, VARIABLE GROUP, MEMBER 1	21.1
AF067063	cDNA SEQUENCE AF067063	20.7
OTTMUSG00000001246	PREDICTED GENE, OTTMUSG00000001246	19.3
EG226955	PREDICTED GENE, EG226955	19.2
Lgals4	LECTIN, GALACTOSE BINDING, SOLUBLE 4	16.2
LOC434660		16.1
Eif1a	EUKARYOTIC TRANSLATION INITIATION FACTOR 1A	16.1
Pif1	PIF1.5'-TO-3' DNA HELICASE HOMOLOG (S. CEREVISIAE)	13.9
	SIMILAR TO NUCLEAR PROTEIN SkiP (Ski-INTERACTING PROTEIN) (SNW1 PROTEIN) (NUCLEAR RECEPTOR COACTIVATOR NCoA- 62)	12.6
LOC673293		12.5
LOC433231		12.1
OTTMUSG00000002043	PREDICTED GENE, OTTMUSG00000002043	10.5
2310065F04Rik	RIKEN cDNA 2310065F04 GENE	9.9
RP23-149D11.4	HYPOTHETICAL LOC381535	9.6
C130092O11Rik	RIKEN cDNA C130092O11 GENE	

FIG. 5AA

LOC671025	SIMILAR TO U2-ASSOCIATED SR140 PROTEIN	9.3
LOC385201	SIMILAR TO SPETEX-2C PROTEIN	9.3
Ly6h	LYMPHOCYTE ANTIGEN 6 COMPLEX, LOCUS H	9.2
LOC240895	SIMILAR TO SKI-INTERACTING PROTEIN	9.2
EG668777	PREDICTED GENE, EG668777	9.1
Calcoco2	CALCIUM BINDING AND COILED-COIL DOMAIN 2	9.0
XM_145358	SEQUENCE XM_145358	9.0
Myo3a	MYOSIN IIIA	8.4
Hist1h4i	HISTONE CLUSTER 1, H4i	8.1
	SIMILAR TO THO COMPLEX SUBUNIT 4 (THO4) (RNA AND EXPORT FACTOR BINDING PROTEIN 1) (REF1-I) (Ally of AML-1 and LEF-1) (Aly/REF)	8.1
LOC669397	REGULATORY FACTOR X-ASSOCIATED PROTEIN	7.9
Rfxap	SIMILAR TO THO COMPLEX SUBUNIT 4 (Tho4) (RNA AND EXPORT FACTOR BINDING PROTEIN 1) (REF1-I) (Ally of AML-1 and LEF-1)(Aly/REF)	7.7
LOC627530	PREDICTED GENE, EG666169	7.7
EG666169	GENE MODEL 257, (NCBI)	7.4
Gm257	RIKEN cDNA 1700066J24 GENE	7.0
1700066l24Rik	T-BOX 3	6.6
Tbx3	MONOGLYCERIDE LIPASE	6.5
Mgll	SIMILAR TO GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE	6.4
LOC226030	TETRASPANIN 1	6.3
Tspan1	PREDICTED GENE, EG627488	6.1
EG627488	SIMILAR TO 2-CELL-STAGE, VARIABLE GROUP, MEMBER 3	6.0
LOC625360	REPRODUCTIVE HOMEOBOX 5	4.9
Rhox5	PREDICTED GENE, EG666099	4.7
EG666099	PREDICTED: MUS MUSCULUS SIMILAR TO OOCYTE SPECIFIC HOMEOBOX 3 (LOC545920), mRNA	4.5
LOC545920	PREDICTED GENE, EG666272	4.4
EG666272	MYOSIN, LIGHT POLYPEPTIDE 4	4.3
Myl4	EXPRESSED SEQUENCE C78283	4.3
C78283		

FIG. 5AB

FIG. 5B	FIG. 5BA
	FIG. 5BB

Sycp1	SYNAPTONEMAL COMPLEX PROTEIN 1	4.3
D13Ertd608e	DNA SEGMENT, Chr 13, ERATO Doi 608, EXPRESSED	4.2
Pdlim3	PDZ AND LIM DOMAIN 3	4.1
Hspa1a	HEAT SHOCK PROTEIN 1A	4.1
EG624262	PREDICTED GENE, EG624262	4.0
Pdgfrl	PLATELET-DERIVED GROWTH FACTOR RECEPTOR-LIKE	4.0
Fer1l3	FER-1-LIKE 3, MYOFERLIN (C. ELEGANS)	3.9
Hist2h2aa1	HISTONE CLUSTER 2, H2aa1	3.9
Rpl39l	RIBOSOMAL PROTEIN L39-LIKE	3.7
BC080695	cDNA SEQUENCE BC080695	3.7
Zfp352	ZINC FINGER PROTEIN 352	3.7
4631416L12Rik	RIKEN cDNA 4631416L12 GENE	3.6
EG238217	PREDICTED GENE, EG238217	3.4
Hist1h2bc	HISTONE CLUSTER 1, H2bc	3.4
Gm428	GENE MODEL 428, (NCBI)	3.4
Lonrf3	LON PEPTIDASE N-TERMINAL DOMAIN AND RING FINGER 3	3.3
Lmx1a	LIM HOMEOBOX TRANSCRIPTION FACTOR 1 ALPHA	3.3
LOC619649	SIMILAR TO TRANSCRIPTION ELONGATION FACTOR B POLYPEPTIDE 3 BINDING PROTEIN 1 ISOFORM 1	3.2
1810062G17Rik	RIKEN cDNA 1810062G17 GENE	3.2
Defcr3	DEFENSIN RELATED CRYPTDIN 3	3.2
LOC666185	SIMILAR TO CG32602-PA	3.2
Sord	SORBITOL DEHYDROGENASE	3.1
EG434050	PREDICTED GENE, EG434050	3.1
Hist1h1c	HISTONE CLUSTER 1, H1c	3.1
D1pas1		3.1

FIG. 5BA

Defcr6	DEFENSIN RELATED CRYPTDIN 6	3.1
Slc4a5	SOLUTE CARRIER FAMILY 4, SODIUM BICARBONATE COTRANSPORTER, MEMBER 5	3.0
Neto2	NEUROPILIN (NRP) AND TOLLOID (TLL)-LIKE 2	3.0
Mael	MAELSTROM HOMOLOG (DROSOPHILA)	3.0
Terc	TELOMERASE RNA COMPONENT	3.0
Avpi1	ARGININE VASOPRESSIN-INDUCED 1	3.0
Zcchc17	ZINC FINGER, CCHC DOMAIN CONTAINING 17	3.0
Hspa1b	HEAT SHOCK PROTEIN 1B	3.0
1700123J19Rik	RIKEN cDNA 1700123J19 GENE	2.9
LOC380994	SIMILAR TO SYCP3 LIKE Y-LINKED	2.9
Thnsl2	THREONINE SYNTHASE-LIKE 2 (BACTERIAL)	2.8
LOC673656	HYPOTHETICAL PROTEIN LOC673656	2.8
Cidea	CELL DEATH-INDUCING DNA FRAGMENTATION FACTOR, ALPHA SUBUNIT-LIKE EFFECTOR A	2.8
Vstm2b	V-SET AND TRANSMEMBRANE DOMAIN CONTAINING 2B	2.8
Jam2	JUNCTION ADHESION MOLECULE 2	2.8
Tuba3a	TUBULIN, ALPHA 3A	2.7
Rhox2a	REPRODUCTIVE HOMEOBOX 2A	2.7
Spic	Spi-C TRANSCRIPTION FACTOR (Spi-1/PU.1 RELATED)	2.7
Xlr4b	X-LINKED LYMPHOCYTE-REGULATED 4B	2.7
EG665954	PREDICTED GENE. EG665954	2.6
Abcb5	ATP-BINDING CASSETTE, SUB-FAMILY B (MDR/TAP), MEMBER 5	2.6
Lonp2	ION PEPTIDASE 2, PEROXISOMAL	2.6
LOC623166		2.5
Kctd8	POTASSIUM CHANNEL TETRAMERISATION DOMAIN CONTAINING 8	2.5
Ankrd22	ANKYRIN REPEAT DOMAIN 22	2.5

FIG. 5BB

FIG. 5C

FIG. 5CA
FIG. 5CB

Defcr23	DEFENSIN RELATED CRYPTDIN 23	2.5
Ccl3	CHEMOKINE (C-C MOTIF) LIGAND 3	2.5
1700052K11Rik	RIKEN cDNA 1700052K11 GENE	2.5
Mylpf	MYOSIN LIGHT CHAIN, PHOSPHORYLATABLE, FAST SKELETAL MUSCLE	2.5
Ak7	ADENYLATE KINASE 7	2.5
Cdc42ep3	CDC42 EFFECTOR PROTEIN (Rho GTPase BINDING) 3	2.5
Ube1y1	UBIQUITIN-ACTIVATING ENZYME E1, Chr Y 1	2.5
Cacna1s	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, L TYPE, ALPHA 1S SUBUNIT	2.5
B020006M18Rik	RIKEN cDNA B020006M18 GENE	2.4
Zfp296	ZINC FINGER PROTEIN 296	2.3
LOC244061	SIMILAR TO SMAD NUCLEAR INTERACTING PROTEIN 1	2.3
Mlana	MELAN-A	2.3
Defa1	DEFENSIN, ALPHA 1	2.3
EG229571	PREDICTED GENE, EG229571	2.3
Gm5	GENE MODEL 5, (NCBI)	2.3
Arl4d	ADP-RIBOSYLATION FACTOR-LIKE 4D	2.3
Myh13	MYOSIN, HEAVY POLYPEPTIDE 13, SKELETAL MUSCLE	2.3
Hormad1	HORMA DOMAIN CONTAINING 1	2.3
Myh4	MYOSIN, HEAVY POLYPEPTIDE 4, SKELETAL MUSCLE	2.3
Syce1	SYNAPTONEMAL COMPLEX CENTRAL ELEMENT PROTEIN 1	2.3
Prkch	PROTEIN KINASE C, ETA	2.2
Pramel3	PREFERENTIALLY EXPRESSED ANTIGEN IN MELANOMA-LIKE 3	2.2
Glrx2	GLUTAREDOXIN 2 (THIOLTRANSFERASE)	2.2
EG620899	PREDICTED GENE, EG620899	2.2
LOC675962	SIMILAR TO TD AND POZ DOMAIN CONTAINING 5	2.2
Tigd2	TIGGER TRANSPOSABLE ELEMENT DERIVED 2	2.2

FIG. 5CA

Phf11	PHD FINGER PROTEIN 11	2.2
Gabarapl2	GAMMA-AMINOBUTYRIC ACID (GABA- A) RECEPTOR -ASSOCIATED PROTEIN-LIKE 2	2.2
Gsta3	GLUTATHIONE S-TRANSFERASE, ALPHA 3	2.2
LOC385262	cDNA SEQUENCE BC003993	2.1
BC003993	TD AND POZ DOMAIN CONTAINING 4	2.1
Tdpoz4	TAF7-LIKE RNA POLYMERASE II, TATA BOX BINDING PROTEIN (TBP)-ASSOCIATED FACTOR	2.1
Taf7l	ISOCHORISMATASE DOMAIN CONTAINING 2b	2.1
Isoc2b	ZINC FINGER PROTEIN 371	2.1
Zfp371	RIKEN cDNA 8030474K03 GENE	2.1
8030474K03Rik	SYNAPTONEMAL COMPLEX PROTEIN 2	2.1
Sycp2	GENE MODEL 1568, (NCBI)	2.1
Gm1568	STORKHEAD BOX 1	2.1
Stox1	DEFENSIN RELATED CRYPTDIN 24	2.1
Defcr24	ARIADNE HOMOLOG 2 (DROSOPHILA)	2.1
Arih2	SERINE (OR CYSTEINE) PEPTIDASE INHIBITOR, CLADE B, MEMBER 1B	2.0
Serpinb1b	Sp110 NUCLEAR BODY PROTEIN	2.0
Sp110	MELANOREGULIN	2.0
Mreg	ACID PHOSPHATASE 6, LYSOPHOSPHATIDIC	2.0
Acp6	MICRORCHIDIA 1	2.0
Morc1	PREDICTED GENE, EG666692	2.0
EG666692	PREDICTED GENE, EG666675	2.0
EG666675	BRAIN DERIVED NEUROTROPHIC FACTOR	2.0
Bdnf		

FIG. 5CB

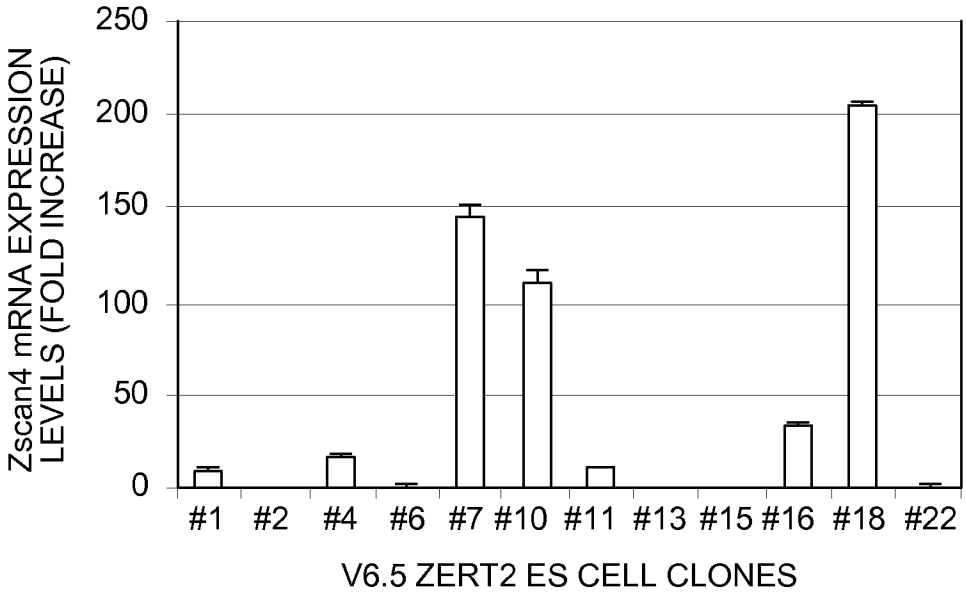
Zscan4 mRNA EXPRESSION LEVELS (FOLD INCREASE)
250
200
150
100
50
0
#1 #2 #4 #6 #7 #10 #11 #13 #15 #16 #18 #22
V6.5 ZERT2 ES CELL CLONES

FIG. 6A

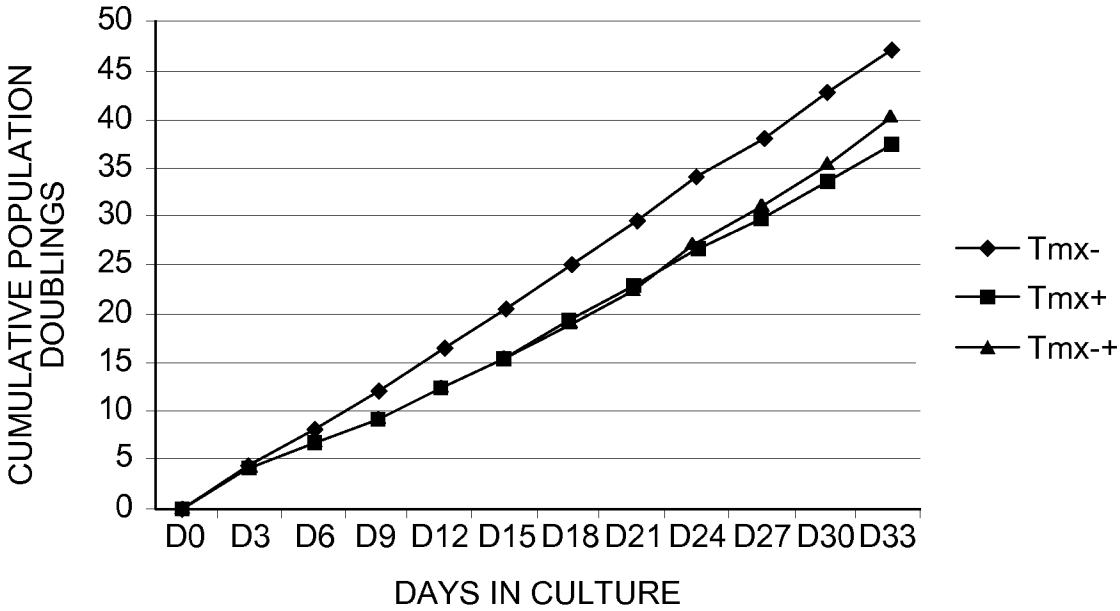
CUMULATIVE POPULATION DOUBLINGS
50
45
40
35
30
25
20
15
10
5
0
D0 D3 D6 D9 D12 D15 D18 D21 D24 D27 D30 D33
DAYS IN CULTURE
Tmx-
Tmx+
Tmx-+

FIG. 6B

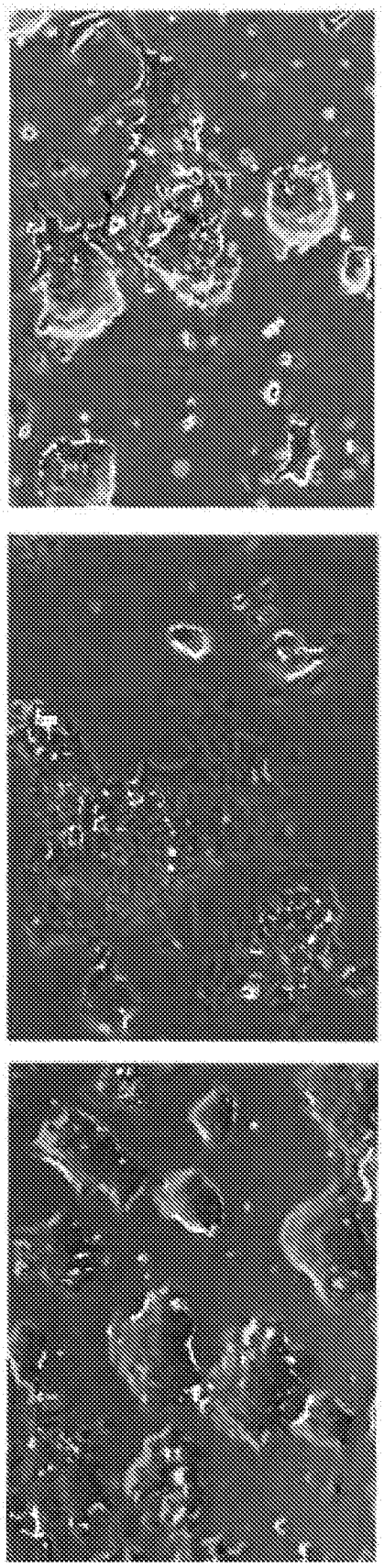

FIG. 6C

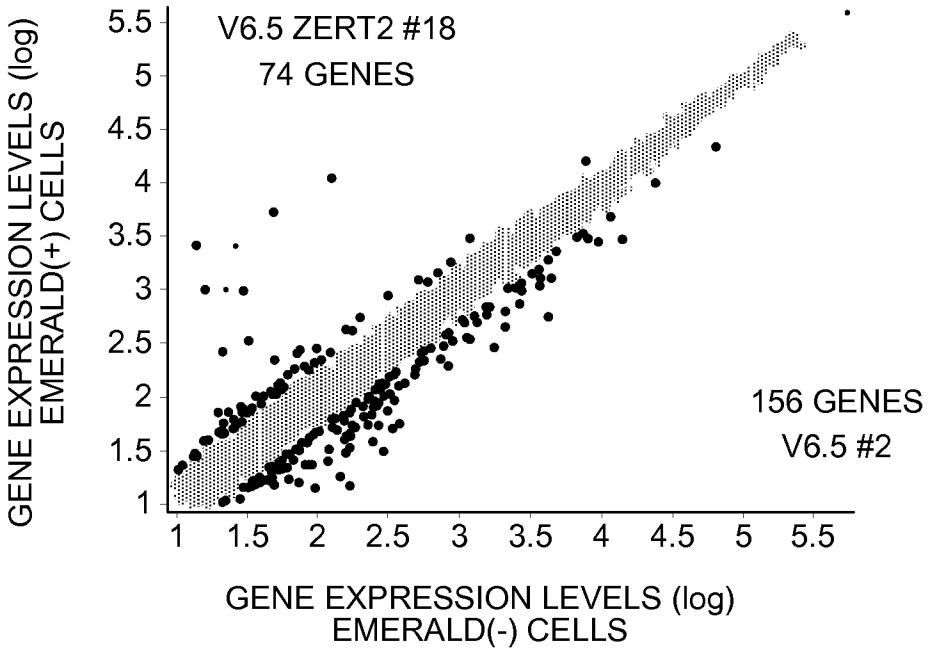
V6.5 ZERT2 #18
74 GENES
156 GENES
V6.5 #2
GENE EXPRESSION LEVELS (log)
EMERALD(+) CELLS
GENE EXPRESSION LEVELS (log)
EMERALD(-) CELLS
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
1 1.5 2 2.5 3 3.5 4 4.5 5 5.5

FIG. 7A

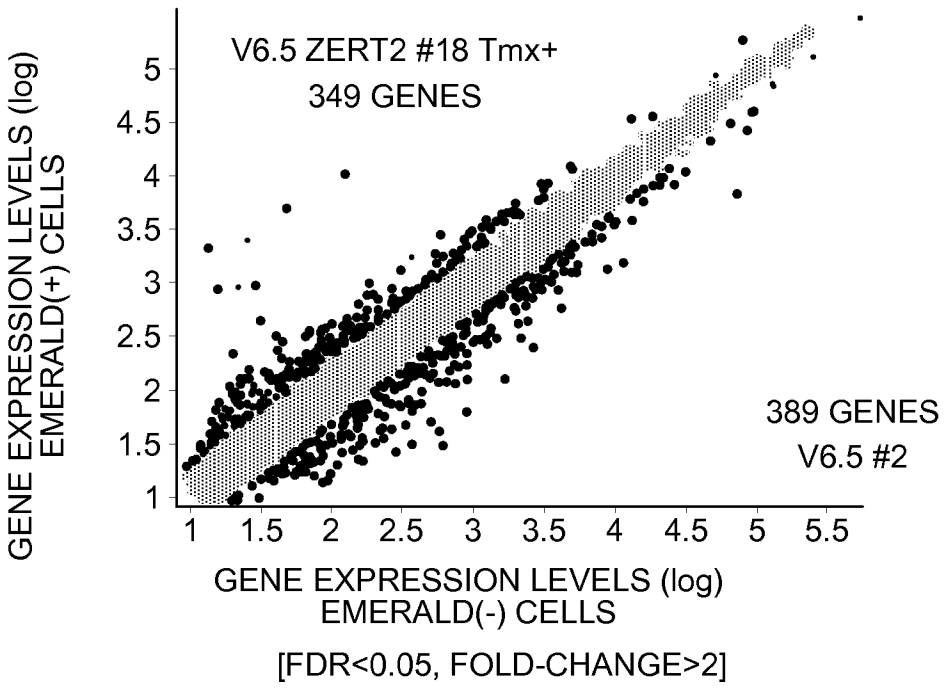
V6.5 ZERT2 #18 Tmx+
349 GENES
389 GENES
V6.5 #2
GENE EXPRESSION LEVELS (log)
EMERALD(+) CELLS
GENE EXPRESSION LEVELS (log)
EMERALD(-) CELLS
[FDR<0.05, FOLD-CHANGE>2]
5
4.5
4
3.5
3
2.5
2
1.5
1
1 1.5 2 2.5 3 3.5 4 4.5 5 5.5

FIG. 7B

FIG. 8 | FIG. 8A | FIG. 8B

GENE SYMBOL	ANNOTATION	FOLD CHANGE
Rian	RNA IMPRINTED AND ACCUMULATED IN NUCLEUS	191.7
Eif2s3y	EUKARYOTIC TRANSLATION INITIATION FACTOR 2, SUBUNIT 3, STRUCTURAL GENE Y-LINKED	113.9
Uty	UBIQUITOUSLY TRANSCRIBED TETRATRICOPEPTIDE REPEAT GENE, Y CHROMOSOME	99.9
Mirg	miRNA CONTAINING GENE	92.1
B830012L14Rik	RIKEN cDNA B830012L14 GENE	64.1
Meg3	MATERNALLY EXPRESSED 3	44.5
Ddx3y	DEAD (Asp-Glu-Ala-Asp) BOX POLYPEPTIDE 3, Y-LINKED	33.5
6430411K18Rik	RIKEN cDNA 6430411K18 GENE	12.9
Ube1y1	UBIQUITIN-ACTIVATING ENZYME E1, Chr Y 1	10.3
Cubn	CUBILIN (INTRINSIC FACTOR-COBALAMIN RECEPTOR)	4.6
Peg13	PATERNALLY EXPRESSED 13	3.7
Tmem92	TRANSMEMBRANE PROTEIN 92	3.7
Zscan4c	ZINC FINGER AND SCAN DOMAIN CONTAINING 4C	3.6
BC080695	cDNA SEQUENCE BC080695	3.1
Ntrk2	NEUROTROPHIC TYROSINE KINASE, RECEPTOR, TYPE 2	2.9
Spon1	SPONDIN 1, (F-SPONDIN) EXTRACELLULAR MATRIX PROTEIN	2.9
AF067061	cDNA SEQUENCE AF067061	2.9
B020031M17Rik	RIKEN cDNA B020031M17 GENE	2.8
Tcstv3	2-CELL-STAGE, VARIABLE GROUP, MEMBER 3	2.8
Rxrg	RETINOID X RECEPTOR GAMMA	2.8
Aadat	AMINOADIPATE AMINOTRANSFERASE	2.7
Tnrc4	TRINUCLEOTIDE REPEAT CONTAINING 4	2.7
OTTMUSG00000010537	PREDICTED GENE, OTTMUSG00000010537	2.7

FIG. 8A

Ebf1	EARLY B-CELL FACTOR 1	2.6
Defb30	DEFENSIN BETA 30	2.6
C130034I18Rik	RIKEN cDNA C130034I18 GENE	2.6
Nudt12	NUDIX (NUCLEOSIDE DIPHOSPHATE LINKED MOIETY X)-TYPE MOTIF 12	2.5
Dnajb13	DnaJ (Hsp40) RELATED, SUBFAMILY B, MEMBER 13	2.5
Tnfrsf17	TUMOR NECROSIS FACTOR RECEPTOR SUPERFAMILY, MEMBER 17	2.5
Gm428	GENE MODEL 428, (NCBI)	2.5
BC061212	cDNA SEQUENCE BC061212	2.5
Zbp1	Z-DNA BINDING PROTEIN 1	2.5
GalNt6	UDP-N-ACETYL-ALPHA-D-GALACTOSAMINE:POLYPEPTIDE N-ACETYLGALACTOSAMINYLTRANSFERASE 6	2.4
Syt4	SYNAPTOTAGMIN IV	2.4
Adm	ADRENOMEDULLIN	2.4
Krt12	KERATIN 12	2.4
Zfp264	ZINC FINGER PROTEIN 264	2.4
OTTMUSG00000001246	PREDICTED GENE, OTTMUSG00000001246	2.4
1700066J24Rik	RIKEN cDNA 1700066J24 GENE	2.3
Gfra1	GLIAL CELL LINE DERIVED NEUROTROPHIC FACTOR FAMILY RECEPTOR ALPHA 1	2.3
Spaca1	SPERM ACROSOME ASSOCIATED 1	2.3
Krt9	KERATIN 9	2.3
Saa2	SERUM AMYLOID A 2	2.3
Tff3	TREFOIL FACTOR 3, INTESTINAL	2.2
4930599N23Rik	RIKEN cDNA 4930599N23 GENE	2.2
Tcstv1	2-CELL-STAGE, VARIABLE GROUP, MEMBER 1	2.2
LOC673293	SIMILAR TO NUCLEAR PROTEIN SKIP (Ski-INTERACTING PROTEIN) (SNW1 PROTEIN) (NUCLEAR RECEPTOR CO	2.2
2410004A20Rik	RIKEN cDNA 2410004A20 GENE	2.2
Rasd1	RAS, DEXAMETHASONE-INDUCED 1	2.2
Ushbp1	USHER SYNDROME 1C BINDING PROTEIN 1	2.2

FIG. 8B

FIG. 9 | FIG. 9A | FIG. 9B

SYMBOL	ANNOTATION	FOLD CHANGE
Rian	RNA IMPRINTED AND ACCUMULATED IN NUCLEUS	154.8
Eif2s3y	EUKARYOTIC TRANSLATION INITIATION FACTOR 2, SUBUNIT 3, STRUCTURAL GENE Y-LINKED	105.9
Uty	UBIQUITOUSLY TRANSCRIBED TETRATRICOPEPTIDE REPEAT GENE, Y CHROMOSOME	96.2
Mirg	miRNA CONTAINING GENE	84.2
B830012L14Rik	RIKEN cDNA B830012L14 GENE	54.7
Meg3	MATERNALLY EXPRESSED 3	40.8
Ddx3y	DEAD (Asp-Glu-Ala-Asp) BOX POLYPEPTIDE 3, Y-LINKED	32.7
Ube1y1	UBIQUITIN-ACTIVATING ENZYME E1, Chr Y 1	14.0
6430411K18Rik	RIKEN cDNA 6430411K18 GENE	10.6
Zbp1	Z-DNA BINDING PROTEIN 1	7.8
Clca3	CHLORIDE CHANNEL CALCIUM ACTIVATED 3	6.1
Cldn10	CLAUDIN 10	5.9
Heph	HEPHAESTIN	5.7
1700013H16Rik	RIKEN cDNA 1700013H16 GENE	5.7
Tmem92	TRANSMEMBRANE PROTEIN 92	5.6
Pof1b	PREMATURE OVARIAN FAILURE 1B	5.5
Gpm6a	GLYCOPROTEIN m6a	5.2
Zscan4c	ZINC FINGER AND SCAN DOMAIN CONTAINING 4C	5.1
BC080695	cDNA SEQUENCE BC080695	5.0
Fbxw16	F-BOX AND WD-40 DOMAIN PROTEIN 16	4.9
Pdzrn3	PDZ DOMAIN CONTAINING RING FINGER 3	4.8
Indo	INDOLEAMINE-PYRROLE 2,3 DIOXYGENASE	4.8
Fgf10	FIBROBLAST GROWTH FACTOR 10	4.8
Ccrl1	CHEMOKINE (C-C motif) RECEPTOR-LIKE 1	4.8
4931407G18Rik	RIKEN cDNA 4931407G18 GENE	4.7

FIG. 9A

Arg2	ARGINASE TYPE II	4.7
Cftr	CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR HOMOLOG	4.6
Atoh1	ATONAL HOMOLOG 1 (DROSOPHILA)	4.6
Rgn	REGUCALCIN	4.6
OTTMUSG00000010537	PREDICTED GENE, OTTMUSG00000010537	4.5
Kctd12b	POTASSIUM CHANNEL TETRAMERISATION DOMAIN CONTAINING 12b	4.5
Gfra1	GLIAL CELL LINE DERIVED NEUROTROPHIC FACTOR FAMILY RECEPTOR ALPHA 1	4.5
Pdzrn4	PDZ DOMAIN CONTAINING RING FINGER 4	4.4
Cyp2j9	CYTOCHROME P450, FAMILY 2, SUBFAMILY j, POLYPEPTIDE 9	4.4
Calcr	CALCITONIN RECEPTOR	4.4
Cttnbp2	CORTACTIN BINDING PROTEIN 2	4.4
AF067061	cDNA SEQUENCE AF067061	4.3
Pramel6	PREFERENTIALLY EXPRESSED ANTIGEN IN MELANOMA LIKE 6	4.2
Aadac	ARYLACETAMIDE DEACETYLASE (ESTERASE)	4.1
Spic	SPL-C TRANSCRIPTION FACTOR (SPI-1/PU.1 RELATED)	4.1
Lrrk2	LEUCINE-RICH REPEAT KINASE 2	4.1
Dll4	DELTA-LIKE 4 (DROSOPHILA)	4.1
Tmem140	TRANSMEMBRANE PROTEIN 140	4.1
B020031M17Rik	RIKEN cDNA B020031M17 GENE	4.0
Tcstv3	2-CELL-STAGE, VARIABLE GROUP, MEMBER 3	4.0
Snph	SYNTAPHILIN	4.0
Akr1c18	ALDO-KETO REDUCTASE FAMILY 1, MEMBER C18	4.0
1110018M03Rik	RIKEN cDNA 1110018M03 GENE	3.9
Lancl3	LanC LANTIBIOTIC SYNTHETASE COMPONENT C-LIKE 3 (BACTERIAL)	3.9
Trpa1	TRANSIENT RECEPTOR POTENTIAL CATION CHANNEL, SUBFAMILY A, MEMBER 1	3.7

FIG. 9B

ID	ATTACHED (DAY2-3)	OUTGROWTH (DAY5-7)	STATUS OF OUTGROWTH *	1ST PICKED UP (DAY6-8)	ES-LIKE COLONIES (DAY10-14)	ES LINE (DAY20)
1	0	0	U	0	0	0
2	0	0	D	X	-	-
3	0	0	U	0	0	0
4	0	0	U/D	0	0	0
5	0	0	D	X	-	-
6	0	0	D	X	-	-
7	0	0	U/D	0	0	0
8	0	0	U/D	0	0	0
9	0	0	D	X	-	-
10	0	0	U	0	0	0
11	0	0	U/D	0	0	0
12	0	0	U	0	0	0
13	0	0	D	X	-	-
14	0	0	U/D	0	0	0
15	0	0	U	0	0	0
16	0	0	U	0	0	0
17	X	X	-	-	-	-
18	0	0	U/D	0	0	0
19	0	0	D	X	-	-
20	0	0	U	0	0	0

FIG. 10A

NO. BLASTOCYSTS USED	NO. BLASTOCYSTS ATTACHED (%)	NO. OUTGROWTH (%)	NO. ES-LIKE COLONIES (%)	NO. ES CELL LINES ESTABLISHED
20	19 (95%)	19 (95%)	13 (68%)	13 (68%)

FIG. 10B

CELL LINE TESTED	NO. 4N BLASTOCYSTS INJECTED	NO. IMPLANTATION (%)	NO. ES-DERIVED FETUS AT E13.5 (%)	NO. LIVE EMBRYOS (%)	SEX OF EMBRYOS
#1	36	21 (58%)	11 (31%)	10 (28%)	M
#3	24	19 (79%)	11 (46%)	10 (42%)	M
#10	20	16 (80%)	13 (65%)	12 (60%)	M
#15	27	24 (89%)	4 (15%)	4 (15%)	F
#16	20	18 (90%)	11 (55%)	10 (50%)	M
#20	20	17 (85%)	11 (55%)	10 (50%)	M

FIG. 11A

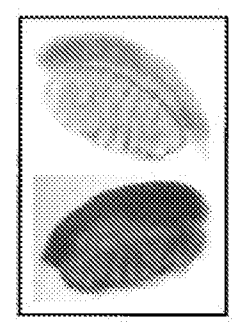

FIG. 11B

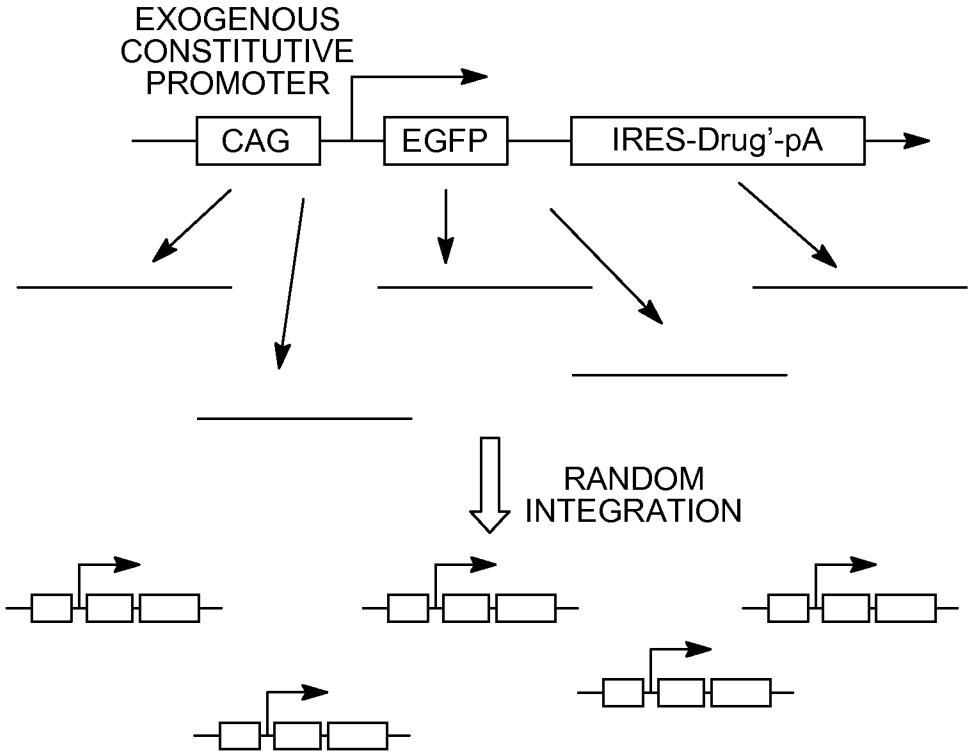

FIG. 12A

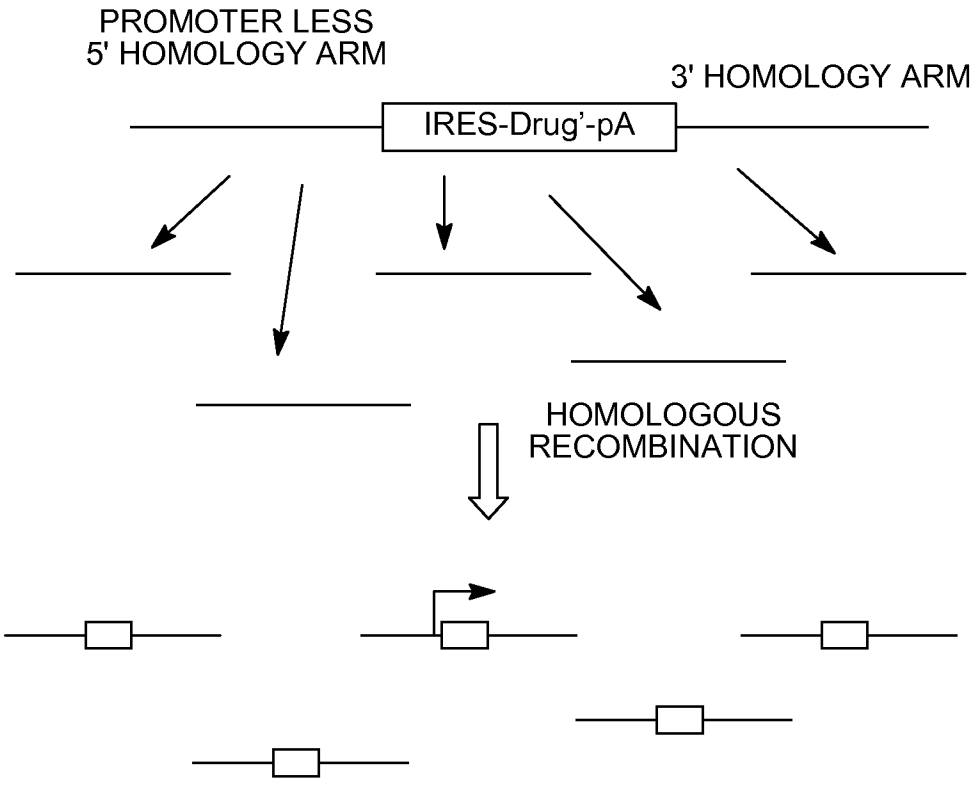

FIG. 12B

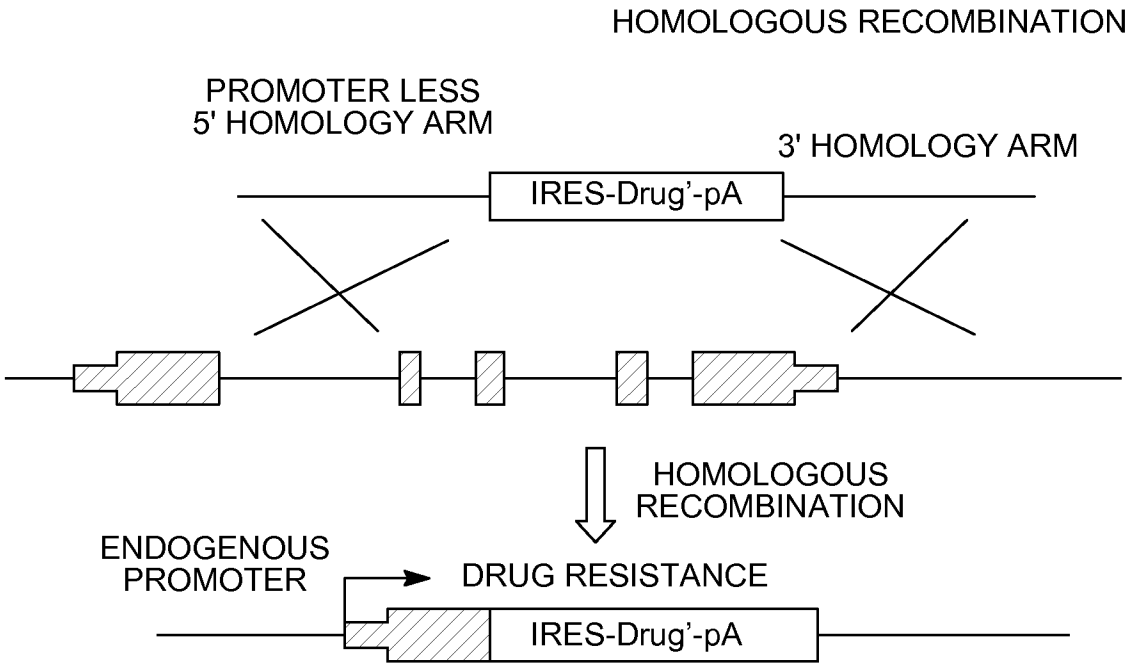

FIG. 12C

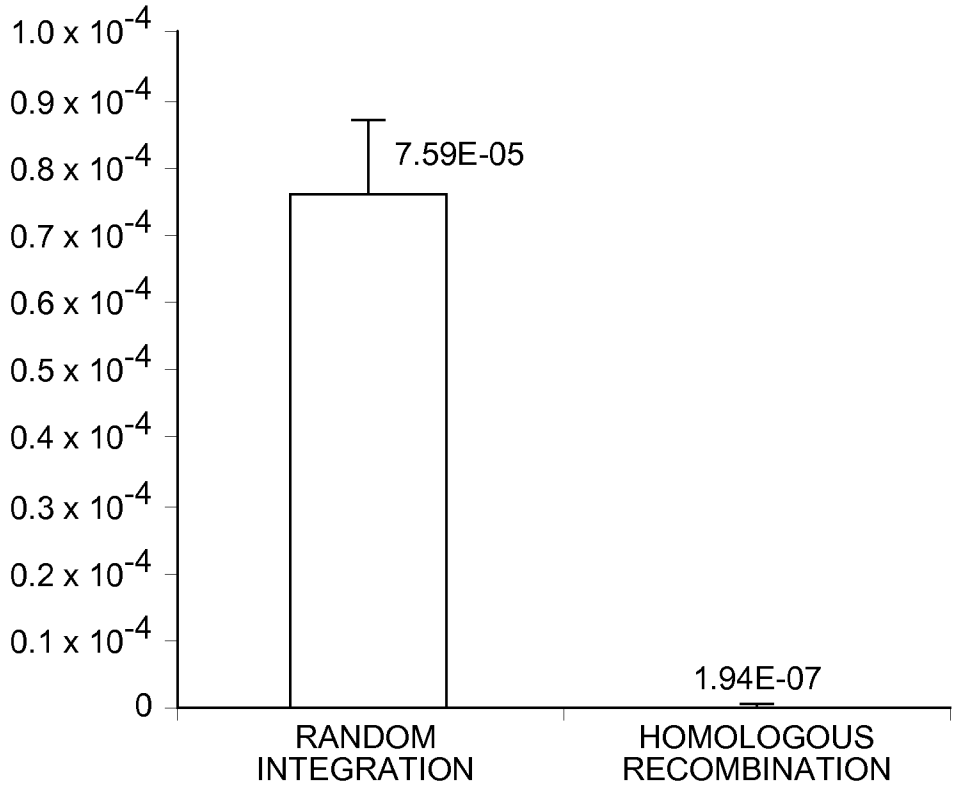

FIG. 12D

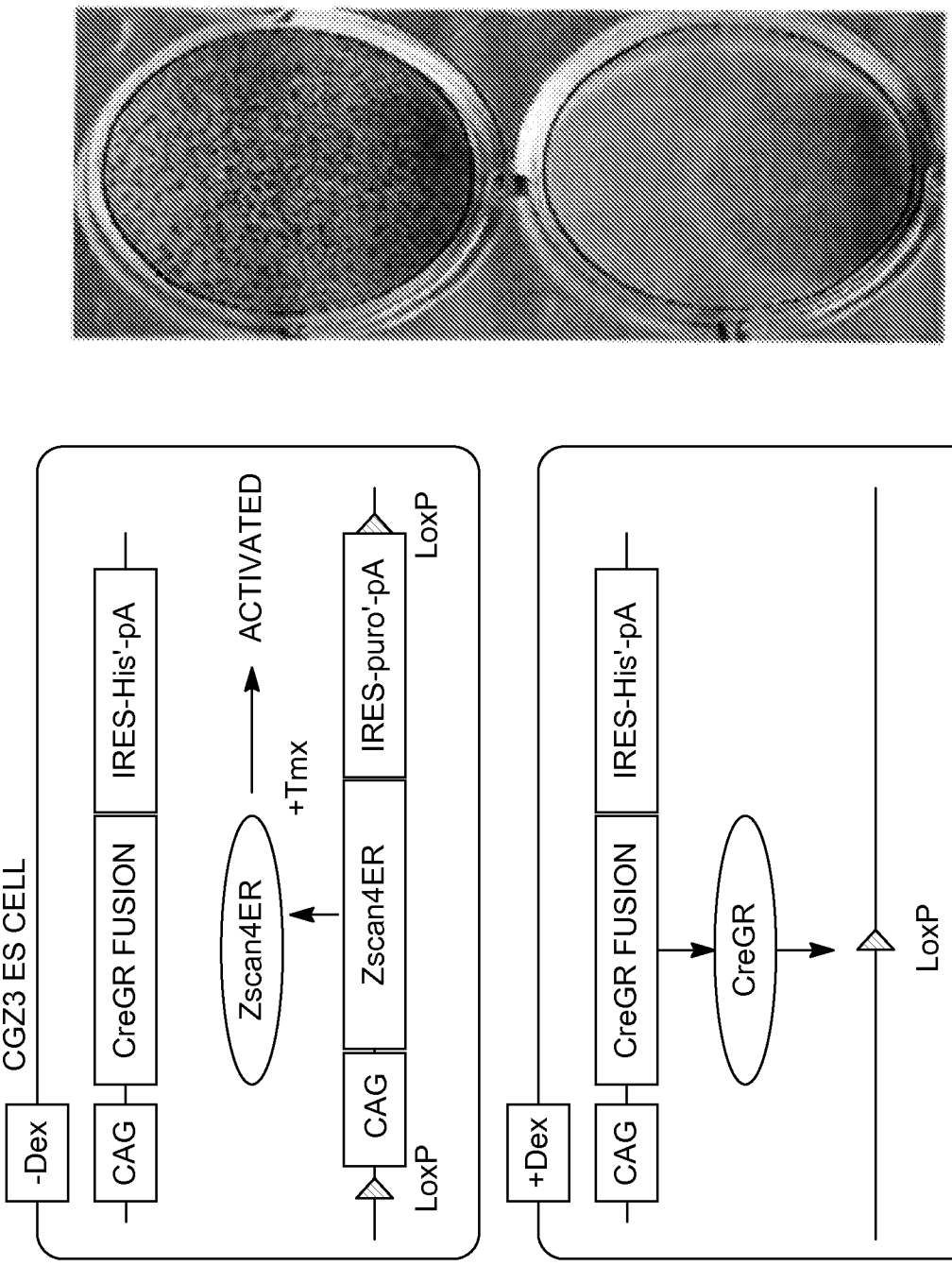
CGZ3 ES CELL
-Dex
CAG
CreGR FUSION
IRES-His'-pA
Zscan4ER
+Tmx
ACTIVATED
CAG
Zscan4ER
IRES-puro'-pA
LoxP
LoxP
+Dex
CAG
CreGR FUSION
IRES-His'-pA
CreGR
LoxP
PUROMYCIN SELECTION

FIG. 13A

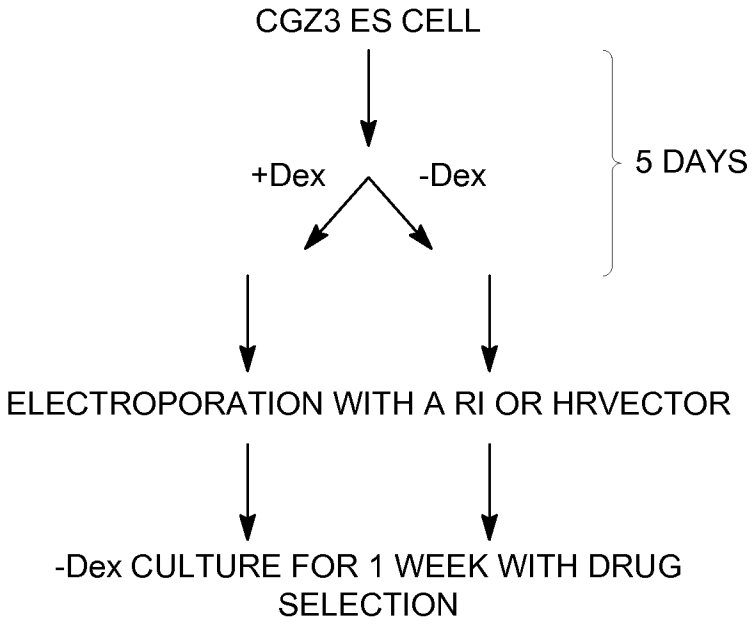

FIG. 13B

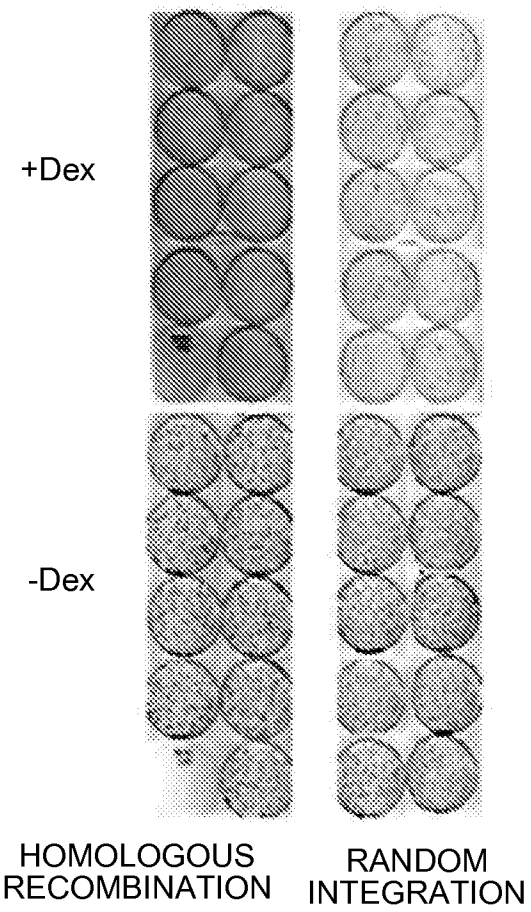

FIG. 13C

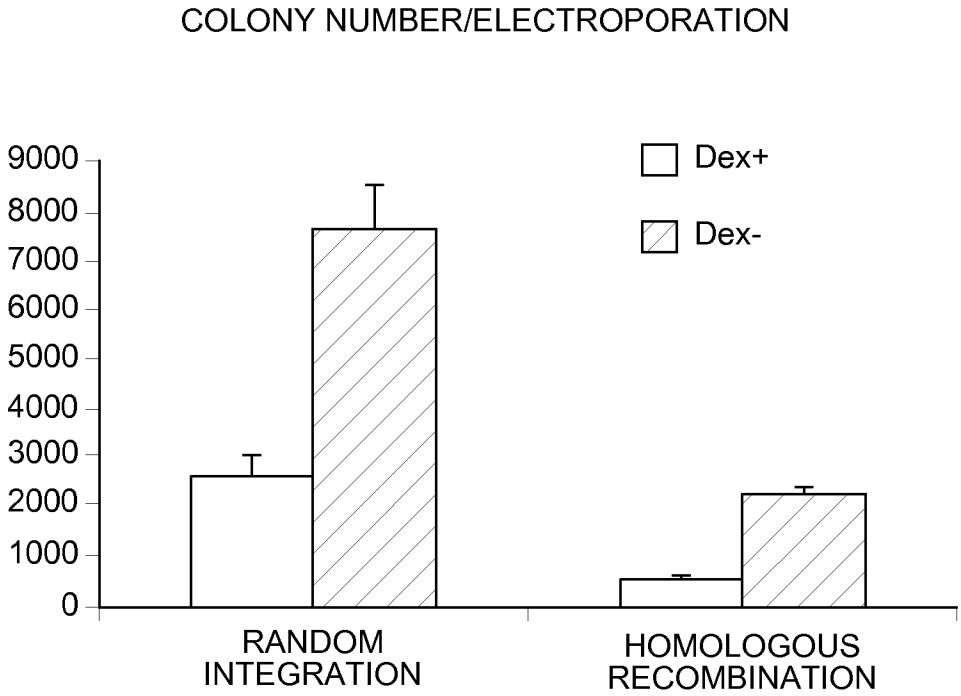
COLONY NUMBER/ELECTROPORATION
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Dex+
Dex-
RANDOM INTEGRATION
HOMOLOGOUS RECOMBINATION

FIG. 13D

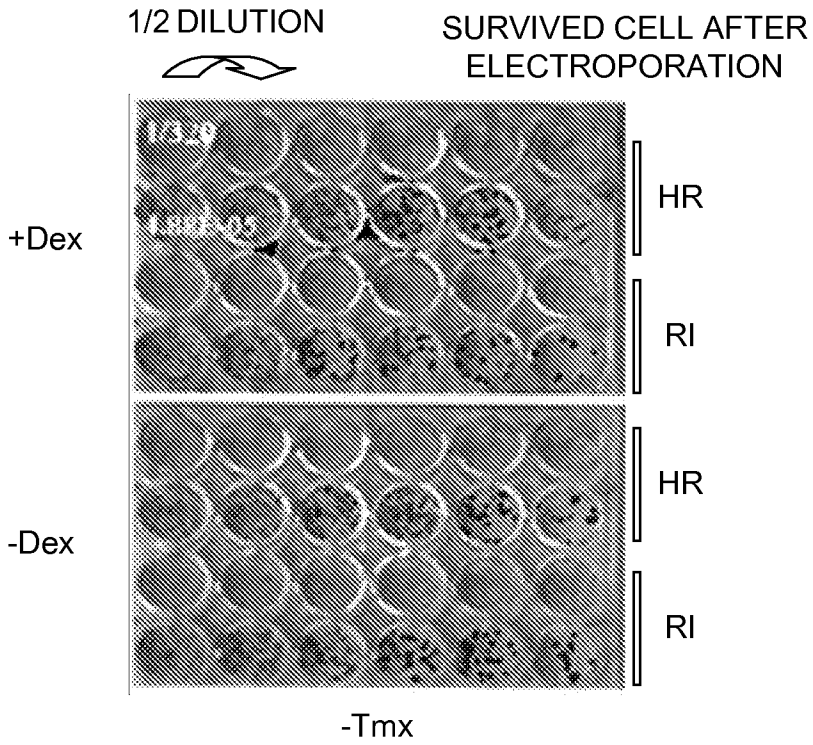

FIG. 13E

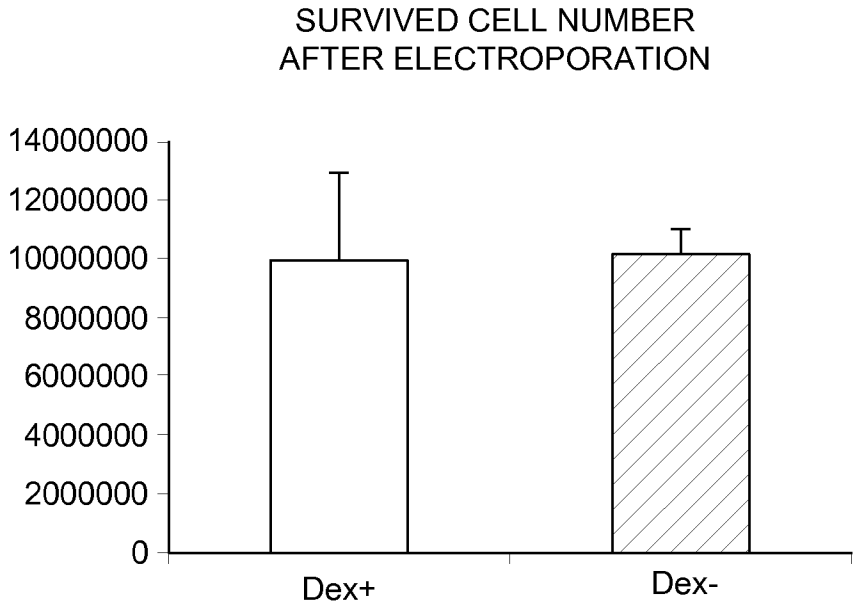

FIG. 13F

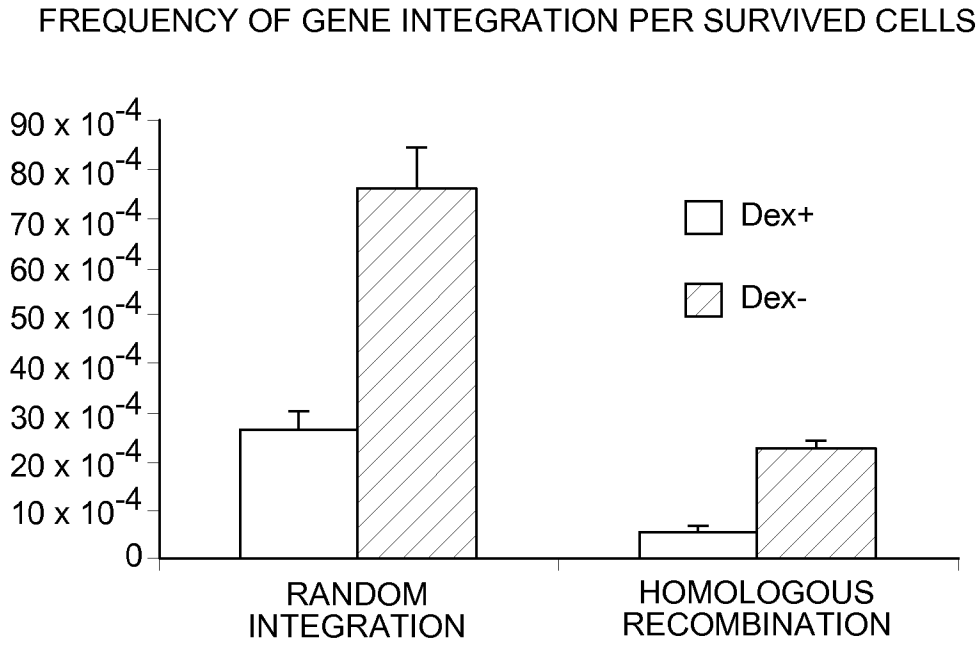

FIG. 13G

	(WILD-TYPE CONTROL ES CELLS) V6.5	(Zscan4-ERT2 REMOVED) CGZ3 Dex+	(Zscan4-ERT2 PRESENT) CGZ3 Dex-
RANDOM INTEGRATION	7.6E-05±1.1E-05	2.3E-04±3.6E-05	7.4E-04±8.9E-05
HOMOLOGOUS RECOMBINATION	1.9E-07±4.1E-07	4.9E-05±31.1E-05	2.2E-04±1.6E-05

FIG. 13H

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2012/037650

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 5/0797 (2012.01)
USPC - 435/455
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - A61K 35/12, 38/02; C12N 5/00, 0797; C12Q 1/68 (2012.01)
USPC - 424/499; 435/6, 325, 366, 455

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase, Google Patents, Google, PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2011/028880 A2 (KO et al) 10 March 2011 (10.03.2011) entire document	1-3, 16-18, 28
A	US 2010/0105043 A1 (KO et al) 29 April 2010 (29.04.2010) entire document	1-3, 16-20, 28
A	ZALZMAN et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464(7290):858-863. 08 April 2010. entire document	1-3, 16-20, 28
A	US 2006/0228798 A1 (VERFAILLIE et al) 12 October 2006 (12.10.2006) entire document	1-3, 16-20, 28

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 August 2012	17 AUG 2012
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2012/037650

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-15, 21-27, 29-31
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.