(12) STANDARD PATENT (11) Application No. **AU 2008230870 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Methods for modulating embryonic stem cell differentiation

(51) International Patent Classification(s)
C12N 5/00 (2006.01) ***C12N 15/00*** (2006.01)

(21) Application No: **2008230870** (22) Date of Filing: **2008.03.26**

(87) WIPO No: **WO08/118957**

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/920,215	**2007.03.26**	**US**

(43) Publication Date: **2008.10.02**
(44) Accepted Journal Date: **2013.09.05**

(71) Applicant(s)
The Government of U.S.A as represented by the Secretary of the Department of Health & Human Services

(72) Inventor(s)
Monti, Manuela;Ko, Minoru S.H.;Lee, Sung-Lim;Stanghellini, Ilaria;Falco, Geppino

(74) Agent / Attorney
Pizzeys, PO Box 291, WODEN, ACT, 2606

(56) Related Art
Zhang et al (2006) Nucleic Acids Research, vol. 34, pps 3780-4790
Edelstein et al (2005) Gene, vol. 359, pps 1-17

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 October 2008 (02.10.2008)

PCT

(10) International Publication Number
WO 2008/118957 A3

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

(21) International Application Number:
PCT/US2008/058261

(22) International Filing Date: 26 March 2008 (26.03.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/920,215 26 March 2007 (26.03.2007) US

(71) Applicant *(for all designated States except US)*: **THE GOVERNMENT OF U.S.A AS REPRESENTED BY THE SECRETARY OF THE DEPARTMENT OF HEALTH & HUMAN SERVICES**; National Institutes Of Health, Office Of Technology Transfer, 6011 Executive Boulevard, Suite 325, Rockville, MD 20852-3804 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KO, Minoru, S.h.** [US/US]; 10516 Willow Vista Way, Cockeysville, MD 21030 (US). **FALCO, Geppino** [IT/US]; 44 Pepperdine Circle, Baltimore, MD 21228 (US). **LEE, Sung-lim** [KR/US]; 912 Lowe Road, Baltimore, MD 21220 (US). **MONTI, Manuela** [IT/US]; 1507 W. Lombard Street, Baltimore, MD 21223 (US). **STANGHELLINI, Ilaria** [IT/US]; 3100 St. Paul Street, Apt. 314, Baltimore, MD 21218 (US).

(74) Agent: **CONNOLLY, Jodi, L.**; Klarquist Sparkman, Llp, One World Trade Center, Suite 1600, 121 Sw Salmon Street, Portland, OR 97204 (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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
5 March 2009

(54) Title: METHODS FOR MODULATING EMBRYONIC STEM CELL DIFFERENTIATION

(57) Abstract: Described herein is Zscan4, a gene exhibiting 2-cell embryonic stage and embryonic stem cell specific expression. Identification of nine Zscan4 co-expressed genes is also described. Inhibition of Zscan4 expression inhibits the 2-cell to 4-cell embryonic transition and prevents blastocyst implantation, expansion and outgrowth. Provided herein are methods of inhibiting differentiation of a stem cell, promoting blastocyst outgrowth of embryonic stem cells and identifying a subpopulation of stem cells expressing Zscan4. Further described is the identification of Trim43 as a gene exhibiting morula-specific expression. Also provided are isolated expression vectors comprising a Zscan4 promoter, or a Trim43 promoter operably linked to a heterologous polypeptide and uses thereof. Further provided are transgenic animals comprising transgenes encoding marker proteins operably linked to Zscan4 and Trim43 promoters.

WO 2008/118957 A3

METHODS FOR MODULATING EMBRYONIC STEM CELL DIFFERENTIATION

CROSS REFERENCE TO RELATED APPLICATION

This application claims the benefit of priority of U.S. Provisional Application No. 60/920,215, filed March 26, 2007, which is herein incorporated by reference in its entirety.

FIELD

This application relates to the field of cellular differentiation, specifically to the methods of identifying and using a subpopulation of stem cells, which can be identified by the expression of Zscan4 or one or more Zscan4 co-expressed genes described herein, and the methods of inhibiting differentiation and prolonging viability by altering Zscan4. This application also relates to the identification of Trim43 as a gene highly expressed at the morula stage.

BACKGROUND

Stem cells have been identified in several somatic tissues including the nervous system, bone marrow, epidermis, skeletal muscle, and liver. This 'set-aside' population of cells is believed to be responsible for maintaining homeostasis within individual tissues in adult animals. The number of stem cells and their decision to differentiate must be tightly controlled during embryonic development and in the adult animal to avoid premature aging or tumor formation. Different somatic stem cells share the properties of self-renewal and multi-developmental potential, suggesting the presence of common cellular machinery.

Embryonic stem (ES) cells can proliferate indefinitely in an undifferentiated state. Furthermore, ES cells are pluripotent cells, meaning that they can generate all of the cells present in the body (bone, muscle, brain cells, etc.). ES cells have been isolated from the inner cell mass of the developing murine blastocyst (Evans *et al.*, *Nature* 292:154-156, 1981; Martin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 78:7634-7636, 1981; Robertson *et al.*, *Nature* 323:445-448, 1986; Doetschman *et al.*, *Nature* 330:576-578, 1987; and Thomas *et al.*, *Cell* 51:503-512, 1987; U.S. Patent No.

5,670,372). Additionally, human cells with ES cell properties have recently been isolated from the inner blastocyst cell mass (Thomson *et al.*, *Science* 282:1145-1147, 1998) and developing germ cells (Shamblott *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 95:13726-13731, 1998) (see also U.S. Patent No. 6,090,622, PCT Publication Nos. WO 00/70021 and WO 00/27995).

There is growing interest in the analysis of patterns of gene expression in cells, such as stem cells. However, few studies have identified an individual gene product that functions in the complex network of signals in developing tissues to inhibit differentiation and increase proliferation.

SUMMARY

Described herein is the identification of Zscan4 as a gene specifically expressed during the 2-cell embryonic stage and in embryonic stem cells. Further described herein is the identification of Zscan4 co-expressed genes which exhibit a similar expression pattern as Zscan4 in the developing embryo. Also described herein is the identification of Trim43 as a gene abundantly expressed at the morula stage of embryonic development.

Provided herein are methods of inhibiting differentiation of a stem cell comprising increasing the expression of Zscan4 in the stem cell. In one embodiment, inhibiting differentiation of the stem cell increases viability of the stem cells. In another embodiment, inhibiting differentiation of the stem cell prevents senescence of the stem cell. As described herein, the stem cell can be any type of stem cell, including, but not limited to, an embryonic stem cell, an embryonic germ cell, a germline stem cell or a multipotent adult progenitor cell.

Also provided herein is a method of promoting blastocyst outgrowth of an embryonic stem cell, comprising increasing the expression of Zscan4 in the embryonic stem cell, thereby promoting blastocyst outgrowth of the embryonic stem cell.

Further provided is a method of identifying an undifferentiated subpopulation of stem cells expressing Zscan4, comprising transfecting stem cells with an expression vector comprising a Zscan4 promoter and a reporter gene, wherein expression of the reporter gene indicates Zscan4 is expressed in the

subpopulation of stem cells. In one embodiment, the promoter is a Zscan4c promoter.

An isolated expression vector comprising a Zscan4 promoter operably linked to a heterologous polypeptide is also provided. In one embodiment, the Zscan4 promoter is a Zscan4c promoter. In another embodiment, the heterologous polypeptide is a marker, enzyme or fluorescent protein. Also provided is an expression vector comprising a Trim43 promoter operably linked to a heterologous polypeptide. In some embodiments, the Trim43 promoter comprises at least a portion of the nucleic acid sequence set forth as SEQ ID NO: 31. Isolated embryonic stem cells comprising the expression vectors described herein are also provided.

Also provided is a method of identifying an undifferentiated subpopulation of stem cells, wherein the stem cells express Zscan4, comprising detecting expression of one or more of AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1. Isolated stem cells identified according to this method are also provided.

The foregoing and other features and advantages will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A is a series of digital images showing the expression profile of Zscan4 during preimplantation development by whole mount *in situ* hybridization. Hybridizations were performed simultaneously under the same experimental conditions for all preimplantation developmental stages. Images were taken at 200x magnification using phase contrast. Zscan4 shows a transient and high expression in the late 2-cell embryos. Such a high level of expression was not observed in 3-cell (two examples indicated by red arrows) and 4-cell embryos. **Figure 1B** shows a graph of the expression levels of Zscan4 during preimplantation development quantitated by qRT-PCR analysis. Three sets of 10 pooled embryos were collected from each stage (O, oocyte; 1, 1-cell embryo; E2, early 2-cell embryo; L2, late 2-cell

embryo; 4, 4-cell embryo; 8, 8-cell embryo; M, morula; and B, blastocyst) and used for qRT-PCR analysis. The expression levels of Zscan4 were normalized to *Chuk* control, and the average expression levels at each stage are represented as a fold change compared to the expression level in oocytes.

Figure 2A shows diagrams of the exon-intron structures of nine Zscan4 paralogs. New proposed gene symbols are shown in bold italics with the current gene symbols. **Figure 2B** illustrates the putative protein structures of Zscan4 paralogs, and shows predicted domains.

Figure 3A is a diagram that illustrates the genomic structure of the Zscan4 locus (encompassing 850 kb on Chromosome 7). The top panel shows genes near the Zscan4 locus. The lower panel shows nine Zscan4 paralogous genes and their characteristic features. Six other genes (LOCs) are predicted in this region, but unrelated to Zscan4. **Figure 3B** is a diagram that depicts the *TaqI*-, *MspI*-, or *TaqI/MspI*-digested DNA fragment sizes predicted from the genome sequences assembled from individual BAC sequences. **Figure 3C** is a digital image that shows the Southern blot analysis of C57BL/6J genomic DNAs digested with *TaqI*, *MspI*, or *TaqI/MspI* restriction enzymes. Sizes of all DNA fragments hybridized with a Zscan4 probe (containing only exon 3 from cDNA clone C0348C03) matched with those predicted in Figure 3B, validating the manually assembled sequences.

Figure 4A is a table showing the three types of siRNA technologies used for the analysis of Zscan4 in preimplantation embryos and their target sequences (SEQ ID NOs: 54-59). **Figure 4B** is a diagram that illustrates the locations of siRNA target sequences in the Zscan4 cDNA. **Figure 4C** is a series of digital images showing the development of shZscan4-injected embryos. The morphology of representative embryos is shown. Stages of shZscan4-injected and shControl-injected embryos were assessed at 61 hrs, 80 hrs, 98 hrs and 108 hrs post-hCG injections. **Figure 4D** is a series of graphs showing the percentage of shZscan4- and shControl-injected embryos at each developmental stage. shZscan4-injected (grey bars) and shControl-injected (white bars) were staged and counted at 61 hrs, 80 hrs, 98 hrs and 108 hrs post-hCG injections (M = morula; B = blastocyst). **Figure 4E** is a graph showing the transcript levels of Zscan4 in shControl-injected and shZscan4-

injected 2-cell embryos by qRT-PCR analysis. The expression levels were normalized by *Eef1a1*.

Figures 5A-5C are a series of graphs indicating the number of embryos at each developmental stage following injection with shZscan4. Embryos received shZscan4-injection in the nucleus of one blastomere of early 2-cell embryos. The stages of shZscan4- (gray) and shControl- (white) microinjected embryos were assessed at 52 hrs, 74 hrs and 96 hrs post-hCG injections. **Figures 5D-5F** show photographs of a 3-cell embryo (D), an unevenly cleaved embryo (E) and a mixed morula and blastocyst like embryo (F). The 3-cell embryo has one blastomere that remained at the size of a 2-cell stage blastomere and two smaller blastomeres with the size of 4-cell stage blastomeres. The 5-cell embryo has one delayed blastomere and four smaller blastomeres with the size of 8-cell blastomeres. These embryos eventually formed blastocyst-like structures, but seemed to be a mixture of a blastocyst-like cell mass and a morula-like cell mass. The morula-like cell mass was developed from one blastomere receiving shZscan4 injection, as shown by the presence of GFP, which was carried in the shZscan4 plasmid (**Figure 5G**). Magnification is 200x.

Figure 6A is an image that illustrates the expression of Zscan4 and *Pou5f1* in blastocysts, blastocyst outgrowth and ES cells by whole mount *in situ* hybridization. **Figure 6B** is a schematic illustration of the Zscan4 expression patterns.

Figures 7A-7E is a series of tables comparing nucleotide and amino acid sequence similarity (percent identity) among human *ZSCAN4*, mouse *Zscan4c, Zscan4d*, and *Zscan4f* genes.

Figure 8 is an illustration showing the Zscan4 syntenic regions of mouse and human genomes.

Figures 9A-9B is a series of graphs and photographs showing the development of embryos that received a siZscan4-injection in the cytoplasm. **Figure 9A** shows the percentage of embryos at each developmental stage for siControl-injected embryos (white bar) and siZscan4-injected embryos (gray bar) at 2.0, 3.5 and 4.0 d.p.c. **Figure 9B** shows the percentage of expanded and hatched

blastocysts at 4.5 d.p.c. in siControl-injected embryos (gray bar; photograph (a)) and siZscan4-injected embryos (black bar; photograph (b)).

Figures 10A-10D are a series of graphs and a table showing the development of embryos that received plus-siZscan4-injection in cytoplasm. **Figure 10A** shows the percentage of embryos at each developmental stage for siControl-injected embryos (white bar) and plus-siZscan4-injected embryos (gray bar) at 2.0, 2.2, 3.0, and 4.0 days post coitus. **Figures 10B and 10C** show the transcript levels of Zscan4 in siControl-injected embryos and plus-siZscan4-injected embryos, measured by qRT-PCR analysis and normalized by *Chuk* (Fig. 10B) and *H2afz* (Fig. 10C). **Figure 10D** provides the raw data of 3 biological replications of qRT-PCR analysis. †, the mean value of the cycle threshold for each biological replicate; ‡, the standard deviation.

Figure 11 is an illustration depicting the expression vector comprising the Zscan4c promoter sequence and reporter gene Emerald. The sequence of the expression vector is set forth as SEQ ID NO: 28.

Figure 12A is a fluorescence activated cell sorting (FACS) graph showing a subpopulation of mouse ES expressing Zscan4. Mouse ES cells were transfected with an expression vector comprising a Zscan4c promoter and a fluorescent reporter gene (Emerald). Expression of the reporter gene in a cell (an Emerald-positive cell) indicates the cell expresses Zscan4. **Figure 12B** is a graph showing expression levels of Zscan4c and Pou5f1 in the subpopulation of ES cells identified as Emerald-positive. The Y-axis represents the fold difference in gene expression between Emerald-positive and Emerald-negative cells.

Figures 13A-G are graphs showing expression profiles of Zscan4 and six genes co-expressed with Zscan4 in a sub-population of ES cells. Shown are the expression profiles of Zscan4 (A), AF067063 (B), Tcstv3 (C), Tho4 (D), Arginase II (E), BC061212 (F) and Gm428 (G)) in metaphase II oocytes (MII), 1 cell embryos, early 2 cell (e 2 cell) embryos, late 2 cell (l 2 cell) embryos, 4 cell embryos, 8 cell embryos, morula (mo) and blastocyts (bl).

SEQUENCE LISTING

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:

SEQ ID NOs: 1 and 2 are the nucleotide sequences of forward and reverse PCR primers for amplification of Zscan4d from 2-cell embryos.

SEQ ID NOs: 3 and 4 are the nucleotide sequences of PCR primers for amplifying a probe designed to contain exon 3 of Zscan4.

SEQ ID NO: 5 is the nucleotide sequence of the Zscan4 PCR and sequencing primer Zscan4_For.

SEQ ID NO: 6 is the nucleotide sequence of the Zscan4 PCR and sequencing primer Zscan4_Rev.

SEQ ID NO: 7 is the nucleotide sequence of the Zscan4 sequencing primer Zscan4_400Rev.

SEQ ID NO: 8 is the nucleotide sequence of the Zscan4 sequencing primer Zscan4_300Rev.

SEQ ID NO: 9 is the nucleotide sequence of the shZscan4 siRNA.

SEQ ID NO: 10 is the nucleotide sequence of the siControl siRNA.

SEQ ID NO: 11 is the nucleotide sequence of Genbank Accession No. BC050218 (deposited April 3, 2003), a cDNA clone derived from ES cells (Clone No. C0348C03).

SEQ ID NO: 12 is the nucleotide sequence of *Zscan4-ps1*.

SEQ ID NO: 13 is the nucleotide sequence of *Zscan4-ps2*.

SEQ ID NO: 14 is the nucleotide sequence of *Zscan4-ps3*.

SEQ ID NOs: 15 and 16 are the nucleotide and amino acid sequences of *Zscan4a*.

SEQ ID NOs: 17 and 18 are the nucleotide and amino acid sequences of *Zscan4b*.

SEQ ID NOs: 19 and 20 are the nucleotide and amino acid sequences of *Zscan4c*.

SEQ ID NOs: 21 and 22 are the nucleotide and amino acid sequences of *Zscan4d*.

SEQ ID NOs: 23 and 24 are the nucleotide and amino acid sequences of *Zscan4e*.

SEQ ID NOs: 25 and 26 are the nucleotide and amino acid sequences of *Zscan4f*.

SEQ ID NO: 27 is the nucleotide sequence of Genbank Accession No. XM_145358, deposited January 10, 2006, incorporated by reference herein.

SEQ ID NO: 28 is the nucleotide sequence of the Zscan4-Emerald expression vector.

SEQ ID NOs: 29 and 30 are the nucleotide and amino acid sequences of human ZSCAN4 (Genbank Accession No. NM_152677, deposited September 6, 2002, incorporated by reference herein).

SEQ ID NO: 31 is the nucleotide sequence of the Trim43 promoter.

SEQ ID NOs: 32 and 33 are the nucleotide and amino acid sequences of Trim43.

SEQ ID NOs: 34 and 35 are the nucleotide and amino acid sequences of AF067063, Genbank Accession No. NM_001001449, deposited May 29, 2004, incorporated by reference herein.

SEQ ID NOs: 36 and 37 are the nucleotide and amino acid sequences of BC061212, Genbank Accession No. NM_198667.1, deposited November 15, 2003, incorporated by reference herein.

SEQ ID NOs: 38 and 39 are the nucleotide and amino acid sequences of Gm428, Genbank Accession No. NM_001081644, deposited February 22, 2007, incorporated by reference herein.

SEQ ID NOs: 40 and 41 are the nucleotide and amino acid sequences of Arginase II, Genbank Accession No. NM_009705, deposited January 26, 2000, incorporated by reference herein.

SEQ ID NOs: 42 and 43 are the nucleotide and amino acid sequences of Tcstv1, Genbank Accession No. NM_018756, deposited July 12, 2007, incorporated by reference herein.

SEQ ID NOs: 44 and 45 are the nucleotide and amino acid sequences of Tcstv3, Genbank Accession No. NM_153523, deposited October 13, 2002, incorporated by reference herein.

SEQ ID NOs: 46 and 47 are the nucleotide and amino acid sequences of Tho4, Genbank Accession No. XM_902103, deposited December 2, 2005, incorporated by reference herein.

SEQ ID NOs: 48 and 49 are the nucleotide and amino acid sequences of Eif1a, Genbank Accession No. NM_010120, deposited August 3, 2002, incorporated by reference herein.

SEQ ID NOs: 50 and 51 are the nucleotide and amino acid sequences of EG668777, Genbank Accession No. XM_001003556, deposited April 27, 2006, incorporated by reference herein.

SEQ ID NOs: 52 and 53 are the nucleotide and amino acid sequences of Pif1, Genbank Accession No. NM_172453, deposited December 24, 2002, incorporated by reference herein.

SEQ ID NO: 54 is the nucleotide sequence of the Plus-siZscan4 (J-064700-05) target sequence.

SEQ ID NO: 55 is the nucleotide sequence of the Plus-siZscan4 (J-064700-06) target sequence.

SEQ ID NO: 56 is the nucleotide sequence of the Plus-siZscan4 (J-064700-07) target sequence.

SEQ ID NO: 57 is the nucleotide sequence of the Plus-siZscan4 (J-064700-08) target sequence.

SEQ ID NO: 58: is the nucleotide sequence of the siZscan4 target sequence.

SEQ ID NO: 59 is the nucleotide sequence of the of shZscan4 target sequence.

SEQ ID NO: 60 is the nucleotide consensus sequence of nucleotides 1-1848 of Zscan4c, Zscan4d and Zscan4f.

DETAILED DESCRIPTION

I. Abbreviations

CDS	Coding sequence
CMV	Cytomegalovirus
DNA	Deoxyribonucleic acid
d.p.c.	Days post coitus
EC	Embryonic carcinoma
EG	Embryonic germ
ES	Embryonic stem
GS	Germline stem
GFP	Green fluorescent protein
hCG	Human chorionic gonadotropin
ICM	Inner cell mass
IVF	*In vitro* fertilization
LIF	Leukemia inhibitory factor
maGSC	Multipotent adult germline stem cell
MAPC	Multipotent adult progenitor cell
PCR	Polymerase chain reaction
qRT-PCR	Quantitative reverse-transcriptase polymerase chain reaction
RNA	Ribonucleic acid
siRNA	small interfering RNA
TS	Trophoblast stem
USSC	Unrestricted somatic stem cell
ZGA	Zygotic genome activation

II. Terms

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).

In order to facilitate review of the various embodiments of the invention, the following explanations of specific terms are provided:

Alter: A change in an effective amount of a substance of interest, such as a polynucleotide or polypeptide. The amount of the substance can be changed by a difference in the amount of the substance produced, by a difference in the amount of the substance that has a desired function, or by a difference in the activation of the substance. The change can be an increase or a decrease. The alteration can be *in vivo* or *in vitro*. In several embodiments, altering an effective amount of a polypeptide or polynucleotide is at least about a 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% increase or decrease in the effective amount (level) of a substance. Altering an effective amount of a polypeptide or polypeptide includes increasing the expression of Zscan4 in a cell. In another embodiment, an alteration in a polypeptide or polynucleotide affects a physiological property of a cell, such as the differentiation, proliferation, or viability of the cell. For example, increasing expression of Zscan4 in a stem cell inhibits differentiation and promotes viability of the stem cell.

Blastocyst: The structure formed in early mammalian embryogenesis, after the formation of the blastocele, but before implantation. It possesses an inner cell mass, or embryoblast, and an outer cell mass, or trophoblast. The human blastocyst comprises 70-100 cells. As used herein, blastocyst outgrowth refers to the process of culturing embryonic stem cells derived from the inner cell mass of a blastocyst. Promoting blastocyst outgrowth refers to enhancing the viability and proliferation of embryonic stem cells derived from the blastocyst.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and regulatory sequences that determine transcription. cDNA is synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

Co-expressed: In the context of the present disclosure, genes that are "co-expressed" with Zscan4 (also referred to as "Zscan4 co-expressed genes") are genes that exhibit a similar expression pattern as Zscan4 during embryonic development

and in ES cells. Specifically, the co-expressed genes are expressed in the same undifferentiated subpopulation of ES cells as Zscan 4, and during embryonic development, these genes are most abundantly expressed at the 2-cell stage. Nine co-expressed genes are described herein, including AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1. However, co-expressed genes are not limited to those disclosed herein, but include any genes exhibiting an expression pattern similar to Zscan4.

AF067063 encodes hypothetical protein LOC380878. The full length cDNA sequence of AF067063 (SEQ ID NO: 34) is 886 base pairs in length and is organized into three exons encoding several hypothetical proteins (for example, SEQ ID NO: 35), which appear to be mouse specific.

BC061212 encodes a protein belonging to the PRAME (preferentially expressed antigen melanoma) family. The full length cDNA sequence of BC061212 (SEQ ID NO: 36) is 1625 base pairs in length and is organized into four exons, encoding a protein of 481 residues in length (SEQ ID NO: 37).

Gm428 (gene model 428) encodes a hypothetical protein. The full length cDNA sequence of Gm428 (SEQ ID NO: 38) is 1325 base pairs in length and is organized into five exons encoding a protein of 360 residues in length (SEQ ID NO: 39).

Arginase II belongs to the Arginase family and may play a role in the regulation of extra-urea cycle arginine metabolism, and in down-regulation of nitric oxide synthesis. The full length cDNA sequence of Arginase II (SEQ ID NO: 40) is 1415 base pairs in length and is organized into eight exons encoding a protein of 354 residues in length (SEQ ID NO: 41).

Tsctv1 and Tsctv3 are splice variants. The full length cDNA of Tsctv1 (SEQ ID NO: 42) is 858 base pairs in length and contains two exons encoding a protein of 171 residues (SEQ ID NO: 43). The full length cDNA sequence of Tsctv3 (SEQ ID NO: 44) is 876 base pairs in length and contains one exon encoding a protein of 169 residues (SEQ ID NO: 45). This family of proteins consists of several hypothetical proteins of approximately 170 residues in length and appears to be mouse-specific.

Tho4 (also called EG627488) encodes a protein with an RNA recognition motif (RRM) involved in regulation of alternative splicing, and protein components

of small nuclear ribonucleoproteins (snRNPs). The full length cDNA sequence of Tho4 (SEQ ID NO: 46) is 811 base pairs in length and is organized into three exons encoding a protein of 163 residues in length (SEQ ID NO: 47).

Eif1a belongs to the eukaryotic translation initiation factor family. The full length cDNA sequence of Eif1a (SEQ ID NO: 48) is 2881 base pairs in length and encodes a protein of 144 amino acids (SEQ ID NO: 49).

EG668777 is a predicted gene having similarity to retinoblastoma-binding protein 6, isoform 2. The full length cDNA sequence of EG668777 is 1918 base pairs in length (SEQ ID NO: 50) and contains one exon encoding a protein of 547 residues (SEQ ID NO: 51).

Pif1 is an ATP-dependent DNA helicase. The full length cDNA sequence of Pif1 (SEQ ID NO: 52) is 3680 base pairs in length and contains 12 exons encoding a protein of 650 amino acids (SEQ ID NO: 53).

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.

Differentiation: Refers to the process by which a cell develops into a specific type of cell (for example, muscle cell, skin cell etc.). In the context of the present disclosure, 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. As used herein, inhibiting differentiation means preventing or slowing the development of a cell into a specific lineage.

Embryonic stem (ES) cells: Pluripotent cells isolated from the inner cell mass of the developing blastocyst. "ES cells" can be derived from any organism. ES cells can be derived from mammals. In one embodiment, ES cells are produced from mice, rats, rabbits, guinea pigs, goats, pigs, cows, monkeys and humans. Human and murine derived ES cells are preferred. 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 in U.S. Patent No. 5,670,372, herein incorporated by reference. Methods for producing human ES cells can be found in U.S. Patent No. 6,090,622, PCT Publication No. WO 00/70021 and PCT Publication No. WO 00/27995, herein incorporated by reference.

Expand: A process by which the number or amount of cells in a cell culture is increased due to cell division. Similarly, the terms "expansion" or "expanded" refers to this process. The terms "proliferate," "proliferation" or "proliferated" may be used interchangeably with the words "expand," "expansion", or "expanded." Typically, during expansion, the cells do not differentiate to form mature cells.

Expression vector: A vector is a nucleic acid molecule allowing insertion of foreign nucleic acid without disrupting the ability of the vector to replicate and/or integrate in a host cell. A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker genes and other genetic elements. An expression vector is a vector that contains the necessary regulatory sequences to allow transcription and translation of inserted gene or genes.

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

Host cells: Cells in which a vector can be propagated and its DNA expressed. The cell may be prokaryotic or eukaryotic. 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.

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.

Multipotent cell: Refers to a cell that can form multiple cell lineages, but not all cell lineages.

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.

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.

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 fusion proteins herein disclosed.

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.

Pharmaceutical agent: A chemical compound, small molecule, or other composition capable of inducing a desired therapeutic or prophylactic effect when properly administered to a subject or a cell. "Incubating" includes a sufficient amount of time for a drug to interact with a cell. "Contacting" includes incubating a drug in solid or in liquid form with a cell.

Pluripotent cell: Refers to 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.

Polynucleotide: A nucleic acid sequence (such as a linear sequence) of any length. Therefore, a polynucleotide includes oligonucleotides, and also gene sequences found in chromosomes. An "oligonucleotide" is a plurality of joined nucleotides joined by native phosphodiester bonds. An oligonucleotide is a polynucleotide of between 6 and 300 nucleotides in length. An oligonucleotide analog refers to moieties that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules.

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.

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.

The term "soluble" refers to a form of a polypeptide that is not inserted into a cell membrane.

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.

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

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 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.

Primers: Short nucleic acids, for example DNA oligonucleotides ten nucleotides or more in length, which are annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Probes and primers as used herein may, for example, include at least 10 nucleotides of the nucleic acid sequences that are shown to encode specific proteins.

In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100 consecutive nucleotides of the disclosed nucleic acid sequences. Methods for preparing and using probes and primers are described in the references, for example Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York; Ausubel *et al.* (1987) *Current Protocols in Molecular Biology*, Greene Publ. Assoc. & Wiley-Intersciences; Innis *et al.* (1990) *PCR Protocols, A Guide to Methods and Applications*, Innis *et al.* (Eds.), Academic Press, San Diego, CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge, MA).

When referring to a probe or primer, the term "specific for (a target sequence)" indicates that the probe or primer hybridizes under stringent conditions substantially only to the target sequence in a given sample comprising the target sequence.

Prolonging viability: As used herein, "prolonging viability" of a stem cell refers to extending the duration of time a stem cell is capable of normal growth and/or survival.

Promoter: A promoter is an array of 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).

Reporter gene: A reporter gene is a gene operably linked to another gene or nucleic acid sequence of interest (such as a promoter sequence). Reporter genes are used to determine whether the gene or nucleic acid of interest is expressed in a cell or has been activated in a cell. Reporter genes typically have easily identifiable characteristics, such as fluorescence, or easily assayed products, such as an enzyme. Reporter genes can also confer antibiotic resistance to a host cell. In one

embodiment, the reporter gene encodes the fluorescent protein Emerald. In another embodiment, the reporter gene encodes the fluorescent protein Strawberry.

Senescence: The inability of a cell to divide further. A senescent cell is still viable, but does not divide.

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 and USSCs. 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.

Subpopulation: An identifiable portion of a population. As used herein, a "subpopulation" of stem cells expressing Zscan4 is the portion of stem cells in a given population that has been identified as expressing Zscan4. In one embodiment, the subpopulation is identified using an expression vector comprising a Zscan4 promoter and a reporter gene, wherein detection of expression of the reporter gene in a cell indicates the cell expresses Zscan4 and is part of the subpopulation. As described herein, the subpopulation of ES cells expressing Zscan4 can further be identified by co-expression of one or more genes disclosed herein, including AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1.

Totipotent cell: Refers to a cell that can form an entire organism autonomously. Only a fertilized egg (oocyte) possesses this ability (stem cells do not).

Transgenic animal: A non-human animal, usually a mammal, having a non-endogenous (heterologous) nucleic acid sequence present as an extrachromosomal element in a portion of its cells or stably integrated into its germ

line DNA (*i.e.*, in the genomic sequence of most or all of its cells). Heterologous nucleic acid is introduced into the germ line of such transgenic animals by genetic manipulation of, for example, embryos or embryonic stem cells of the host animal according to methods well known in the art. A "transgene" is meant to refer to such heterologous nucleic acid, such as, heterologous nucleic acid in the form of an expression construct (such as for the production of a "knock-in" transgenic animal) or a heterologous nucleic acid that upon insertion within or adjacent to a target gene results in a decrease in target gene expression (such as for production of a "knock-out" transgenic animal).

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.

Trim43 (tripartite motif-containing protein 43): A gene identified herein as exhibiting morula-specific expression during embryonic development. The nucleotide and amino acid sequences of Trim43 are provided herein as SEQ ID NO: 32 and SEQ ID NO: 33, respectively.

Zscan4: A group of genes identified herein as exhibiting 2-cell embryonic stage and ES cell-specific expression. In the mouse, the term "Zscan4" refers to a collection of genes including three pseudogenes (Zscan1-ps1, Zscan4-ps2 and Zscan4-ps3) and six expressed genes (Zscan4a, Zscan4b, Zscan4c, Zscan4d, Zscan4e and Zscan4f). As used herein, Zscan4 also includes human ZSCAN4. Zscan4 refers to Zscan4 polypeptides and Zscan4 polynucleotides encoding the Zscan4 polypeptides.

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 invention 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 invention, 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. Overview of Several Embodiments

Disclosed herein are Zscan4 polypeptides and polynucleotides encoding these polypeptides, which are of use in inhibiting differentiation and increasing proliferation of cells, such as stem cells, including embryonic stem cells. Stem cells, especially ES cells in the undifferentiated condition, were previously considered to be a relatively homogenous cell population. However, described herein is the unique expression of Zscan4 in a subpopulation of stem cells, which establishes the presence of a unique cell population among undifferentiated ES cells and provides the means to identify and isolate these cells. Also described herein is the identification of nine genes co-expressed with Zscan4 in the undifferentiated ES cell subpopulation. These genes include AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1. Further described herein is the identification of Trim43 as a gene exhibiting morula-specific gene expression.

It is disclosed herein that Zscan4 is specifically expressed during the 2-cell embryonic stage and in a subpopulation of embryonic stem cells. There is a genus of Zscan4-related genes, including three pseudogenes (*Zscan4-ps1*, *Zscan4-ps2* and *Zscan4-ps3*) and six expressed genes (*Zscan4a*, *Zscan4b*, *Zscan4c*, *Zscan4d*, *Zscan4e* and *Zscan4f*). The Zscan4 genus also includes human ZSCAN4. It is further disclosed herein that AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1 are co-expressed with Zscan4 during embryonic development. Like Zscan4, during embryonic development, these genes are expressed most abundantly at the 2-cell stage.

Methods are provided herein for inhibiting differentiation of a stem cell comprising increasing the expression of Zscan4 in the stem cell. As described herein, the use of Zscan4 includes the use of any Zscan4 gene, including Zscan4a, Zscan4b, Zscan4c, Zscan4d, Zscan4e, Zscan4f and human ZSCAN4. In some embodiments, the Zscan4 gene is at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to Zscan4c (SEQ ID NO: 19), Zscan4d (SEQ ID NO: 21) or Zscan4f (SEQ ID NO: 25). In another embodiment, the Zscan4 gene comprises SEQ ID NO: 60.

Increasing expression of Zscan4 in a cell, such as a stem cell, can be achieved according to any number of methods well known in the art. In one embodiment, increasing expression of Zscan4 in a stem cell comprises transfecting the stem cell with a nucleotide encoding Zscan4 operably linked to a promoter. The promoter can be any type of promoter, including a constitutive promoter or an inducible promoter. In one embodiment, the stem cells are transfected with a vector comprising the nucleotide sequence encoding Zscan4 operably linked to the promoter. The vector can be any type of vector, such as a viral vector or a plasmid vector. In one embodiment, the Zscan4 nucleotide sequence is at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to Zscan4c (SEQ ID NO: 19), Zscan4d (SEQ ID NO: 21) or Zscan4f (SEQ ID NO: 25). In another embodiment, the Zscan4 nucleotide sequence comprises SEQ ID NO: 60.

In one embodiment of the methods described herein, inhibiting differentiation of the stem cell increases viability of the stem cells. In another embodiment, inhibiting differentiation of the stem cell prevents senescence of the stem cell. As described herein, the stem cell can be any type of stem cell, including, but not limited to, an embryonic stem cell, an embryonic germ cell, a germline stem cell or a multipotent adult progenitor cell.

Also provided herein is a method of promoting blastocyst outgrowth of an embryonic stem cell, comprising increasing the expression of Zscan4 in the embryonic stem cell, thereby promoting blastocyst outgrowth of the embryonic stem cell. Promoting blastocyst outgrowth can include increasing the efficiency of outgrowth or increasing the number of embryonic stem cells resulting from blastocyst outgrowth. In one embodiment, the method comprises increasing

expression of Zscan4 in the cells during the early stages of blastocyst outgrowth, such as prior to proliferation of the stem cells. As described herein, Zscan4 includes any Zscan4 gene, including Zscan4a, Zscan4b, Zscan4c, Zscan4d, Zscan4e, Zscan4f, and human ZSCAN4. In one embodiment, the Zscan4 gene is at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to Zscan4c (SEQ ID NO: 19), Zscan4d (SEQ ID NO: 21) or Zscan4f (SEQ ID NO: 25). In another embodiment, the Zscan4 gene comprises SEQ ID NO: 60.

In one embodiment, increasing the expression of Zscan4 in the stem cell comprises transfecting the stem cell with a nucleotide sequence encoding a Zscan4 operably linked to a promoter. The promoter can be any type of promoter, including an inducible promoter or a constitutive promoter. In one embodiment, the cells are transfected with a vector comprising the nucleotide encoding Zscan4 operably linked to a promoter. The vector can be any type of vector, including a viral vector or a plasmid vector.

A method is also provided for identifying a subpopulation of stem cells expressing Zscan4, comprising transfecting the cells with an expression vector comprising a Zscan4 promoter and a reporter gene, wherein expression of the reporter gene indicates Zscan4 is expressed in the subpopulation of stem cells. In one embodiment, the promoter is a Zscan4c promoter. In another embodiment, the Zscan4c promoter includes the nucleic acid sequence set forth as nucleotides 1-2540 of SEQ ID NO: 28, such as nucleotides 1-2643, 1-3250, or 1-3347 of SEQ ID NO: 28. In another embodiment, the expression vector comprises the nucleic acid sequence set forth as SEQ ID NO: 28. As described herein, the subpopulation of ES cells expressing Zscan4 are in an undifferentiated state. Further provided is a method of identifying the undifferentiated subpopulation of ES cells by detecting expression of one or more Zscan4 co-expressed genes, such as AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1. Detection of expression of these genes can be accomplished using any means well known in the art, such as, for example, RT-PCR, Northern blot or *in situ* hybridization. Further provided are isolated stem cells identified according to this method.

An isolated expression vector comprising a Zscan4 promoter operably linked to a nucleic acid sequence encoding a heterologous polypeptide is also provided. In one embodiment, the Zscan4 promoter is a Zscan4c promoter. In another embodiment, the Zscan4c promoter comprises the nucleic acid sequence set forth as nucleotides 1-2540 of SEQ ID NO: 28, such as nucleotides 1-2643, 1-3250, or 1-3347 of SEQ ID NO: 28. In another embodiment, the heterologous polypeptide is a marker, enzyme or fluorescent protein. The expression vector can be any type of vector, including, but not limited to a viral vector or a plasmid vector.

Further provided herein is an ES cell line comprising an expression vector comprising a Zscan4 promoter operably linked to a heterologous polypeptide. In one embodiment, the Zscan4 promoter is a Zscan4c promoter. In another embodiment, the Zscan4c promoter comprises the nucleic acid sequence set forth as nucleotides 1-2540 of SEQ ID NO: 28, such as nucleotides 1-2643, 1-3250, or 1-3347 of SEQ ID NO: 28. In another embodiment, the heterologous polypeptide is a marker, enzyme or fluorescent protein. In one example, the fluorescent protein is Emerald.

An isolated expression vector comprising a Trim43 promoter operably linked to a nucleic acid sequence encoding a heterologous polypeptide is also provided. In one embodiment, the Trim43 promoter comprises at least a portion of the nucleic acid sequence set forth as SEQ ID NO: 31. The portion of SEQ ID NO: 31 to be included in the expression vector is at least a portion of SEQ ID NO: 31 that is capable of promoting transcription of the heterologous polypeptide in a cell in which Trim43 is expressed. In some embodiments, the Trim43 promoter sequence is at least 70%, at least 80%, at least 90%, at least 95% or at least 99% identical to SEQ ID NO: 31. In another embodiment, the Trim43 promoter comprises SEQ ID NO: 31. In another embodiment, the Trim43 promoter consists of SEQ ID NO: 31. In some embodiments, the heterologous polypeptide is a marker, enzyme or fluorescent protein. In one example the fluorescent protein is Strawberry. The expression vector can be any type of vector, including, but not limited to a viral vector or a plasmid vector.

Further provided herein is an ES cell line containing an expression vector comprising a Trim43 promoter operably linked to a heterologous polypeptide. In

one embodiment, the Trim43 promoter comprises at least a portion of the nucleic acid sequence set forth as SEQ ID NO: 31. In some embodiments, the Trim43 promoter sequence is at least 70%, at least 80%, at least 90%, at least 95% or at least 99% identical to SEQ ID NO: 31. In another embodiment, the Trim43 promoter comprises SEQ ID NO: 31. In another embodiment, the Trim43 promoter consists of SEQ ID NO: 31. In another embodiment, the heterologous polypeptide is a marker, enzyme or fluorescent protein. In one example, the fluorescent protein is Strawberry.

Provided herein are antibodies specific for Zscan4. In one embodiment, the Zscan4 antibodies specifically recognize Zscan4a, Zscan4b, Zscan4c, Zscan4d, Zscan4e, Zscan4f or human ZSCAN4. Also provided are antibodies specific for each Zscan4 co-expressed gene, including antibodies raised against at least a portion of a polypeptide encoded by AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 or Pif1.

Also described herein are transgenic animals harboring a transgene that includes the Zscan4 polynucleotide sequences disclosed herein. Also provided are transgenic animals harboring a transgene that includes polynucleotide sequences of one or more of the Zscan4 co-expressed genes. Such transgenic animals include, but are not limited to, transgenic mice.

Further provided is a transgenic non-human animal comprising a nucleic acid sequence (a transgene) encoding a heterologous polypeptide operably linked to a Zscan4 promoter. In some embodiments, the heterologous polypeptide is a marker, enzyme or fluorescent protein. In one embodiment, the heterologous polypeptide is fluorescent protein. In one example, the fluorescent protein is Emerald. In one embodiment, the Zscan4 promoter is a Zscan4c promoter. In another embodiment, the Zscan4c promoter comprises the nucleic acid sequence set forth as nucleotides 1-2540 of SEQ ID NO: 28, such as nucleotides 1-2643, 1-3250, or 1-3347 of SEQ ID NO: 28.

In another embodiment, the transgenic non-human animal further comprises a nucleic acid sequence encoding a heterologous polypeptide operably linked to a Trim43 promoter. In one embodiment, the Trim43 promoter comprises the nucleic acid sequence set forth as SEQ ID NO: 31. The heterologous polypeptide can be,

for example, a marker, enzyme or fluorescent protein. In one embodiment, the heterologous polypeptide is a fluorescent protein. In one example, the fluorescent protein is Strawberry. In some embodiments, the transgenic non-human animal is a transgenic mouse.

Also provided herein are isolated embryonic stem cells obtained from an embryo of the transgenic non-human animal. In one embodiment, the transgenic non-human animal is a transgenic mouse.

IV. Methods of Inducing Differentiation and/or Inhibiting Proliferation of Stem Cells

A method for inhibiting differentiation of a stem cell is disclosed herein. A method for increasing viability and/or inducing proliferation of a stem cell is also disclosed herein. A method is also provided herein for inhibiting senescence of a stem cell. The methods include altering the level of a Zscan4 polypeptide in the cell, thereby inhibiting differentiation and/or inducing proliferation of the cell, and/or inhibiting senescence of the cell. The cell can be *in vivo* or *in vitro*.

It is shown herein that inhibiting Zscan4 in embryos blocks the 2- to 4-cell stage embryonic transition. Inhibition of Zscan4 expression also prevents blastocysts from expanding and implanting and prevents the outgrowth of embryonic stem cells from blastocysts. In addition, in embryonic stem cells, Zscan4 expression is only detected in a subpopulation of undifferentiated stem cells. Thus, expression of Zscan4 plays an important role in maintaining ES cells in an undifferentiated state, which is necessary for ES cell viability and proliferation. Zscan4 is also important in allowing outgrowth of ES cells from blastocysts. Therefore, provided herein are methods of increasing expression of Zscan4 in a stem cell to inhibit differentiation, increase viability and prevent senescence of a stem cell. The methods provided herein also include increasing expression of Zscan4 to promote blastocyst outgrowth of ES cells.

Expression of Zscan4 can be increased to inhibit differentiation and/or induce proliferation. In one example, expression of Zscan4 is increased as compared to a control. Increased expression includes, but is not limited to, at least a 20% increase in the amount of Zscan4 mRNA or polypeptide in a cell as compared

to a control, such as, but not limited to, at least about a 30%, 50%, 75%, 100%, or 200% increase of Zscan4 mRNA or polypeptide. Suitable controls include a cell not contacted with an agent that alters Zscan4 expression, or not transfected with a vector encoding Zscan4, such as a wild-type stem cell. Suitable controls also include standard values. Exemplary Zscan4 amino acid sequences are set forth in the Sequence Listing as SEQ ID NO: 16 (Zscan4a), SEQ ID NO: 18 (Zscan4b), SEQ ID NO: 20 (Zscan4c), SEQ ID NO: 22 (Zscan4d), SEQ ID NO: 24 (Zscan4e), SEQ ID NO: 26 (Zscan4f) and SEQ ID NO: 30 (human ZSCAN4).

Specific, non-limiting examples of Zscan4 polypeptides include polypeptides including an amino acid sequence at least about 80%, 85%, 90%, 95%, or 99% homologous to the amino acid sequence set forth in SEQ ID NO: 16, 18, 20, 22, 24, 26 or 30. In a further embodiment, a Zscan4 polypeptide is a conservative variant of SEQ ID NO: 16, 18, 20, 22, 24, 26 or 30, 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: 16, 18, 20, 22, 24, 26 or 30. In another embodiment, a Zscan4 polypeptide has an amino acid sequence as set forth in SEQ ID NO: 16, 18, 20, 22, 24, 26 or 30.

Fragments and variants of a Zscan4 polypeptide can readily be prepared by one of skill in the art using molecular techniques. In one embodiment, a fragment of a Zscan4 polypeptide includes at least 8, 10, 15, or 20 consecutive amino acids of the Zscan4 polypeptide. In another embodiment, a fragment of a Zscan4 polypeptide includes a specific antigenic epitope found on a full-length Zscan4. In a further embodiment, a fragment of Zscan4 is a fragment that confers a function of Zscan4 when transferred into a cell of interest, such as, but not limited to, inhibiting differentiation or increasing proliferation of the cell.

One skilled in the art, given the disclosure herein, can purify a Zscan4 polypeptide using standard techniques for protein purification. The substantially pure polypeptide will yield a single major band on a non-reducing polyacrylamide gel. The purity of the Zscan4 polypeptide can also be determined by amino-terminal amino acid sequence analysis.

Minor modifications of the Zscan4 polypeptide primary amino acid sequences may result in peptides which have substantially equivalent activity as compared to the unmodified counterpart polypeptide described herein. Such modifications may be deliberate, as by site-directed mutagenesis, or may be spontaneous. All of the polypeptides produced by these modifications are included herein.

One of skill in the art can readily produce fusion proteins including a Zscan4 polypeptide and a second polypeptide of interest. Optionally, a linker can be included between the Zscan4 polypeptide and the second polypeptide of interest. Fusion proteins include, but are not limited to, a polypeptide including a Zscan4 polypeptide and a marker protein. In one embodiment, the marker protein can be used to identify or purify a Zscan4 polypeptide. Exemplary fusion proteins include, but are not limited to, green fluorescent protein, six histidine residues, or myc and a Zscan4 polypeptide.

Polynucleotides encoding a Zscan4 polypeptide are also provided, and are termed *Zscan4* polynucleotides. These polynucleotides include DNA, cDNA and RNA sequences which encode a Zscan4. It is understood that all polynucleotides encoding a Zscan4 polypeptide are also included herein, as long as they encode a polypeptide with the recognized activity, such as the binding to an antibody that recognizes a Zscan4 polypeptide, or modulating cellular differentiation or proliferation. The polynucleotides 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. A Zscan4 polynucleotide encodes a Zscan4 polypeptide, as disclosed herein. Exemplary polynucleotide sequences encoding Zscan4 are set for in the Sequence Listing as SEQ ID NO: 12 (Zscan4-ps1), SEQ ID NO: 13 (Zscan4-ps2), SEQ ID NO: 14 (Zscan4-ps3), SEQ ID NO: 15 (Zscan4a), SEQ ID NO: 17 (Zscan4b), SEQ ID NO: 19(Zscan4c), SEQ ID NO: 21 (Zscan4d), SEQ ID NO: 23 (Zscan4e), SEQ ID NO: 25 (Zscan4f) and SEQ ID NO: 29 (human ZSCAN4).

In some embodiments, the Zscan4 polynucleotide sequence is at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to Zscan4c (SEQ ID NO: 19), Zscan4d (SEQ ID NO: 21) or Zscan4f (SEQ ID NO: 25). In another embodiment, the Zscan4 gene comprises SEQ ID NO: 60.

The *Zscan4* polynucleotides include recombinant DNA which is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA) independent of other sequences. The nucleotides can be ribonucleotides, deoxyribonucleotides, or modified forms of either nucleotide. The term includes single- and double-stranded forms of DNA. Also included in this disclosure are fragments of the above-described nucleic acid sequences that are at least 15 bases in length, which is sufficient to permit the fragment to selectively hybridize to DNA that encodes the disclosed *Zscan4* polypeptide (e.g., a polynucleotide that encodes SEQ ID NO: 16, 18, 20, 22, 24, 26 or 30) under physiological conditions. The term "selectively hybridize" refers to hybridization under moderately or highly stringent conditions, which excludes non-related nucleotide sequences.

Also contemplated herein is the use of a Zscan4 polynucleotide, or the complement of a Zscan4 polynucleotide, for RNA interference. Fragments of Zscan4 polynucleotides or their complements can be designed as siRNA molecules to inhibit expression of one or more Zscan4 proteins. In one embodiment, the siRNA compounds are fragments of a Zscan4 pseudogene. Methods of preparing and using siRNA are generally disclosed in U.S. Patent 6,506,559, incorporated herein by reference (*see also* reviews by Milhavet *et al.*, *Pharmacological Reviews* 55:629-648, 2003; and Gitlin *et al.*, *J. Virol.* 77:7159-7165, 2003; incorporated herein by reference). The double-stranded structure of siRNA can be formed by a single self-complementary RNA strand or two complementary RNA strands.

The siRNA can comprise one or more strands of polymerized ribonucleotide, and may include modifications to either the phosphate-sugar backbone or the nucleoside. For example, the phosphodiester linkages of natural RNA can be modified to include at least one of a nitrogen or sulfur heteroatom. Modifications in RNA structure can be tailored to allow specific genetic inhibition while avoiding a

general panic response in some organisms which is generated by dsRNA. Likewise, bases can be modified to block the activity of adenosine deaminase.

Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition. Nucleic acid containing a nucleotide sequence identical to a portion of a target sequence can be used for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Sequence identity may be optimized by alignment algorithms known in the art and calculating the percent difference between the nucleotide sequences. Alternatively, the duplex region of the RNA can be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript.

Sequence identity can optimized by sequence comparison and alignment algorithms known in the art (see Gribskov and Devereux, Sequence Analysis Primer, Stockton Press, 1991, and references cited therein) and calculating the percent difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters *(e.g.*, University of Wisconsin Genetic Computing Group). Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory RNA and the portion of particular target gene sequence is preferred. Alternatively, the duplex region of the RNA can be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the particular target gene (*e.g.*, 400 mM NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA, 50°C. or 70°C. hybridization for 12-16 hours; followed by washing). The length of the identical nucleotide sequences may be at least 20, 25, 50, 100, 200, 300 or 400 bases. A 100% sequence identity between the RNA and Zscan4 is not required to practice the present methods.

For siRNA (RNAi), the RNA can be directly introduced into the cell (such as intracellularly); or introduced extracellularly into a cavity, interstitial space, into the circulation of an organism, introduced orally, or may be introduced by bathing an organism in a solution containing RNA. Physical methods of introducing nucleic acids include injection of a solution containing the RNA, bombardment by particles covered by the RNA, soaking the cell or organism in a solution of the RNA, or

electroporation of cell membranes in the presence of the RNA. A viral construct packaged into a viral particle can efficiently introduce an expression construct into the cell can provide transcription of RNA encoded by the expression construct. Other methods known in the art for introducing nucleic acids to cells may be used, such as lipid-mediated carrier transport, chemical-mediated transport, such as calcium phosphate, and the like. Thus, the RNA may be introduced along with components that perform one or more of the following activities: enhance RNA uptake by the cell, promote annealing of the duplex strands, stabilize the annealed strands, or other-wise increase inhibition of the target gene.

RNA may be synthesized either *in vivo* or *in vitro*. Endogenous RNA polymerase of the cell can mediate transcription *in vivo*, or cloned RNA polymerase can be used for transcription *in vivo* or *in vitro*. For transcription from a transgene *in vivo* or an expression construct, a regulatory region can be used to transcribe the RNA strand (or strands). RNA may be chemically or enzymatically synthesized by manual or automated reactions. The RNA may be synthesized by a cellular RNA polymerase or a bacteriophage RNA polymerase (for example, T3, T7, SP6). The use and production of expression constructs are known in the art (for example, PCT Publication No. WO 97/32016; U.S. Pat. Nos: 5,593,874, 5,698,425, 5,712,135, 5,789,214, and 5,804,693; and the references cited therein). If synthesized chemically or by *in vitro* enzymatic synthesis, the RNA can be purified prior to introduction into the cell. For example, RNA can be purified from a mixture by extraction with a solvent or resin, precipitation, electrophoresis, chromatography, or a combination thereof. Alternatively, the RNA can be used with no or a minimum of purification to avoid losses due to sample processing. The RNA can be dried for storage or dissolved in an aqueous solution. The solution can contain buffers or salts to promote annealing, and/or stabilization of the duplex strands.

A polynucleotide encoding Zscan4 can be included in an expression vector to direct expression of the Zscan4 nucleic acid sequence. Thus, other expression control sequences including appropriate promoters, enhancers, transcription terminators, a start codon (i.e., ATG) in front of a protein-encoding gene, splicing signal for introns, maintenance of the correct reading frame of that gene to permit proper translation of mRNA, and stop codons can be included in an expression

vector. Generally expression control sequences include a promoter, a minimal sequence sufficient to direct transcription.

The expression vector typically contains an origin of replication, a promoter, as well as specific genes which allow phenotypic selection of the transformed cells (e.g. an antibiotic resistance cassette). Vectors suitable for use include, but are not limited, to the pMSXND expression vector for expression in mammalian cells (Lee and Nathans, *J. Biol. Chem.* 263:3521, 1988). Generally, the expression vector will include a promoter. The promoter can be inducible or constitutive. The promoter can be tissue specific. Suitable promoters include the thymidine kinase promoter (TK), metallothionein I, polyhedron, neuron specific enolase, thyrosine hyroxylase, beta-actin, or other promoters. In one embodiment, the promoter is a heterologous promoter.

In one example, the polynucleotide encoding Zscan4 is located downstream of the desired promoter. Optionally, an enhancer element is also included, and can generally be located anywhere on the vector and still have an enhancing effect. However, the amount of increased activity will generally diminish with distance.

Expression vectors including a polynucleotide encoding Zscan4 can be used to transform host cells. Hosts can include isolated microbial, yeast, insect and mammalian cells, as well as cells located in the organism. Biologically functional viral and plasmid DNA vectors capable of expression and replication in a host are known in the art, and can be used to transfect any cell of interest. Where the cell is a mammalian cell, the genetic change is generally achieved by introduction of the DNA into the genome of the cell (i.e., stable) or as an episome. Thus, host cells can be used to produce Zscan4 polypeptides. Alternatively, expression vectors can be used to transform host cells of interest, such as stem cells.

A "transfected cell" is a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule encoding Zscan4. Transfection of a host cell with recombinant DNA may be carried out by conventional techniques as are well known in the art. Where the host is prokaryotic, such as *E. coli*, competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the $CaCl_2$ method using procedures well known in the art. Alternatively,

$MgCl_2$ or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell if desired, or by electroporation.

When the host is a eukaryote, such as a stem cell, such methods of transfection of DNA as calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors may be used. Eukaryotic cells can also be cotransformed with DNA sequences encoding Zscan4, and a second foreign DNA molecule encoding a selectable phenotype, such as neomycin resistance. Another method is to use a eukaryotic viral vector, such as simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the protein (see for example, *Eukaryotic Viral Vectors*, Cold Spring Harbor Laboratory, Gluzman ed., 1982). Other specific, non-limiting examples of viral vectors include adenoviral vectors, lentiviral vectors, retroviral vectors, and pseudorabies vectors.

Differentiation can be induced, or proliferation decreased, of any cell, either *in vivo* or *in vitro*, using the methods disclosed herein. In one embodiment, the cell is a stem cell, such as, but not limited to, an embryonic stem cell, a germline stem cell or a multipotent adult progenitor cell. In several examples, a Zscan4 polypeptide, or a polynucleotide encoding the Zscan4 polypeptide, is introduced into a stem cell to decrease differentiation and/or increase proliferation.

In one example, the cells are stem cells, such as embryonic stem cells. For example, murine, primate or human cells can be utilized. ES cells can proliferate indefinitely in an undifferentiated state. Furthermore, ES cells are totipotent cells, meaning that they can generate all of the cells present in the body (bone, muscle, brain cells, etc.). ES cells have been isolated from the inner cell mass (ICM) of the developing murine blastocyst (Evans *et al.*, *Nature* 292:154-156, 1981; Martin *et al.*, *Proc. Natl. Acad. Sci.* 78:7634-7636, 1981; Robertson *et al.*, *Nature* 323:445-448, 1986). Additionally, human cells with ES properties have been isolated from the inner blastocyst cell mass (Thomson *et al.*, *Science* 282:1145-1147, 1998) and developing germ cells (Shamblott *et al.*, *Proc. Natl. Acad. Sci. USA* 95:13726-13731, 1998), and human and non-human primate embryonic stem cells have been produced (see U.S. Patent No. 6,200,806, which is incorporated by reference herein).

As disclosed in U.S. Patent No. 6,200,806, ES cells can be produced from human and non-human primates. In one embodiment, primate ES cells are isolated "ES medium" that express SSEA-3; SSEA-4, TRA-1-60, and TRA-1-81 (see U.S. Patent No. 6,200,806). ES medium consists of 80% Dulbecco's modified Eagle's medium (DMEM; no pyruvate, high glucose formulation, Gibco BRL), with 20% fetal bovine serum (FBS; Hyclone), 0.1 mM ß-mercaptoethanol (Sigma), 1% non-essential amino acid stock (Gibco BRL). Generally, primate ES cells are isolated on a confluent layer of murine embryonic fibroblast in the presence of ES cell medium. In one example, embryonic fibroblasts are obtained from 12 day old fetuses from outbred mice (such as CF1, available from SASCO), but other strains may be used as an alternative. Tissue culture dishes treated with 0.1% gelatin (type I; Sigma) can be utilized. Distinguishing features of ES cells, as compared to the committed "multipotential" stem cells present in adults, include the capacity of ES cells to maintain an undifferentiated state indefinitely in culture, and the potential that ES cells have to develop into every different cell types. Unlike mouse ES cells, human ES (hES) cells do not express the stage-specific embryonic antigen SSEA-1, but express SSEA-4, which is another glycolipid cell surface antigen recognized by a specific monoclonal antibody (see, e.g., Amit *et al.*, *Devel. Biol.* 227:271-278, 2000).

For rhesus monkey embryos, adult female rhesus monkeys (greater than four years old) demonstrating normal ovarian cycles are observed daily for evidence of menstrual bleeding (day 1 of cycle=the day of onset of menses). Blood samples are drawn daily during the follicular phase starting from day 8 of the menstrual cycle, and serum concentrations of luteinizing hormone are determined by radioimmunoassay. The female is paired with a male rhesus monkey of proven fertility from day 9 of the menstrual cycle until 48 hours after the luteinizing hormone surge; ovulation is taken as the day following the leutinizing hormone surge. Expanded blastocysts are collected by non-surgical uterine flushing at six days after ovulation. This procedure generally results in the recovery of an average 0.4 to 0.6 viable embryos per rhesus monkey per month (Seshagiri *et al.*, *Am J Primatol.* 29:81-91, 1993).

For marmoset embryos, adult female marmosets (greater than two years of age) demonstrating regular ovarian cycles are maintained in family groups, with a fertile male and up to five progeny. Ovarian cycles are controlled by intramuscular injection of 0.75 g of the prostaglandin PGF2a analog cloprostenol (Estrumate, Mobay Corp, Shawnee, KS) during the middle to late luteal phase. Blood samples are drawn on day 0 (immediately before cloprostenol injection), and on days 3, 7, 9, 11, and 13. Plasma progesterone concentrations are determined by ELISA. The day of ovulation is taken as the day preceding a plasma progesterone concentration of 10 ng/ml or more. At eight days after ovulation, expanded blastocysts are recovered by a non-surgical uterine flush procedure (Thomson *et al.*, *J Med Primatol.* 23:333-336, 1994). This procedure results in the average production of 1.0 viable embryos per marmoset per month.

The zona pellucida is removed from blastocysts, such as by brief exposure to pronase (Sigma). For immunosurgery, blastocysts are exposed to a 1:50 dilution of rabbit anti-marmoset spleen cell antiserum (for marmoset blastocysts) or a 1:50 dilution of rabbit anti-rhesus monkey (for rhesus monkey blastocysts) in DMEM for 30 minutes, then washed for 5 minutes three times in DMEM, then exposed to a 1:5 dilution of Guinea pig complement (Gibco) for 3 minutes. After two further washes in DMEM, lysed trophoectoderm cells are removed from the intact inner cell mass (ICM) by gentle pipetting, and the ICM plated on mouse inactivated (3000 rads gamma irradiation) embryonic fibroblasts.

After 7-21 days, ICM-derived masses are removed from endoderm outgrowths with a micropipette with direct observation under a stereo microscope, exposed to 0.05% Trypsin-EDTA (Gibco) supplemented with 1% chicken serum for 3-5 minutes and gently dissociated by gentle pipetting through a flame polished micropipette.

Dissociated cells are re-plated on embryonic feeder layers in fresh ES medium, and observed for colony formation. Colonies demonstrating ES-like morphology are individually selected, and split again as described above. The ES-like morphology is defined as compact colonies having a high nucleus to cytoplasm ratio and prominent nucleoli. Resulting ES cells are then routinely split by brief trypsinization or exposure to Dulbecco's Phosphate Buffered Saline (PBS, without

calcium or magnesium and with 2 mM EDTA) every 1-2 weeks as the cultures become dense. Early passage cells are also frozen and stored in liquid nitrogen.

Cell lines may be karyotyped with a standard G-banding technique (such as by the Cytogenetics Laboratory of the University of Wisconsin State Hygiene Laboratory, which provides routine karyotyping services) and compared to published karyotypes for the primate species.

Isolation of ES cell lines from other primate species would follow a similar procedure, except that the rate of development to blastocyst can vary by a few days between species, and the rate of development of the cultured ICMs will vary between species. For example, six days after ovulation, rhesus monkey embryos are at the expanded blastocyst stage, whereas marmoset embryos do not reach the same stage until 7-8 days after ovulation. The rhesus ES cell lines can be obtained by splitting the ICM-derived cells for the first time at 7-16 days after immunosurgery; whereas the marmoset ES cells were derived with the initial split at 7-10 days after immunosurgery. Because other primates also vary in their developmental rate, the timing of embryo collection, and the timing of the initial ICM split, varies between primate species, but the same techniques and culture conditions will allow ES cell isolation (see U.S. Patent No. 6, 200,806, which is incorporated herein by reference for a complete discussion of primate ES cells and their production).

Human ES cell lines exist and can be used in the methods disclosed herein. Human ES cells can also be derived from preimplantation embryos from *in vitro* fertilized (IVF) embryos. Experiments on unused human IVF-produced embryos are allowed in many countries, such as Singapore and the United Kingdom, if the embryos are less than 14 days old. Only high quality embryos are suitable for ES isolation. Present defined culture conditions for culturing the one cell human embryo to the expanded blastocyst have been described (see Bongso *et al.*, *Hum Reprod.* 4:706-713, 1989). Co-culturing of human embryos with human oviductal cells results in the production of high blastocyst quality. IVF-derived expanded human blastocysts grown in cellular co-culture, or in improved defined medium, allows isolation of human ES cells with the same procedures described above for non-human primates (see U.S. Patent No. 6,200,806).

Precursor cells can also be utilized with the methods disclosed herein.

The precursor cells can be isolated from a variety of sources using methods known to one skilled in the art. The precursor cells can be of ectodermal, mesodermal or endodermal origin. Any precursor cells which can be obtained and maintained *in vitro* can potentially be used in accordance with the present methods. Such cells include cells of epithelial tissues such as the skin and the lining of the gut, embryonic heart muscle cells, and neural precursor cells (Stemple and Anderson, 1992, Cell 71:973-985).

In one example, the cells are mesenchymal progenitor cells. Mesenchymal progenitors give rise to a very large number of distinct tissues (Caplan, J. Orth. Res 641-650, 1991). Mesenchymal cells capable of differentiating into bone and cartilage have also been isolated from marrow (Caplan, J. Orth. Res. 641-650, 1991). U.S. Pat. No. 5,226,914 describes an exemplary method for isolating mesenchymal stem cells from bone marrow.

In other examples, the cells are epithelial progenitor cells or keratinocytes can be obtained from tissues such as the skin and the lining of the gut by known procedures (Rheinwald, Meth. Cell Bio. 21A:229, 1980). In stratified epithelial tissue such as the skin, renewal occurs by mitosis of precursor cells within the germinal layer, the layer closest to the basal lamina. Precursor cells within the lining of the gut provide for a rapid renewal rate of this tissue. The cells can also be liver stem cells (see PCT Publication No. WO 94/08598) or kidney stem cells (see Karp *et al.*, *Dev. Biol. 91*:5286-5290, 1994).

In one non-limited example, neuronal precursor cells are utilized. Undifferentiated neural stem cells differentiate into neuroblasts and glioblasts which give rise to neurons and glial cells. During development, cells that are derived from the neural tube give rise to neurons and glia of the CNS. Certain factors present during development, such as nerve growth factor (NGF), promote the growth of neural cells. Methods of isolating and culturing neural stem cells and progenitor cells are well known to those of skill in the art (Hazel and Muller, 1997; U.S. Pat. No. 5,750,376). Methods for isolating and culturing neuronal precursor cells are disclosed, for example, in U.S. Patent No. 6,610,540.

V. Zscan4 and Trim43 Promoter Sequences

A Zscan4 promoter or a Trim43 promoter can be included in an expression vector to direct expression of a heterologous nucleic acid sequence. Other expression control sequences including appropriate enhancers, transcription terminators, a start codon (i.e., ATG) in front of a protein-encoding gene, splicing signal for introns, maintenance of the correct reading frame of that gene to permit proper translation of mRNA, and stop codons can be included with the Zscan4 or Trim43 promoter in an expression vector. Generally the promoter includes at least a minimal sequence sufficient to direct transcription of a heterologous nucleic acid sequence. In several examples, the heterologous nucleic acid sequence encodes a polypeptide. However, the heterologous nucleic acid can be any RNA sequence of interest, such as an inhibitory RNA.

Expression vectors typically contain an origin of replication as well as specific genes which allow phenotypic selection of the transformed cells. Vectors suitable for use include, but are not limited to the pMSXND expression vector for expression in mammalian cells (Lee and Nathans, *J. Biol. Chem.* 263:3521, 1988). In one example, an enhancer is located upstream of the Zscan4 or Trim43 promoter, but enhancer elements can generally be located anywhere on the vector and still have an enhancing effect. However, the amount of increased activity will generally diminish with distance. Additionally, two or more copies of an enhancer sequence can be operably linked one after the other to produce an even greater increase in promoter activity.

Generally, an expression vector includes a nucleic acid sequence encoding a polypeptide of interest. A polypeptide of interest can be a heterologous polypeptide, such as a polypeptide that affects a function of the transfected cell. Polypeptides of interest include, but are not limited to, polypeptides that confer antibiotic resistance, receptors, oncogenes, and neurotransmitters. A polypeptide of interest can also be a marker polypeptide, which is used to identify a cell of interest. Marker polypeptides include fluorescent polypeptides, enzymes, or antigens that can be identified using conventional molecular biology procedures. For example, the polypeptide can be a fluorescent marker (such as green fluorescent protein, Emerald (Invitrogen, Carlsbad, CA), Strawberry (Clontech, Mountain View, CA), *Aequoria victoria*, or

Discosoma DSRed); an antigenic marker (such as human growth hormone, human insulin, human HLA antigens); a cell-surface marker (such as CD4, or any cell surface receptor); or an enzymatic marker (such as lacZ, alkaline phosphatase). Techniques for identifying these markers in host cells include immunohistochemistry and fluorescent microscopy, and are well known in the art.

RNA molecules transcribed from an expression vector need not always be translated into a polypeptide to express a functional activity. Specific non-limiting examples of other molecules of interest include antisense RNA molecules complementary to an RNA of interest, ribozymes, small inhibitory RNAs, and naturally occurring or modified tRNAs.

Expression vectors including a Zscan4 or Trim43 promoter can be used to transform host cells. Hosts can include isolated microbial, yeast, insect and mammalian cells, as well as cells located in the organism. Biologically functional viral and plasmid DNA vectors capable of expression and replication in a host are known in the art, and can be used to transfect any cell of interest. Where the cell is a mammalian cell, the genetic change is generally achieved by introduction of the DNA into the genome of the cell (stable integration). However, the vector can also be maintained as an episome.

A "transfected cell" is a host cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule including a Zscan4 promoter element. Transfection of a host cell with recombinant DNA may be carried out by conventional techniques as are well known to those skilled in the art. Where the host is prokaryotic, such as *E. coli*, competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the $CaCl_2$ method using procedures well known in the art. Alternatively, $MgCl_2$ or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell if desired, or by electroporation.

When the host is a eukaryote, such methods of transfection of DNA as calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors may be used. Eukaryotic cells can also be cotransformed with DNA

sequences including the Zscan4 promoter, and a second foreign DNA molecule encoding a selectable phenotype, such as neomycin resistance. Another method is to use a eukaryotic viral vector, such as simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the protein (see for example, *Eukaryotic Viral Vectors*, Cold Spring Harbor Laboratory, Gluzman ed., 1982). Other specific, non-limiting examples of viral vectors include adenoviral vectors, lentiviral vectors, retroviral vectors, and pseudorabies vectors.

In one embodiment described in the Examples below, an expression vector comprising a Zsan4 promoter sequence operably linked to a heterologous polypeptide is used to identify cells that express Zscan4. In one embodiment, the Zscan4 promoter is a Zscan4c promoter. In some embodiments, the Zscan4c promoter comprises Zsan4c exon and/or intron sequence. The heterologous protein is typically a marker, an enzyme, or a fluorescent protein. In one embodiment, the heterologous protein is green fluorescent protein (GFP), or a variant of GFP, such as Emerald.

Provided herein is a method of identifying a subpopulation of stem cells expressing Zscan4. In one embodiment, the subpopulation is identified by transfecting the stem cells with an expression vector, wherein the expression vector comprises a Zscan4 promoter sequence and a reporter gene. In one embodiment, the Zscan4 promoter is a Zscan4c promoter. In another embodiment, the Zscan4c promoter comprises the nucleic acid sequence set forth as nucleotides 1-2540 of SEQ ID NO: 28, such as nucleotides 1-2643, 1-3250, or 1-3347 of SEQ ID NO: 28.

The reporter gene can be any type of identifiable marker, such as an enzyme or a fluorescent protein. In one embodiment, the reporter gene is GFP or a variant of GFP, such as Emerald. Expression of the reporter gene indicates the cell expresses Zscan4. Methods of detecting expression of the reporter gene vary depending upon the type of reporter gene and are well known in the art. For example, when a fluorescent reporter is used, detection of expression can be achieved by fluorescence activated cell sorting or fluorescence microscopy. Identification of a subpopulation of stem cells expressing Zscan4 can be achieved with alternative methods, including, but not limited to, using antibodies specific for Zscan4 or by *in situ* hybridization. In one embodiment, the subpopulation of ES cells expressing Zscan4 is identified by

detecting expression of one or more Zscan4 co-expressed genes, including AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1.

Also described herein is an expression vector comprising a Trim43 promoter sequence operably linked to a heterologous polypeptide. The heterologous protein is typically a marker, an enzyme, or a fluorescent protein. In one embodiment, the heterologous protein is the fluorescent protein Strawberry. In some embodiments, the Trim43 promoter sequence is at least 70%, at least 80%, at least 90%, at least 95% or at least 99% identical to SEQ ID NO: 31. In another embodiment, the Trim43 promoter comprises SEQ ID NO: 31. In another embodiment, the Trim43 promoter consists of SEQ ID NO: 31.

Also provided herein are isolated ES cells comprising the Zscan4 or Trim43 expression vectors described herein. In one embodiment, the ES cells are a stable cell line.

VI. Transgenic animals

The Zscan4 polynucleotide sequences disclosed herein can also be used in the production of transgenic animals such as transgenic mice, as described below. Transgenic animals can also be produced that contain polynucleotide sequences of one or more Zscan4 co-expressed genes, including AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1.

In one embodiment, a non-human animal is generated that carries a transgene comprising a nucleic acid encoding Zscan4 operably linked to a promoter. Specific promoters of use include, but are not limited to, a tissue specific promoter such as, but not limited to, an immunoglobulin promoter, a neuronal specific promoter, or the insulin promoter. Specific promoters of use also include a constitutive promoter, such as, but not limited to, the thymidine kinase promoter or the human β-globin minimal, or an actin promoter, amongst others. The Zscan4 promoter can also be used.

In another embodiment, the transgenic non-human animal carries a transgene comprising a nucleic acid encoding a heterologous peptide, such as a marker, enzyme or fluorescent protein, operably linked to a Zscan4 promoter. In one

example, the Zscan4 promoter is a Zscan4c promoter, or a portion thereof. In another embodiment, the Zscan4c promoter comprises the nucleic acid sequence set forth as nucleotides 1-2540 of SEQ ID NO: 28, such as nucleotides 1-2643, 1-3250, or 1-3347 of SEQ ID NO: 28. In one example, the heterologous peptide is the fluorescent protein Emerald.

In another embodiment, the transgenic non-human animal carries a transgene comprising a nucleic acid encoding a heterologous peptide, such as a marker, enzyme or fluorescent protein, operably linked to a Trim43 promoter. In one example, the Trim43 promoter comprises the nucleotide sequence of SEQ ID NO: 31, or a portion thereof. The portion of SEQ ID NO: 31 to be included in the expression vector is at least a portion of SEQ ID NO: 31 that is capable of promoting transcription of the heterologous polypeptide in a cell in which Trim43 is expressed. In some embodiments, the Trim43 promoter sequence is at least 70%, at least 80%, at least 90%, at least 95% or at least 99% identical to SEQ ID NO: 31. In another embodiment, the Trim43 promoter comprises SEQ ID NO: 31. In another embodiment, the Trim43 promoter consists of SEQ ID NO: 31. In one example, the heterologous peptide is the fluorescent protein Strawberry.

In another embodiment, the transgenic non-human animal carries two transgenes, a transgene comprising the Zscan4 promoter linked to a nucleic acid sequence encoding a heterologous peptide, and a transgene comprising the Trim43 promoter linked to a nucleic acid sequence encoding a heterologous peptide, as described above. In some cases, the transgenic non-human animal is a mouse comprising the Zscan4 promoter transgene and the Trim43 promoter transgene. In one specific example, the heterologous polypeptide operably linked to the Zscan4 promoter sequence is the fluorescent protein Emerald and the heterologous polypeptide operably linked to the Trim43 promoter sequence is the fluorescent protein Strawberry. This mouse is referred to as a "rainbow" mouse (see Example 10 below).

Embryos obtained from transgenic "rainbow" animals exhibit green color at the late 2-cell stage and red color at the 4-cell to morula stages (with strongest expression at the morula stage). The expression of these colors at the proper timing and intensity indicates the progress of a correct genetic program, and thus, can be

used as indicators of proper development of preimplantation embryos. These embryos have a variety of applications, including, but not limited to development of optimized culture media for human embryos for *in vitro* fertilization (IVF); training of technicians and clinicians in the IVF clinic and research laboratories; testing of chemical compounds and drugs for embryo toxicity; and as indicators of successful nuclear reprogramming for nuclear transplantation/cloning procedures.

The nucleic acid sequences described herein can be introduced into a vector to produce a product that is then amplified, for example, by preparation in a bacterial vector, according to conventional methods (see, for example, Sambrook *et al.*, *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Press, 1989). The amplified construct is thereafter excised from the vector and purified for use in producing transgenic animals.

Any transgenic animal can be of use in the methods disclosed herein, provided the transgenic animal is a non-human animal. 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, or a zoo animal such as lions, tigers or bears. In one specific, non-limiting example, the non-human animal is a transgenic animal, such as, but not limited to, a transgenic mouse, cow, sheep, or goat. In one specific, non-limiting example, the transgenic animal is a mouse. In a particular example, the transgenic animal has altered proliferation and/or differentiation of a cell type as compared to a non-transgenic control (wild-type) animal of the same species.

A transgenic animal contains cells that bear genetic information received, directly or indirectly, by deliberate genetic manipulation at the subcellular level, such as by microinjection or infection with a recombinant virus, such that a recombinant DNA is included in the cells of the animal. This molecule can be integrated within the animal's chromosomes, or can be included as extrachromosomally replicating DNA sequences, such as might be engineered into yeast artificial chromosomes. A transgenic animal can be a "germ cell line" transgenic animal, such that the genetic information has been taken up and incorporated into a germ line cell, therefore conferring the ability to transfer the

information to offspring. If such offspring in fact possess some or all of that information, then they, too, are transgenic animals.

Transgenic animals can readily be produced by one of skill in the art. For example, transgenic animals can be produced by introducing into single cell embryos DNA encoding a marker, in a manner such that the polynucleotides are stably integrated into the DNA of germ line cells of the mature animal and inherited in normal Mendelian fashion. Advances in technologies for embryo micromanipulation permit introduction of heterologous DNA into fertilized mammalian ova. For instance, totipotent or pluripotent stem cells can be transformed by microinjection, calcium phosphate mediated precipitation, liposome fusion, retroviral infection or other means. The transformed cells are then introduced into the embryo, and the embryo then develops into a transgenic animal. In one non-limiting method, developing embryos are infected with a retrovirus containing the desired DNA, and a transgenic animal is produced from the infected embryo.

In another specific, non-limiting example, the appropriate DNA(s) are injected into the pronucleus or cytoplasm of embryos, preferably at the single cell stage, and the embryos are allowed to develop into mature transgenic animals. These techniques are well known. For instance, reviews of standard laboratory procedures for microinjection of heterologous DNAs into mammalian (mouse, pig, rabbit, sheep, goat, cow) fertilized ova include: Hogan *et al.*, *Manipulating the Mouse Embryo*, Cold Spring Harbor Press, 1986; Krimpenfort *et al.*, *Bio/Technology* 9:86, 1991; Palmiter *et al.*, *Cell* 41:343, 1985; Kraemer *et al.*, *Genetic Manipulation of the Early Mammalian Embryo*, Cold Spring Harbor Laboratory Press, 1985; Hammer *et al.*, *Nature* 315:680, 1985; Purcel *et al.*, *Science* 244:1281, 1986; U.S. Patent No. 5,175,385; U.S. Patent No. 5,175,384.

VII. Antibodies

A Zscan4 polypeptide or a fragment or conservative variant thereof can be used to produce antibodies which are immunoreactive or specifically bind to an epitope of a Zscan4. Polyclonal antibodies, antibodies which consist essentially of pooled monoclonal antibodies with different epitopic specificities, as well as distinct

monoclonal antibody preparations are included. In one embodiment, the Zscan4 antibodies recognize all Zscan4 proteins, including Zscan4a, Zscan4b, Zscan4c, Zscan4d, Zscan4e, Zscan4f and human ZSCAN4. In another embodiment, the antibodies specifically recognize only one Zscan4 protein. As used herein, the ability of an antibody to specifically a particular Zscan4 protein means that the antibody detects expression of one Zscan4 protein, but none of the other Zscan4 proteins. In an alternative embodiment, the antibodies recognize two or more different Zscan4 proteins. For example, a Zscan4 antibody may recognize only the Zscan4 proteins comprising a SCAN domain (*e.g.*, Zscan4c, Zscan4d, Zscan4f). Or, a Zscan4 antibody may recognize only the Zscan4 proteins comprising the zinc finger domains, but lacking the SCAN domain (*e.g.*, Zscan4a, Zscan4b, Zscan4e).

Antibodies can also be raised against one or more proteins encoded by genes identified herein as Zscan4 co-expressed genes. Thus, in some embodiments, a polypeptide encoded by AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 or Pif1, or a fragment or conservative variant thereof, can be used to produce antibodies which are immunoreactive or specifically bind to an epitope of the polypeptide.

In addition, antibodies can be generated that specifically bind Trim43. In one embodiment, a Trim43 polypeptide, or a fragment or conservative variant thereof, can be used to produce antibodies which are immunoreactive or specifically bind to an epitope of Trim43.

The preparation of polyclonal antibodies is well known to those skilled in the art. See, for example, Green *et al.*, "Production of Polyclonal Antisera," in: *Immunochemical Protocols*, pages 1-5, Manson, ed., Humana Press, 1992; Coligan *et al.*, "Production of Polyclonal Antisera in Rabbits, Rats, Mice and Hamsters," in: *Current Protocols in Immunology*, section 2.4.1, 1992.

The preparation of monoclonal antibodies likewise is conventional. See, for example, Kohler & Milstein, *Nature* 256:495, 1975; Coligan *et al.*, sections 2.5.1-2.6.7; and Harlow *et al.* in: *Antibodies: a Laboratory Manual*, page 726, Cold Spring Harbor Pub., 1988. Briefly, monoclonal antibodies can be obtained by injecting mice with a composition comprising an antigen, verifying the presence of antibody production by removing a serum sample, removing the spleen to obtain B

lymphocytes, fusing the B lymphocytes with myeloma cells to produce hybridomas, cloning the hybridomas, selecting positive clones that produce antibodies to the antigen, and isolating the antibodies from the hybridoma cultures. Monoclonal antibodies can be isolated and purified from hybridoma cultures by a variety of well-established techniques. Such isolation techniques include affinity chromatography with Protein-A Sepharose, size-exclusion chromatography, and ion-exchange chromatography. See, e.g., Coligan *et al.*, sections 2.7.1-2.7.12 and sections 2.9.1-2.9.3; Barnes *et al.*, Purification of Immunoglobulin G (IgG), in: *Methods in Molecular Biology*, Vol. 10, pages 79-104, Humana Press, 1992.

Methods of *in vitro* and *in vivo* multiplication of monoclonal antibodies are well known to those skilled in the art. Multiplication *in vitro* may be carried out in suitable culture media such as Dulbecco's Modified Eagle Medium or RPMI 1640 medium, optionally supplemented by a mammalian serum such as fetal calf serum or trace elements and growth-sustaining supplements such as normal mouse peritoneal exudate cells, spleen cells, thymocytes or bone marrow macrophages. Production *in vitro* provides relatively pure antibody preparations and allows scale-up to yield large amounts of the desired antibodies. Large-scale hybridoma cultivation can be carried out by homogenous suspension culture in an airlift reactor, in a continuous stirrer reactor, or in immobilized or entrapped cell culture. Multiplication *in vivo* may be carried out by injecting cell clones into mammals histocompatible with the parent cells, such as syngeneic mice, to cause growth of antibody-producing tumors. Optionally, the animals are primed with a hydrocarbon, especially oils such as pristane (tetramethylpentadecane) prior to injection. After one to three weeks, the desired monoclonal antibody is recovered from the body fluid of the animal.

Antibodies can also be derived from a subhuman primate antibody. General techniques for raising therapeutically useful antibodies in baboons can be found, for example, in PCT Publication No. WO 91/11465, 1991; and Losman *et al.*, *Int. J. Cancer* 46:310, 1990.

Alternatively, an antibody that specifically binds a Zscan4 polypeptide can be derived from a humanized monoclonal antibody. Humanized monoclonal antibodies are produced by transferring mouse complementarity determining regions from heavy and light variable chains of the mouse immunoglobulin into a human

variable domain, and then substituting human residues in the framework regions of the murine counterparts. The use of antibody components derived from humanized monoclonal antibodies obviates potential problems associated with the immunogenicity of murine constant regions. General techniques for cloning murine immunoglobulin variable domains are described, for example, by Orlandi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 86:3833, 1989. Techniques for producing humanized monoclonal antibodies are described, for example, by Jones *et al.*, *Nature* 321:522, 1986; Riechmann *et al.*, *Nature* 332:323, 1988; Verhoeyen *et al.*, *Science* 239:1534, 1988; Carter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89:4285, 1992; Sandhu, *Crit. Rev. Biotech.* 12:437, 1992; and Singer *et al.*, *J. Immunol.* 150:2844, 1993.

Antibodies can be derived from human antibody fragments isolated from a combinatorial immunoglobulin library. See, for example, Barbas *et al.*, in: *Methods: a Companion to Methods in Enzymology*, Vol. 2, page 119, 1991; Winter *et al.*, *Ann. Rev. Immunol.* 12:433, 1994. Cloning and expression vectors that are useful for producing a human immunoglobulin phage library can be obtained, for example, from STRATAGENE Cloning Systems (La Jolla, CA).

In addition, antibodies can be derived from a human monoclonal antibody. Such antibodies are obtained from transgenic mice that have been "engineered" to produce specific human antibodies in response to antigenic challenge. In this technique, elements of the human heavy and light chain loci are introduced into strains of mice derived from embryonic stem cell lines that contain targeted disruptions of the endogenous heavy and light chain loci. The transgenic mice can synthesize human antibodies specific for human antigens, and the mice can be used to produce human antibody-secreting hybridomas. Methods for obtaining human antibodies from transgenic mice are described by Green *et al.*, *Nature Genet.* 7:13, 1994; Lonberg *et al.*, *Nature* 368:856, 1994; and Taylor *et al.*, *Int. Immunol.* 6:579, 1994.

Antibodies include intact molecules as well as fragments thereof, such as Fab, $F(ab')_2$, and Fv which are capable of binding the epitopic determinant. These antibody fragments retain some ability to selectively bind with their antigen or receptor and are defined as follows:

(1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule, can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain;

(2) Fab', the fragment of an antibody molecule can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule;

(3) $(Fab')_2$, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; $F(ab')_2$ is a dimer of two Fab' fragments held together by two disulfide bonds;

(4) Fv, defined as a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and

(5) Single chain antibody (SCA), defined as a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule.

Methods of making these fragments are known in the art (see for example, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York, 1988). An epitope is any antigenic determinant on an antigen to which the paratope of an antibody binds. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics.

Antibody fragments can be prepared by proteolytic hydrolysis of the antibody or by expression in *E. coli* of DNA encoding the fragment. Antibody fragments can be obtained by pepsin or papain digestion of whole antibodies by conventional methods. For example, antibody fragments can be produced by enzymatic cleavage of antibodies with pepsin to provide a 5S fragment denoted $F(ab')_2$. This fragment can be further cleaved using a thiol reducing agent, and optionally a blocking group for the sulfhydryl groups resulting from cleavage of

disulfide linkages, to produce 3.5S Fab' monovalent fragments. Alternatively, an enzymatic cleavage using pepsin produces two monovalent Fab' fragments and an Fc fragment directly (see U.S. Patents No. 4,036,945 and U.S. Patent No. 4,331,647, and references contained therein; Nisonhoff *et al.*, *Arch. Biochem. Biophys.* 89:230, 1960; Porter, *Biochem. J.* 73:119, 1959; Edelman *et al.*, *Methods in Enzymology*, Vol. 1, page 422, Academic Press, 1967; and Coligan *et al.* at sections 2.8.1-2.8.10 and 2.10.1-2.10.4).

Other methods of cleaving antibodies, such as separation of heavy chains to form monovalent light-heavy chain fragments, further cleavage of fragments, or other enzymatic, chemical, or genetic techniques may also be used, so long as the fragments bind to the antigen that is recognized by the intact antibody.

For example, Fv fragments comprise an association of V_H and V_L chains. This association may be noncovalent (Inbar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 69:2659, 1972). Alternatively, the variable chains can be linked by an intermolecular disulfide bond or cross-linked by chemicals such as glutaraldehyde. See, e.g., Sandhu, *supra*. Preferably, the Fv fragments comprise V_H and V_L chains connected by a peptide linker. These single-chain antigen binding proteins (sFv) are prepared by constructing a structural gene comprising DNA sequences encoding the V_H and V_L domains connected by an oligonucleotide. The structural gene is inserted into an expression vector, which is subsequently introduced into a host cell such as *E. coli*. The recombinant host cells synthesize a single polypeptide chain with a linker peptide bridging the two V domains. Methods for producing sFvs are known in the art (see Whitlow *et al.*, *Methods: a Companion to Methods in Enzymology*, Vol. 2, page 97, 1991; Bird *et al.*, *Science* 242:423, 1988; U.S. Patent No. 4,946,778; Pack *et al.*, *Bio/Technology* 11:1271, 1993; and Sandhu, *supra)*.

Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. Such genes are prepared, for example, by using the polymerase chain reaction to synthesize the variable region from RNA of antibody-producing cells (Larrick *et al.*, *Methods: a Companion to Methods in Enzymology*, Vol. 2, page 106, 1991).

Antibodies can be prepared using an intact polypeptide or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or a peptide used to immunize an animal can be derived from substantially purified polypeptide produced in host cells, *in vitro* translated cDNA, or chemical synthesis which can be conjugated to a carrier protein, if desired. Such commonly used carriers which are chemically coupled to the peptide include keyhole limpet hemocyanin (KLH), thyroglobulin, bovine serum albumin (BSA), and tetanus toxoid. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

Polyclonal or monoclonal antibodies can be further purified, for example, by binding to and elution from a matrix to which the polypeptide or a peptide to which the antibodies were raised is bound. Those of skill in the art will know of various techniques common in the immunology arts for purification and/or concentration of polyclonal antibodies, as well as monoclonal antibodies (see, for example, Coligan *et al.*, Unit 9, *Current Protocols in Immunology*, Wiley Interscience, 1991).

It is also possible to use the anti-idiotype technology to produce monoclonal antibodies which mimic an epitope. For example, an anti-idiotypic monoclonal antibody made to a first monoclonal antibody will have a binding domain in the hypervariable region that is the "image" of the epitope bound by the first monoclonal antibody.

Binding affinity for a target antigen is typically measured or determined by standard antibody-antigen assays, such as competitive assays, saturation assays, or immunoassays such as ELISA or RIA. Such assays can be used to determine the dissociation constant of the antibody. The phrase "dissociation constant" refers to the affinity of an antibody for an antigen. Specificity of binding between an antibody and an antigen exists if the dissociation constant ($K_D = 1/K$, where K is the affinity constant) of the antibody is, for example $< 1\mu M$, < 100 nM, or < 0.1 nM. Antibody molecules will typically have a K_D in the lower ranges. $K_D = [Ab\text{-}Ag]/[Ab][Ag]$ where [Ab] is the concentration at equilibrium of the antibody, [Ag] is the concentration at equilibrium of the antigen and [Ab-Ag] is the concentration at equilibrium of the antibody-antigen complex. Typically, the binding interactions

between antigen and antibody include reversible noncovalent associations such as electrostatic attraction, Van der Waals forces and hydrogen bonds.

Effector molecules, e.g., therapeutic, diagnostic, or detection moieties can be linked to an antibody that specifically binds Zscan4, using any number of means known to those of skill in the art. Exemplary effector molecules include, but not limited to, radiolabels, fluorescent markers, or toxins (e.g. *Pseudomonas* exotoxin (PE), see "*Monoclonal Antibody-Toxin Conjugates: Aiming the Magic Bullet*," Thorpe *et al.*, "Monoclonal Antibodies in Clinical Medicine," Academic Press, pp. 168-190, 1982; Waldmann, *Science*, 252: 1657, 1991; U.S. Patent No. 4,545,985 and U.S. Patent No. 4,894,443, for a discussion of toxins and conjugation). Both covalent and noncovalent attachment means may be used. The procedure for attaching an effector molecule to an antibody varies according to the chemical structure of the effector. Polypeptides typically contain a variety of functional groups; e.g., carboxylic acid (COOH), free amine ($-NH_2$) or sulfhydryl (-SH) groups, which are available for reaction with a suitable functional group on an antibody to result in the binding of the effector molecule. Alternatively, the antibody is derivatized to expose or attach additional reactive functional groups. The derivatization may involve attachment of any of a number of linker molecules such as those available from Pierce Chemical Company, Rockford, IL. The linker can be any molecule used to join the antibody to the effector molecule. The linker is capable of forming covalent bonds to both the antibody and to the effector molecule. Suitable linkers are well known to those of skill in the art and include, but are not limited to, straight or branched-chain carbon linkers, heterocyclic carbon linkers, or peptide linkers. Where the antibody and the effector molecule are polypeptides, the linkers may be joined to the constituent amino acids through their side groups (e.g., through a disulfide linkage to cysteine) or to the alpha carbon amino and carboxyl groups of the terminal amino acids.

In some circumstances, it is desirable to free the effector molecule from the antibody when the immunoconjugate has reached its target site. Therefore, in these circumstances, immunoconjugates will comprise linkages that are cleavable in the vicinity of the target site. Cleavage of the linker to release the effector molecule from the antibody may be prompted by enzymatic activity or conditions to which the

immunoconjugate is subjected either inside the target cell or in the vicinity of the target site.

In view of the large number of methods that have been reported for attaching a variety of radiodiagnostic compounds, radiotherapeutic compounds, label (e.g. enzymes or fluorescent molecules) drugs, toxins, and other agents to antibodies, one skilled in the art will be able to determine a suitable method for attaching a given agent to an antibody or other polypeptide.

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

EXAMPLES

The characterization of Zscan4 is disclosed herein. Zscan4 is shown herein to exhibit transient and specific expression at the late 2-cell embryonic stage and in embryonic stem cells. Without being bound by theory, Zscan4 is the only gene that is exclusively expressed during the first wave of *de novo* transcription, zygotic genome activation.

Zscan4 was identified from a cDNA clone derived from ES cells (clone number C0348C03) and subsequently sequenced by the Mammalian Gene Collection project (Gerhard *et al. Genom Res. 14*:2121-2127, 2004). The cDNA sequence, deposited under Genbank Accession No. BC050218 (SEQ ID NO: 11), comprised 2292 bp organized into 4 exons encoding a protein of 506 amino acids. As described in the Examples below, using this cDNA clone as a probe, a high level of Zscan4 transcript was detected in late 2-cell stage embryos. Since the original cDNA was isolated from ES cells, RT-PCR was performed on RNAs derived from late 2-cell stage embryos and the amplification product was sequenced, as described in the Examples below. The amplified sequence was 2268 bp in length and like the cDNA isolated from ES cells, encoded a protein of 506 amino acids. Analysis of the nucleotide and amino acid sequences of the cDNA clones isolated from ES cells and late 2-cell embryos showed they were two different, but similar genes.

As described in the Examples below, nine Zscan4 gene copies were identified in the mouse genome. Three copies are pseudogenes and were designated *Zscan4-ps1* (SEQ ID NO: 12), *Zscan4-ps2* (SEQ ID NO: 13) and *Zscan4-ps3* (SEQ ID NO: 14), according to the convention of mouse gene nomenclature. The remaining six gene copies are transcribed and encode ORFs, thus they were named *Zscan4a* (SEQ ID NOs: 15 and 16), *Zscan4b* (SEQ ID NOs: 17 and 18), *Zscan4c* (SEQ ID NOs: 19 and 20), *Zscan4d* (SEQ ID NOs: 21 and 22), *Zscan4e* (SEQ ID NOs: 23 and 24) and *Zscan4f* (SEQ ID NOs: 25 and 26). Zscan4c, Zscan4d and Zscan4f encode proteins of 506 amino acids, while Zscan4a, Zscan4b and Zscan4e encode shorter proteins of 360, 195 and 195 amino acids, respectively. A polypeptide comprising any of the amino acid sequences set forth as SEQ ID NOs: 16, 18, 20, 22, 24, 26 or 30, or a polynucleotide encoding these polypeptides, are of use in the methods disclosed herein. A polynucleotide encoding a Zscan4 pseudogene set forth as SEQ ID NOs: 12, 13 or 14 are also of use in the methods disclosed herein.

Analysis of the expression levels of Zscan4 demonstrated that expression of each of the six Zscan4 genes could be detected in ES cells with Zscan4c being the predominant transcript. Zscan4d was the predominant transcript in 2-cell stage embryos; however, low levels of Zscan4a Zscan4e and Zscan4f could also be detected. These findings are consistent with the origin of each cDNA clone since Zscan4c was derived from the ES cell cDNA library and Zscan4d was derived from the 2-cell embryo cDNA library. Furthermore, expression of Zscan4 was not detected in blastocysts (including the inner cell mass) or early blastocyst outgrowth. After approximately six days of outgrowth, Zscan4 expression was detected in a subpopulation of undifferentiated ES cells.

It is shown herein that expression of Zscan4 is temporally regulated and its expression or lack of expression at different embryonic stages is critical to proper development. As described in the Examples below, inhibition of Zscan4 expression in embryos blocked the 2- to 4-cell embryonic transition, prevented blastocysts from expanding, prevented blastocysts from implanting and prevented proliferation of ES cells from blastocyst outgrowths.

Also described herein is the development of a mouse ES cell line expressing a heterologous protein, Emerald, under the control of a Zscan4 promoter. Further described is the identification of nine Zscan4 co-expressed genes exhibiting 2-cell stage specific expression.

Also shown herein is the identification of Trim43 as a gene exhibiting expression during the 4-cell to morula embryonic stages, with the highest level of expression observed at the morula stage. Also described herein is the development of a transgenic mouse, which comprises two transgenes, the first comprising Emerald operably linked to the Zscan4c promoter and the second comprising Strawberry operably linked to the Trim43 promoter.

Example 1: Materials and Methods

Identification and cloning of the mouse Zscan4d gene

Using DNA microarray data of mouse preimplantation embryos (Hamatani *et al.*, *Dev. Cell 6*:117-131, 2004), *Zscan4d* gene was identified for its specific expression in 2-cell embryos. A corresponding cDNA clone (no. C0348C03; R1 ES cells, 129 strain ; Genbank Accession No. BC050218, SEQ ID NO: 11) was identified in the mouse cDNA collection described previously (Sharov *et al.*, *PLoS Bio. 1*:E74, 2003). Based on this full-length cDNA sequence, a primer pair (5'-cctccctgggcttcttggcat-3', SEQ ID NO: 1; 5'-agctgccaaccagaaagacactgt-3', SEQ ID NO: 2) was designed and used to PCR-amplify the full-length cDNA sequence of this gene from 2-cell embryos (B6D2F1 mouse). In brief, mRNA was extracted from 2-cell embryos and treated with DNAase (DNA-free, Ambion). The mRNA was annealed with an oligo-dT primer and reverse-transcribed into cDNA with ThermoScript Reverse Transcriptase (Invitrogen). A full-length cDNA clone was PCR-amplified with Ex Taq Polymerase (Takara Mirus Bio, Madison, WI), purified with the Wizard SV Gel and PCR Clean-Up System (Promega Biosciences, San Luis Obispo, CA), cloned into a pENTR plasmid vector with the Directional TOPO Cloning Kit (Invitrogen), and completely sequenced using BigDye Terminator kit (PE Applied Biosystems, Foster City, CA) and DyeEX 96 Kit (Qiagen Valencia,

CA) on ABI 3100 Genetic Analyzer (PE Applied Biosystems). The sequence is set forth herein as SEQ ID NO: 21).

The WU-BLAST (available online) and UCSC genome browser were used to obtain Zscan4 orthologs in the human genome sequence. Open reading frames (ORFs) were deduced by ORF finder (available online from the National Center for Biotechnology Information) and protein domains were identified by Pfam HMM database (available online). Orthologous relationships were assessed with the phylogenetic tree of amino acid sequences determined by a sequence distance method and the Neighbor Joining (NJ) algorithm (Saitou and Nei, 1987) using Vector NTI software (Invitrogen, Carlsbad, CA).

All gene names and gene symbols were consulted with and approved by the mouse gene nomenclature committee.

Southern blot analysis

Southern blot analysis was carried out to validate the genome sequence of the Zscan4 locus assembled using individual BAC clone sequences downloaded from the public database (RPCI-23 and RPCI-24 BAC libraries: C57BL/6J strain). A probe containing exon 3 was designed and amplified from mouse DNA extracted from ES cells (C57BL/6) using a primer pair (5'-gcattcctacataccaatta-3', SEQ ID NO: 3; 5'-gatttaatttagctgggctg-3', SEQ ID NO: 4). The PCR product was purified using GFX PCR DNA and Gel band purification kit (GE Healthcare). Fifteen μg of mouse genomic DNA extracted from ES cells (BL6.9 line derived from C57BL/6 strain) was digested overnight with restriction enzymes (*MspI*, *TaqI*, and *MspI*/*TaqI*, see Fig. 3B), fractionated on a 1% (w/v) agarose gel, transferred and immobilized onto nitrocellulose membranes. Blots were hybridized with random-primed ^{32}P-labeled DNA probes under standard conditions. Membranes were subjected to 3 washes of 30 min each (2xSSC/0.1% (w/v) SDS at room temperature, 0.5xSSC/0.1% (w/v) SDS at 42°C, and 0.1xSSC/0.1% (w/v) SDS at room temperature) and autoradiographed for 48 hours at -80°C.

Measurement of gene expression levels

cDNAs from ES cells (129.3 ES cells purchased from the Transgenic Core Laboratory of the Johns Hopkins University School of Medicine, Baltimore, MD)

and 2-cell embryos (B6D2F1 mice) were synthesized. Zscan4 cDNA fragments were amplified using a Zscan4-specific primer pair (Zscan4_For:5'-cagatgccagtagacaccac-3', SEQ ID NO: 5; Zscan4_Rev 5'-gtagatgttccttgacttgc-3', SEQ ID NO: 6), which were100%-matched to all Zscan4 paralogs. These cDNA fragments were sequenced using the following primers: Zscan4_For, 5'-cagatgccagtagacaccac-3', SEQ ID NO: 5; Zscan4_400Rev, 5'-ggaagtgttatagcaattgttc-3', SEQ ID NO: 7; Zscan4_Rev, 5'-gtagatgttccttgacttgc-3', SEQ ID NO: 6; and Zscan4_300Rev, 5'-gtgttatagcaattgttcttg-3', SEQ ID NO: 8. Electropherograms of these sequences were used to calculate the relative expression levels of nine paralogous copies of Zscan4 in the following manner. Based on sequence information of transcripts (either predicted from the genome sequence or determined by sequencing cDNA clones), nucleotide positions were identified where one or a few paralogous copies can be distinguished based on the nucleotide mismatches. The phred base calling program (version 0.020425.c (Ewing *et al.*, *Genome Res. 8*:175-185, 1998)) was used to obtain the amplitudes of all four bases in the electropherogram for those nucleotide sites. After subtracting the noise level (i.e., the average of amplitudes of the bases that are not present in any of the nine paralogous copies), the amplitudes of each base (A, T, G, C) were obtained. The expression levels of each of the paralogous copies were calculated by the least square fitting, which found the expression levels that are most consistent with all mismatched nucleotide positions.

Collection and manipulation of embryos

Four- to six-week old B6D2F1 mice were superovulated by injecting 5 IU pregnant mare serum gonadotropin (PMS; Sigma, St Louis, MO, USA) and 5 IU human chorionic gonadotropin (HCG; Sigma) after 46–47 h (Protocol#220MSK-Mi approved by the National Institute on Aging Animal Care and Use Committee). Unfertilized eggs were harvested at 21 h post-HCG according to the standard method (Nagy *et al.*, 2003, "Manipulation of the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory Press, New York). After removing cumulus cells by incubation in M2 medium (MR-015-D) supplemented with bovine testicular hyaluronidase (HY, 0.1% (w/v), 300 Umg-1), unfertilized eggs were

thoroughly washed, selected for good morphology and collected. Fertilized eggs (1-cell embryos) were also harvested from mated superovulated mice in the same way as unfertilized eggs. Fertilized eggs (1-cell embryos) were cultured in synthetic oviductal medium enriched with potassium (KSOMaa MR-121-D) at 37°C in an atmosphere of 5% CO2. For the embryo transfer procedure, 3.5 d.p.c. blastocysts were transferred into the uteri of 2.5 d.p.c. pseudopregnant ICR female mice.

To synchronize *in vitro* embryo development, embryos with two pronuclei (PN) were selected. When some of these 1-cell stage embryos started to cleave, the early 2-cell stage embryos were selected and transferred to another microdrop culture. The early 2-cell stage embryos were cultured until some of them started 2nd cleavage and the embryos that were still at the 2-cell stage were collected. These embryos were synchronized at the late 2-cell stage.

DNA was microinjected into embryos according to the following procedures.

(1) Pronuclear injection: Plasmid vectors constitutively expressing a siRNA against mouse Zscan4 were constructed by inserting the following target sequences in a pRNAT-U6.1/Neo vector (GenScript Corp., Scotch Plains, NJ, USA), shZscan4 (gagtgaattgctttgtgtc, SEQ ID NO: 9) and siControl (randomized 21-mer, agagacatagaatcgcacgca, SEQ ID NO: 10). This vector contains a green fluorescence protein (GFP) marker under a cytomegalovirus (CMV) promoter. For RNA interference experiments, 1-2 pl (2-3 ng/μl) of a linearized vector DNA (shZscan4 or shControl) was microinjected into the male pronucleus of zygotes. A plasmid vector constitutively expressing the Zscan4d gene was constructed by cloning the CDS of Zscan4d into a plasmid pPyCAGIP (Chambers *et al.*, *Cell 113*:643-655, 2003). For overexpression experiments, 1-2 pl (2-3 ng/l) of plasmid DNA (Zscan4d-inserted or no insert pPyCAGIP vector) linearized by *ScaI* was microinjected into the male pronucleus of zygotes.

(2) Cytoplasmic injection: Transient RNA interference experiments were carried out by microinjecting ~10 pl (5 ng/μl) of oligonucleotide (siZscan4, plus-siZscan4, and siControl) into the cytoplasm of zygotes. The optimal amount of siRNA was determined by testing different concentrations of siRNA (4, 20, and 100 ng/μl).

All siRNAs were resuspended and diluted with the microinjection buffer (Specialty Media). The transfer of cultured blastocysts into pseudopregnant recipients was done according to the standard protocol (Nagy *et al.*, 2003, "Manipulation of the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory Press, New York). All media were purchased from Specialty Media (Phillipsburg, NJ).

Culture of ES cells and blastocyst outgrowth

A mouse ES cell line (129.3 line derived from strain 129 and purchased from The Transgenic Core Laboratory of the Johns Hopkins University School of Medicine, Baltimore, MD, USA) was first cultured for two passages into a gelatin-coated culture dish in the presence of leukemia inhibitory factor (LIF) to remove contaminating feeder cells. Cells were then seeded on gelatin coated 6-well plates at the density of 1-2 x 10^5/well (1-2 x 10^4/cm^2) and cultured for 3 days with complete ES medium (DMEM, 15% FBS; 1000 U/ml ESGRO (mLIF; Chemicon, Temecula, CA); 1mM sodium pyruvate; 0.1 mM NEAA; 2 mM glutamate; 0.1 mM beta-mercapto ethanol and 50 U/50 µg per ml penicillin/streptomycin).

For the outgrowth experiments, blastocysts at 3.5 days post coitum (d.p.c.) were cultured individually in DMEM (Gibco catalog no. 10313-021) supplemented with 15% fetal bovine serum, 15 mM HEPES buffer, 100 units/ml of penicillin, 100 µg/ml of streptomycin, 100 µM nonessential amino acids, 4.5 mM of L-glutamine, and 100 µM of ß-mercapto ethanol on gelatinized chamber slides at 37°C in 5% CO2.

Whole Mount In situ hybridization (WISH)

A plasmid DNA (clone C0348C03) was digested with *SalI/NotI* and transcribed *in vitro* into digoxigenin-labeled antisense and sense probe as control. Embryos obtained from young (7 weeks old) B6D2F1/J mice were fixed in 4% paraformaldehyde and used to perform whole mount *in situ* hybridization (WISH) according to the previously described protocol. WISH was also carried out on cultured ES cells according to the same protocol (Yoshikawa *et al.*, *Gene Expr. Patterns 6*:213-224, 2006).

Quantitative reverse transcriptase PCR

Embryos for quantitative reverse transcriptase (qRT)-PCR experiments were collected as described above and harvested at 23, 43, 55, 66, 80 and 102 hours post-hCG for 1-cell, early 2 cell, late 2-cell, 4-cell, 8-cell, morula and blastocyst embryos, respectively. Three subsets of 10 synchronized and intact embryos were transferred in PBT 1X (PBS supplemented 0.1% Tween X20) and stored in liquid nitrogen. These pools of embryos were mechanically ruptured by a freeze/thaw and directly used as a template for cDNA preparations. The Ovation system (NuGen technologies, San Carlos, CA, USA) was used to synthesize cDNAs from each pool. The cDNAs were then diluted to 1:25 in a total of 1000 µl and 2 µl was used as a template for qPCR. The qPCR was performed on the ABI 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA) as previously described (Falco *et al.*, *Reprod. Biomed. Online 13*:394-403, 2006) and data were normalized by *Chuk* and *H2afz* with the ΔΔCt method (Falco *et al.*, *Reprod. Biomed. Online 13*:394-403, 2006; Livak and Schmittgen, *Methods 25*:402-408, 2001). Embryos subjected to RNA interference experiments were analyzed in the same way as described above for the normal preimplantation embryos

Example 2: Identification of 2-cell-specific genes during preimplantation development

After fertilization, the maternal genetic program governed by maternally stored RNAs and proteins must be switched to the embryonic genetic program governed by *de novo* transcription, called zygotic genome activation (ZGA), from the newly-formed zygotic genome (DePamphilis *et al.*, "Activation of Zygotic Gene Expression" *In* Advances in Developmental Biology and Biochemistry, Vol. 12, pp. 56-84, Elsevier Science B.V., 2002; Latham and Schultz, *Front Biosci. 6*:D748-759, 2001). The ZGA is one of the first and most critical events in animal development. Earlier reports have established that the ZGA begins during the 1-cell stage (Aoki *et al.*, *Dev. Biol. 181*:296-307, 1997) (Nothias *et al.*, *J. Biol. Chem. 270*:22077-22080, 1995; Ram and Schultz, *Dev. Biol. 156*:552-556, 1993). However, global gene expression profiling by DNA microarray analysis has recently revealed that nearly all genes identified for their increase of expression at the 1-cell stage were

insensitive to alpha-amanitin treatment, which blocks RNA polymerase II (Hamatani *et al.*, *Dev. Cell 6*:117-131, 2004; Zeng and Schultz, *Dev. Biol. 283*:40-57, 2005). Thus, these studies not only identified many ZGA genes, but also revealed that *de novo* transcription of the zygotic genome begins during the 2-cell stage of mouse preimplantation development (Hamatani *et al.*, *Dev. Cell 6*:117-131, 2004; Zeng and Schultz, *Dev. Biol. 283*:40-57, 2005). Furthermore, it has been shown that the major burst of ZGA does not occur at the early 2-cell stage, but during the late 2-cell stage (Hamatani *et al.*, *Dev. Cell 6*:117-131, 2004).

Arrest of development at the 2-cell stage has been reported for the loss-of-function mutants of *Mater/Nalp5* (Tong *et al.*, *Nat. Genet. 26*:267-268, 2000), *Mhr6a/Ube2a* (Roest *et al.*, *Mol. Cell Biol. 24*:5485-5495, 2004) and *Brg1/Smarca4* (Bultman *et al.*, *Genes Dev. 20*:1744-1754, 2006). Although the timing of the developmental arrest coincides with that of the ZGA, these genes are expressed during oogenesis and stored in oocytes, but are not transcribed in the 2-cell stage. Therefore, these maternal effect genes are not suitable for the study of the ZGA. Previously the ZGA has been studied using either exogenous plasmid-borne reporter genes Nothias *et al.*, *J. Biol. Chem. 270*:22077-22080), or endogenous, but rather ubiquitously expressed genes, such as *Hsp70.1* (Christians *et al.*, 1995), *eIF-4C* (Davis *et al.*, *Dev. Biol. 174*:190-201, 1996), *Xist* (Zuccotti *et al.*, *Mol. Reprod. Dev. 61*:14-20, 2002) and other genes (DePamphilis *et al.*, "Activation of Zygotic Gene Expression" *In* Advances in Developmental Biology and Biochemistry, Vol. 12, pp. 56-84, Elsevier Science B.V., 2002). Although TEAD-2/TEF-4 (Kaneko *et al.*, *Development 124*:1963-1973, 1997) and Pou5f1/Oct4 (Palmieri *et al.*, *Dev. Biol. 166*:259-267, 1994) are considered as transcription factors selectively expressed at ZGA (DePamphilis *et al.*, "Activation of Zygotic Gene Expression" *In* Advances in Developmental Biology and Biochemistry, Vol. 12, pp. 56-84, Elsevier Science B.V., 2002), these genes are known to be expressed in cells other than 2-cell embryos. It is thus important to identify and study individual ZGA genes, especially the genes expressed exclusively at the 2-cell stage.

Global gene expression profiling of preimplantation embryos was previously carried out and a group of genes was identified that showed transient spike-like expression in the 2-cell embryo (Hamatani *et al.*, *Dev. Cell 6*:117-131, 2004). By

examining the expression of these genes in the public expressed sequence tag (EST) database (NCBI/NIH), a novel gene was identified represented by only 29 cDNA clones out of 4.7 million mouse ESTs. These cDNA clones have been isolated from cDNA libraries derived from ES cells and preimplantation embryos. Furthermore, the previous DNA microarray data showed that the expression of this gene is detected in ES cells, but not in embryonal carcinoma (EC) cells (F9 and P19), trophoblast stem (TS) cells, or neural stem/progenitor (NS) cells (Aiba *et al.*, *Stem Cells 24*:889-895, 2006).

One of the cDNA clones derived from ES cells (clone number C0348C03; (Sharov *et al.*, *PLoS Biol. 1*:E74, 2003)) was completely sequenced by the Mammalian Gene Collection (MGC) project (Genbank Accession No. BC050218; SEQ ID NO: 11 (Gerhard *et al.*, *Genome Res. 14*:2121-2127, 2004)). Whole mount *in situ* hybridization (WISH) using this cDNA clone as a probe detected high level of transcripts in late 2-cell embryos (Fig. 1A). The transcript was not detected in unfertilized eggs and embryos in other preimplantation stages including 3-cell embryos, suggesting a high specificity of gene expression at the late 2-cell stage and a relatively short half-life of the transcripts. Quantitative reverse-transcriptase PCR (qRT-PCR) analysis confirmed the WISH results (Fig. 1B). Previous microarray analysis showed that the expression of this gene at the late 2-cell stage was suppressed in embryos treated with α-amanitin (a blocker of RNA pol II-based transcription) (Hamatani *et al.*, *Dev. Cell 6*:117-131, 2004), confirming that this gene is transcribed *de novo* during the major burst of ZGA. The transient expression pattern was observed in both *in vitro* cultured embryos and freshly isolated *in vivo* embryos (Hamatani *et al.*, *Dev. Cell 6*:117-131, 2004).

Example 3: Structure and expression of Zscan4 paralogous genes

The full-length cDNA sequence (BC050218; SEQ ID NO: 11) of 2292 bp was organized into 4 exons, encoding a protein of 506 amino acids (Fig. 2A). Because this cDNA clone was isolated from a cDNA library made from ES cells (Sharov *et al.*, *PLoS Biol. 1*:E74, 2003), another cDNA clone was isolated by performing RT-PCR on RNAs isolated from late 2 cell-stage embryos and completely sequenced (SEQ ID NO: 21). This 2268 bp cDNA clone encoded a protein of 506 amino acids. DNA

sequence and protein sequences clearly showed that these two cDNAs (SEQ ID NOs: 11 and 21) were two different genes with close similarity. Domain prediction analysis revealed a SCAN (Leucine Rich Element) domain and four zinc finger domains at the N- and C-terminal ends, respectively (Fig. 2B). A hypothetical human ortholog - zinc finger and SCAN domain containing 4 (*ZSCAN4*) was also identified that shares 45% of amino acid sequence similarity with the high conservation in SCAN (50%) and zinc finger domains (59%) (Fig. 7).

Alignment of full-length cDNA sequences (SEQ ID NOs: 11 and 21) to the mouse genome sequence (mm7) revealed multiple hits in the proximal region of chromosome 7, the syntenic region of human *ZSCAN4* (Fig. 8). One notable feature of this genome region was repetitions of a very similar sequence segment. The sequences of each copy of Zscan4 and the surrounding region were very similar to each other, leaving the assembled genome sequences of this region less accurate than those of other regions. To understand the genome structure of this region better, individual BAC clone sequences were manually reassembled from this region into ~850 kb genome sequence contigs (Fig. 3A). Because it was difficult to find a hybridization probe or oligonucleotides to distinguish each copy, restriction enzymes were used that can distinguish small sequence differences among gene copies. Southern blot analysis was carried out by digesting C57BL/6J mouse genomic DNAs with *TaqI* alone, *MspI* alone, or *TaqI/MspI* (Fig. 3B and C). All the detected DNA fragments confirmed nine paralogous Zscan4 genes predicted in the assembled genome sequences.

The full-length cDNA sequence (BC050218; SEQ ID NO: 11) was then aligned to the assembled genome sequence and nine gene copies were found, all of which had multi-exon gene organizations (Fig. 2, 3A). Three gene copies were apparently pseudogenes as no evidence was found that they were transcribed based on available EST information and sequencing analysis of RT-PCR products. Therefore, the genes were named *Zscan4-ps1* (SEQ ID NO: 12), *Zscan4-ps2* (SEQ ID NO: 13), and *Zscan4-ps3* (SEQ ID NO: 14), according to the convention of mouse gene nomenclature. Because the remaining 6 gene copies were transcribed and encoded ORFs, they were named *Zscan4a* (SEQ ID NO: 15), *Zscan4b* (SEQ ID NO: 17), *Zscan4c* (SEQ ID NO: 19), *Zscan4d* (SEQ ID NO: 21), *Zscan4e* (SEQ ID NO: 23) and

Zscan4f (SEQ ID NO: 25). Three of the these genes, *Zscan4a*, *Zscan4b*, and *Zscan4e*, encoded ORFs of 360, 195 and 195 amino acids, respectively, which included the SCAN domain, but not the four zinc finger domains (Fig. 2B).

The remaining three genes, *Zscan4c*, *Zscan4d* and *Zscan4f*, encoded full-length ORFs (506 amino acids). The main features of these genes are summarized in Fig. 3A. *Zscan4c* corresponds to the cDNA clone isolated from ES cells (C0348C03; Genbank Accession No. BC050218; Gm397; SEQ ID NO: 11). *Zscan4d* corresponds to the cDNA clone isolated from 2-cell embryos (SEQ ID NO: 21). *Zscan4f* corresponds to a gene predicted from the genome sequence (Genbank Accession No. XM_145358; SEQ ID NO: 27). Similarities of both ORFs and mRNAs between these three genes were very high (Fig. 7). Thus, it is most likely that these three genes have the same function. To measure the expression levels of each paralog, DNA sequences of the nine Zscan4 paralogs were analyzed by the Clustal X multiple-sequence alignment program, which showed the presence of sequence differences specific to each paralog. To examine the expression levels of each gene in 2-cell embryos and ES cells, cDNA fragments amplified by RT-PCR from 2-cell embryos and ES cells were sequenced. The expression level of each paralog were estimated based on the amplitudes of each nucleotide at polymorphic sites. The results are summarized in Fig. 3A. In 2-cell embryos, *Zscan4d* was a predominant transcript (90%). In contrast, in ES cells, *Zscan4c* was a predominant transcript (40%), although *Zscan4f* was a lesser, but significant transcript (24%). These results were consistent with the origin of each cDNA clone; *Zscan4c* was derived from the ES cell cDNA library, whereas *Zscan4d* was derived from the 2-cell embryo library.

Example 4: Function of Zscan4 in preimplantation development

As a first step to characterize the function of Zscan4 genes, the studies focused on preimplantation development. Initially a possibility to carry out a standard gene targeting strategy was explored, but it was difficult for the following three reasons. First, sequences of Zscan4 paralogs and surrounding genomic regions are too similar to design targeting constructs for specific genes. Second, it is highly likely that Zscan4d$^{-/-}$ phenotype can be compensated functionally by other Zscan4 paralogs, because in addition to predominantly-expressed *Zscan4d*, at least 3 other similar

copies (*Zscan4a, Zscan4e*, and *Zscan4f*) were also transcribed in 2-cell embryos. Third, the presence of other predicted genes, though not annotated as genes yet, within ~850 kb Zscan4 locus makes a strategy to delete the entire Zscan4 locus less attractive. Therefore, siRNA technology was used. Although RNAi and siRNA technology has been successfully used for blocking the expression of specific genes in preimplantation embryos (Kim *et al.*, *Biochem. Biopys. Res. Commun. 296*:1372-1377, 2002; Stein *et al.*, *Dev. Biol. 286*:464-471, 2005), widely-recognized off-target effects are generally a major concern (Jackson *et al.*, *Rna 12*:1179-1187, 2006; Scacheri *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 101*:1892-1897, 2004; Semizarov *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 100*:6347-6352, 2003). To increase the confidence of the effects by siRNA against Zscan4, the siRNA experiments were carried out by three independent siRNA technologies, an oligonucleotide-based siRNA (denoted here siZscan4 and obtained from Invitrogen); a vector-based shRNA (denoted here shZscan4 and obtained from Genscript); and a mixture of oligonucleotide siRNAs (denoted here plus-siZscan4 and obtained from Dharmacon) (Fig. 4A, B). Oligonucleotide sequences used for siZscan4, shZscan4, plus-siZscan4 matched 100% with cDNA sequences of *Zscan4a, Zscan4b, Zscan4c*, *Zscan4d*, *Zscan4e* and *Zscan4f*, except for shZscan4 with 2 bp mismatches with *Zscan4b* and *Zscan4e* (Fig. 4A, B).

A shZscan4 vector was microinjected into the male pronucleus of zygotes at 21-23 hours after the hCG injection and embryos were observed during preimplantation development (Fig. 4C and D). At 61 hours post-hCG, when the majority (58.8%) of shControl-injected embryos have already reached the 4-cell stage, the majority (78.8%) of shZscan4-injected embryos remained at the 2-cell stage. By 98 hours post-hCG, when the majority (70.0%) of shControl-injected embryos have reached blastocyst stage, the majority (52.5%) of shZscan4-injected embryos reached only morula stage. A significant reduction (~95%) of Zscan4 RNA levels was confirmed by the qRT-PCR analysis (Fig. 4E). Taken together, these results indicate that the development of shZscan4-injected embryos was delayed for about 24 hrs between the 2- and 4-cell stages, followed by progression to the later stages at a speed comparable to that of shControl-injected embryos. Essentially the same results were obtained using two different siRNA technologies: siZscan4 (Fig. 9) and plus-siZscan4 (Fig. 10).

siZscan4-injected embryos formed normal looking early blastocysts (3.5 d.p.c.), but often failed to form expanded blastocysts (4.5 d.p.c.; 45% of siZscan4-injected embryos versus 6% of siControl-injected embryos; Fig. 9B). To test whether these blastocysts had any compromise even at 3.5 d.p.c., shZscan4-injected blastocysts were transferred to the uterus of pseudo-pregnant mice. None of the shZscan4-injected blastocysts implanted, whereas most shControl-injected embryos implanted (Table 1). *In vitro* blastocyst outgrowth experiments determined that cells of shZscan4-injected blastocysts failed to proliferate in culture (Table 1). These results clearly demonstrated that the transient expression of Zscan4 at the late 2-cell stage is required for the development of proper blastocysts.

Table 1

Blastocyst outgrowth (A) and post-implantation development (B) of embryos received pronuclear injection of shZscan4 or shControl

A Blastocyst Outgrowth	Number of tested blastocysts	Number of successful outgrowth
shZscan4	16	0
shControl	17	7
B Embryo Transfer	Number of blastocysts transferred to pseudo-pregnant mother	Number of pups born
shZscan4	8	0
shControl	10	4

*A shZscan4 or shControl vector was microinjected into the male pronucleus of zygotes at 21-23 hours after the hCG injection. Early blastocysts (3.5 d.p.c.) formed from these embryos were subjected to tests of blastocysts outgrowth (A) and embryo transfer (B). In the outgrowth assay, the presence of proliferating cells after 6 days in culture was considered as successful outgrowth.

The notion that the reduction of Zscan4 expression level delays the development of preimplantation embryos at the 2-cell stage was further supported by the fact that when shZscan4 was injected into one of the blastomeres of early 2-cell stage embryos, ~28% of embryos became 3-cell embryos (Fig. 5A). One blastomere

that received shZscan4 injection remained as a 2-cell blastomere, whereas the other blastomere cleaved into two smaller blastomeres with the size of 4-cell blastomeres (Fig. 5D). Subsequently, these embryos (24%) became unevenly cleaved embryos, typically 5-cell embryos, with one 2-cell-sized blastomere and four 8-cell-sized blastomeres (Figure 5B, E). These embryos eventually formed blastocyst-like structures, but they seemed to be the mixtures of blastocyst-like cell mass and morula-like cell mass, which was often GFP-positive, a marker for shRNA-injected blastomere (Fig. 5C, F, G). In contrast, when shControl was injected into one of the blastomeres at the early 2-cell stage, nearly all embryos cleaved normally (Fig. 5A, B, C).

To investigate the effect of prolonged *Zscan4d* expression on preimplantation development, *Zscan4d* was overexpressed by microinjecting a Zscan4d-expressing plasmid into the male pronucleus of zygotes. Although the Zscan4d plasmid-injected embryos showed a rate of development similar to control plasmid-injected embryos, the former blastocysts failed to produce the outgrowth (Table 2A) and failed to implant (Table 2B). The results suggest that the timely downregulation of *Zscan4d* is also important for the proper development of blastocysts.

Table 2

Blastocyst outgrowth (A) and post-implantation development (B) of embryos received pronuclear injection of a *Zscan4d*-expressing plasmid or a control plasmid

A Blastocyst Outgrowth	Number of tested blastocysts	Number of successful outgrowth
Zscan4d-expressing plasmid	10	2
Control plasmid	15	11
B Embryo Transfer	Number of blastocysts transferred to pseudo-pregnant mother	Number of pups
Zscan4d-expressing plasmid	10	0
Control plasmid	14	5

*A plasmid vector constitutively expressing Zscan4d gene or control empty vector was microinjected into the male pronucleus of zygotes at 21-23 hours after the hCG injection. Early blastocysts (3.5 d.p.c.) formed from these embryos were subjected to the same tests as described in Table 1.

Example 5: Analysis of Zscan4 expression using the Whole Mount *In situ* Hybridization (WISH)

One intriguing aspect of the expression pattern of Zscan4 is the exclusive expression in late 2-cell embryos and ES cells. This appears to be counter-intuitive, because ES cells are derived from the ICM and many genes that are expressed in ES cells are also expressed in the ICM (e.g., Yoshikawa *et al.*, *Gene Expr. Patterns* *6*:213-224, 2006). Therefore the expression of Zscan4 in blastocysts, blastocyst outgrowth, and ES cells was examined using WISH. The results demonstrated that the expression of Zscan4 was not detected anywhere in blastocysts, including the ICM and the early blastocyst outgrowth (Fig. 6A). However, the expression of Zscan4 began to be detected in a small fraction of cells by the day 6 of the outgrowth. Surprisingly, the strong expression of Zscan4 was detected in only a small fraction of ES cells in undifferentiated colonies. In contrast, the expression of *Pou5f1* (*Oct3/4*), a well-known marker for pluripotency, was detected in the ICM of blastocysts, a large fraction of the cells in the blastocyst outgrowth, and the majority of ES cells in undifferentiated colonies (Fig. 6A). Due to the close similarity of cDNA sequences, each Zscan4 paralog could not be distinguished by WISH, but the expression analysis by sequencing RT-PCR products mentioned above indicates that *Zscan4c* and *Zscan4f* were the genes detected in the subpopulation of the cells in blastocyst outgrowth and ES cells by WISH.

Example 6: Zscan4 promoter expression vector

As described in previous Examples herein, Zscan4 expression is only detected in a subpopulation of undifferentiated ES cells. In order to identify this subpopulation of ES cells, and to identify any other cell expressing Zscan4, an expression plasmid was developed which comprises a Zscan4c promoter sequence and the Emerald reporter gene (a variant of green fluorescent protein). The

components and orientation of the expression vector are illustrated in Figure 11. The sequence of the Zscan4c promoter-Emerald expression vector is set forth as SEQ ID NO: 28. The nucleotide ranges of SEQ ID NO: 28 of the components of the expression vector are provided in Table 3.

Table 3

Zscan4c Promoter-Emerald Expression Vector

Component	Nucleotides of SEQ ID NO: 28
Zscan4c promoter	1-3347
TATA box	2483-2489
Zscan4c exon 1	2541-2643
Zscan4c intron 1	2644-3250
Zscan4c exon 2 (partial)	3251-3347
Emerald start codon	3398-3400
Emerald reporter gene	3398-4117
TK poly A signal	4132-4403
EM7 promoter	5257-5323
Blasticidin selection gene	5330-5722
SV40 polyA signal	5880-6010

Mouse ES cells were transfected with the Zscan4c promoter expression vector and analyzed by fluorescence activated cell sorting to identify Emerald-positive cells and Emerald-negative cells. If Zscan4 is expressed in a cell, it is Emerald-positive. The results show approximately 3-5% of mouse ES cells express Zscan4 (Figure 12).

Sorted cells were collected and analyzed by quantitative real time PCR (qPCR) for expression of Zscan4c and Pou5f1 (also known as Oct3, Oct4, Oct3/4), a well known marker for pluripotency. As shown in Figure 12, Pou5f1 is expressed at the same level in both Emerald-positive and Emerald-negative cells, whereas Zscan4c is more highly expressed in Emerald-positive cells than in Emerald-negative cells. The data indicate that the Zscan4c promoter sequence used in this vector can reproduce the expression of endogenous Zscan4c gene, and thus the Zscan4c promoter-Emerald expression vector can be used to purify Zscan4-expressing cells. The data also indicate that both Zscan4-expressing cells and non-expressing cells retain the pluripotency-marker Pou5f1 expression, thus this subpopulation of ES cells cannot be identified by a standard pluripotency marker.

Example 7: Mouse ES cell line expressing Emerald under control of the Zscan4 promoter

A mouse ES cell line was established in which the Zscan4c promoter expression vector described in Example 6 was stably incorporated into the cells. The ES cell line expresses Emerald under control of the Zscan4c promoter. After transfecting a linearlized plasmid DNA into mouse ES cells, the cells were cultured in the presence of the selectable marker (blasticidin). The blasticidin-resistant ES cell clones were isolated and used for further analysis.

As described herein, Zscan4 is only expressed in a subpopulation of undifferentiated ES cells (approximately 3-5% of ES cells). Accordingly, the ES cell line incorporating the Zscan4 promoter expression vector exhibits expression in only a small percentage, approximately three percent, of cells.

Example 8: Identification of nine genes co-expressed with Zscan4 in a sub-population of ES cells

Using the mouse ES cell line stably transfected with the Zscan4c promoter (as described in Example 7), DNA microarray analysis was performed to compare gene expression patterns of Emerald(+) and Emerald(-) cells. Emerald(+) and Emerald(-) cells were sorted by FACS and total RNAs were isolated from each cell population. These RNAs were labeled and hybridized to the NIA-Agilent 44K DNA microarray (Agilent Technologies).

Nine genes were identified as being co-expressed with Zscan4: AF067063, Tcstv1/Tcstv3, Tho4, Arginase II, BC061212 and Gm428, Eif1a, EG668777 and Pif1. *In situ* hybridization was performed to confirm expression of these genes in mouse ES cells. The 2-cell embryo-specific expression profiles of six of these genes (AF067063, Tcstv3, Tho4, Arginase II, BC061212 or Gm428) are shown in Figures 13A-G.

Example 9: Trim43 is specifically expressed in 4-cell to morula stage embryos

To identify genes that are specifically expressed at the 8-cell and morula stages, publicly available EST frequency data (TIGR Mouse Gene Index; MGI Library Expression Search; NIA Mouse Gene Index (Sharov *et al.*, *PLoS Bio. 1*:E74, 2003)) and microarray data from mouse preimplantation embryos (Hamatani *et al.*, *Dev. Cell* 6(1):117-31, 2004) were used. After selecting candidate genes, quantitative RT-PCR analysis was carried out to confirm the specific expression pattern of Trim43 (tripartite motif-containing protein 43).

Trim43 expression was detected beginning at the 4-cell embryonic stage and peaked at the morula stage. A low level of Trim43 expression was detected in blastocysts. The function of the Trim43 protein is unknown. The nucleotide and amino acid sequences of Trim43 are provided herein as SEQ ID NO: 32 and SEQ ID NO: 33, respectively. The nucleic acid sequence of the Trim43 promoter is provided herein as SEQ ID NO: 31.

Example 10: Transgenic "rainbow" mouse

As described herein, an expression vector comprising a Zscan4c promoter operably linked to a first heterologous polypeptide (Emerald) and an expression vector comprising a Trim43 promoter operably linked to a second heterologous polypeptide (Strawberry), have been generated. A transgenic mouse (a "rainbow" mouse) can be generated which incorporates both of these expression constructs.

A 7155 base pair DNA fragment containing the Insulator-Zscan4 promoter-emerald and TK polyA and a 8672 base pair DNA fragment containing the Insulator-Trim43 promoter-Strawberry are co-injected into the pronucleus of fertilized mouse eggs (B6C3 X B6).

Embryos obtained from the rainbow mouse will exhibit green color (as a result of expression of Emerald) at the late 2-cell stage, and red color (due to expression of Strawberry) from the 4-cell stage to the morula stage (with peak expression at the morula stage). The expression of Emerald and Strawberry at the appropriate stage of embryonic development indicates proper development of the embryo. Thus, these embryos will be useful for a number of research and clinical purposes. For example, embryos obtained from the rainbow mouse can be used to

develop optimized culture conditions for embryos, which can be applied to human embryos used in the IVF clinic. In addition, these embryos can be used to test chemical compounds or drugs for toxicity to the embryo. The embryos can also be used as indicators of successful nuclear reprogramming for nuclear transplantation procedures.

This disclosure provides methods of inhibiting differentiation of stem cells and promoting blastocyst outgrowth of ES cells. The disclosure further provides a Zscan4 promoter sequence and methods of use, including identification of a subpopulation of stem cells expressing Zscan4. It will be apparent that the precise details of the methods described may be varied or modified without departing from the spirit of the described invention. We claim all such modifications and variations that fall within the scope and spirit of the claims below.

2008230870 09 Aug 2013

CLAIMS

1. A method of inhibiting differentiation of a stem cell, comprising increasing the expression of Zscan4 in the stem cell, thereby inhibiting differentiation of the stem cell, wherein increasing the expression of Zscan4 in the cell comprises transfecting the stem cells with a nucleic acid encoding Zscan4 operably linked to a promoter.

2. The method of claim 1 wherein increasing the expression of Zscan4 comprises increasing the expression of Zscan4a, Zscan4b, Zscan4c, Zscan4d, Zscan4e, Zscan4f, human ZSCAN4 or a combination thereof.

3. The method of claim 1 or claim 2, wherein inhibiting the differentiation of the stem cell increases the viability of the stem cell.

4. The method of any of claims 1-3, wherein inhibiting the differentiation of the stem cell prevents senescence of the stem cell.

5. The method of any of claims 1-4, wherein the stem cell is an embryonic stem cell, an embryonic germ cell, a germline stem cell or a multipotent adult progenitor cell.

6. A method of promoting blastocyst outgrowth of embryonic stem cells, comprising increasing the expression of Zscan4 in the embryonic stem cells, thereby promoting blastocyst outgrowth of the embryonic stem cells, wherein increasing the expression of Zscan4 in the cell comprises transfecting the stem cells with a nucleic acid encoding Zscan4 operably linked to a promoter.

7. The method of claim 6, wherein increasing the expression of Zscan4 comprises increasing the expression of Zscan4a, Zscan4b, Zscan4c, Zscan4d, Zscan4e, Zscan4f, human ZSCAN4 or a combination thereof.

8. The method of claim 1 or claim 6, wherein the nucleic acid encoding Zscan4 comprises SEQ ID NO: 60.

9. The method of claim 1 or claim 6, wherein the nucleic acid encoding Zscan4 is at least 95% identical to Zscan4c (SEQ ID NO: 19), Zscn4d (SEQ ID NO: 21) or Zscan4f (SEQ ID NO: 25).

10. The method of any of claims 1 to 9, wherein the promoter is a constitutive promoter.

2008230870 09 Aug 2013

11. The method of any of claims 1 to 9, wherein the promoter is an inducible promoter.

12. The method of any of claims 1 to 11, comprising transfecting the cells with a vector comprising the nucleic acid encoding Zscan 4.

13. The method of claim 12, wherein the vector is a viral vector.

14. A method of identifying an undifferentiated subpopulation of stem cells expressing Zscan4, comprising transfecting the cells with an expression vector comprising a Zscan4 promoter and a reporter gene, wherein expression of the reporter gene indicates Zscan4 is expressed in the subpopulation of stem cells.

15. The method of claim 14, wherein the Zscan4 promoter is a Zscan4c promoter.

16. The method of claim 14, wherein the expression vector comprises the nucleic acid sequence set forth as SEQ ID NO: 28.

17. An isolated expression vector comprising a Zscan4c promoter operably linked to a nucleic acid encoding a heterologous polypeptide, when used for identifying an undifferentiated subpopulation of stem cells expressing Zscan4.

18. The method of claim 15 or the isolated expression vector of claim 17, wherein the Zscan4c promoter comprises the nucleic acid sequence set forth as nucleotides 1-2540 of SEQ ID NO: 28, nucleotides 1-2643 of SEQ ID NO: 28, nucleotides 1-3250 of SEQ ID NO: 28, or nucleotides 1-3347 of SEQ ID NO: 28.

19. The isolated expression vector of claims 17 or 18, wherein the polypeptide is a marker, enzyme, or fluorescent protein.

20. The isolated expression vector of any of claims 17 to 19, wherein the vector is a viral vector.

21. The isolated expression vector of any of claims 17 to 19, wherein the vector is a plasmid vector.

22. An isolated embryonic stem cell comprising the expression vector of any of claims 17 to 21.

2008230870 09 Aug 2013

23. A method of inhibiting differentiation of a stem cell, or a method of identifying an undifferentiated subpopulation of stem cells expressing Zscan4, substantially as hereinbefore described with reference to the Examples.

24. An isolated expression vector comprising a Zscan4c promoter operably linked to a nucleic acid encoding a heterologous polypeptide substantially as hereinbefore described with reference to Example 6, when used for inhibiting differentiation of a stem cell, promoting blastocyst outgrowth of embryonic stem cells or identifying an undifferentiated subpopulation of stem cells expressing Zscan4.

FIG. 1A

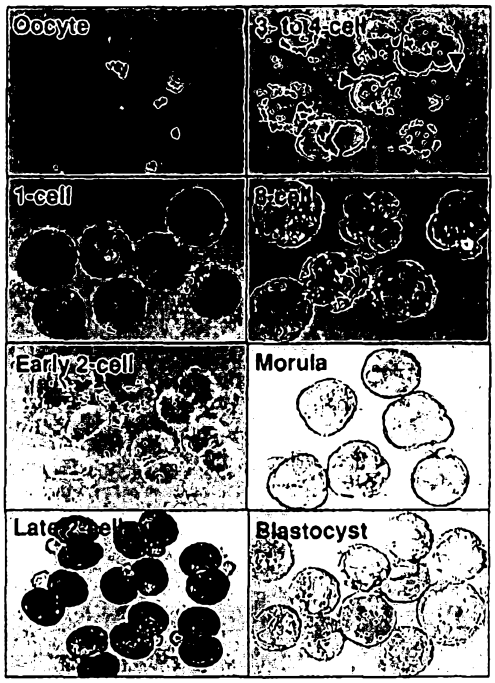

Oocyte
1-cell
Early 2-cell
Morula
Blastocyst

FIG. 1B

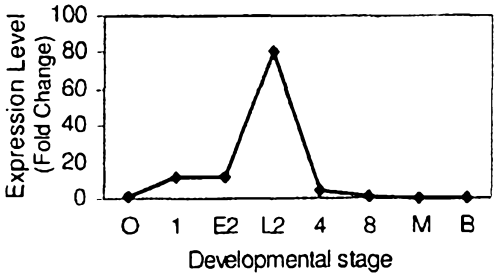

Expression Level
(Fold Change)
100
80
60
40
20
0
O
1
E2
L2
4
8
M
B
Developmental stage

FIG. 2A

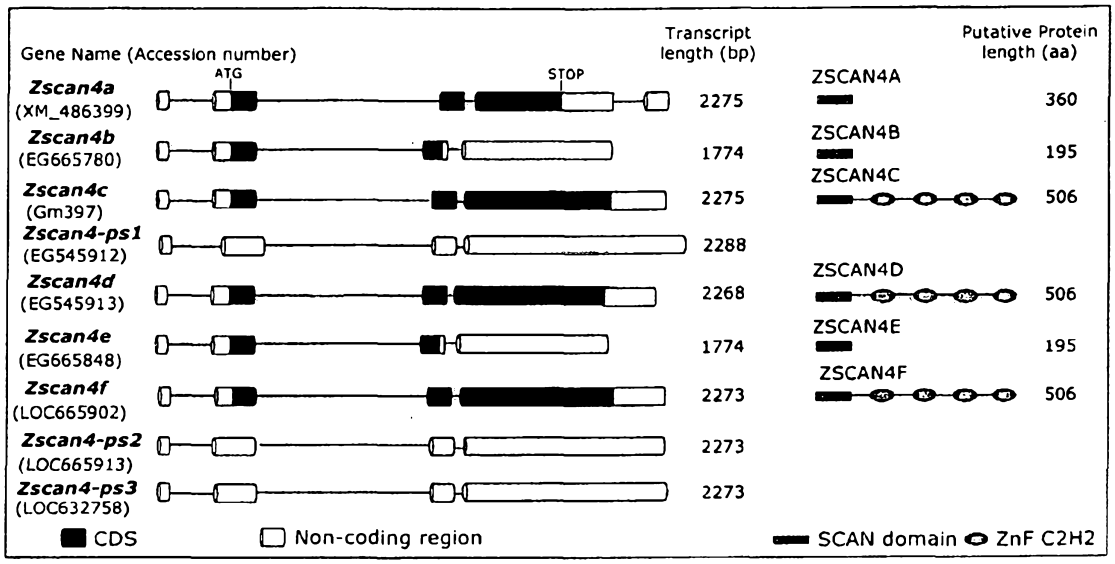
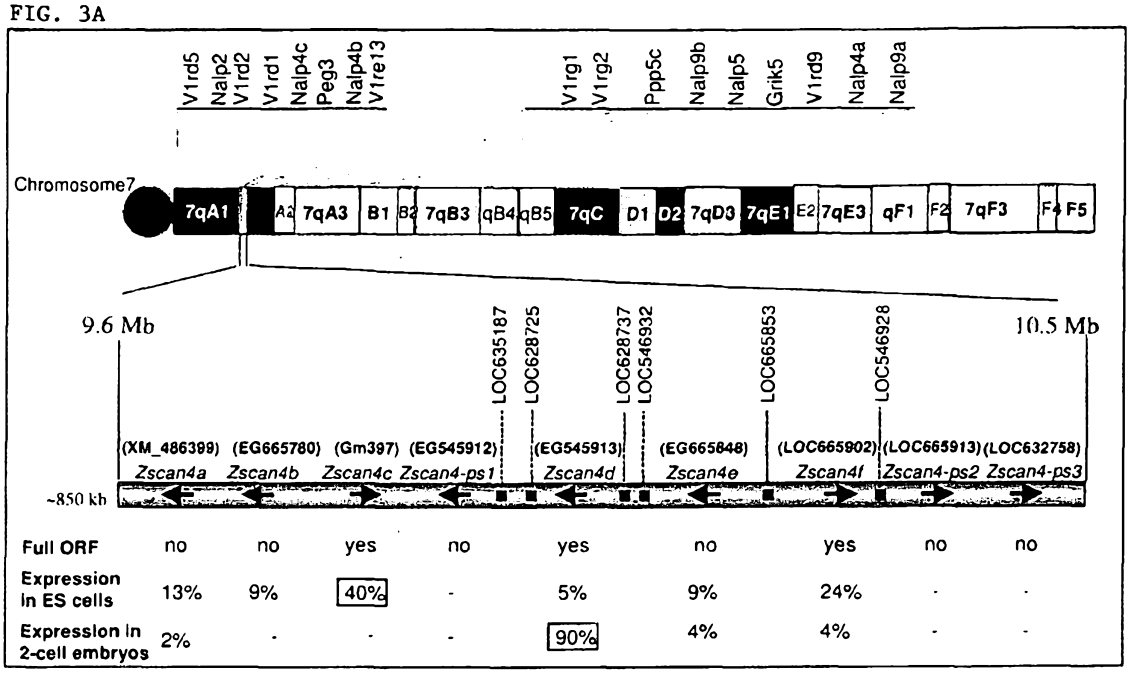
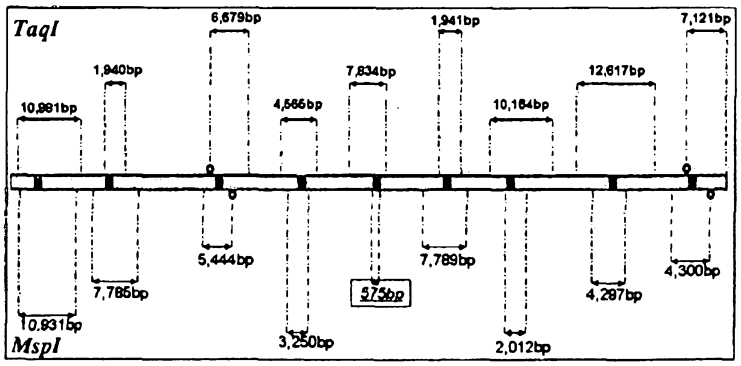

Gene Name (Accession number)
Transcript length (bp)
ATG
STOP
Zscan4a (XM_486399) 2275
Zscan4b (EG665780) 1774
Zscan4c (Gm397) 2275
Zscan4-ps1 (EG545912) 2288
Zscan4d (EG545913) 2268
Zscan4e (EG665848) 1774
Zscan4f (LOC665902) 2273
Zscan4-ps2 (LOC665913) 2273
Zscan4-ps3 (LOC632758) 2273
CDS
Non-coding region

FIG. 2B

Putative Protein length (aa)
ZSCAN4A 360
ZSCAN4B 195
ZSCAN4C 506
ZSCAN4D 506
ZSCAN4E 195
ZSCAN4F 506
SCAN domain
ZnF C2H2

FIG. 3A





♦ = Target sequences of the probe used for Southern Blot hybridization
○ = Restriction sites that generate extra bands in double digestion with *MspI/TaqI*

FIG. 3B

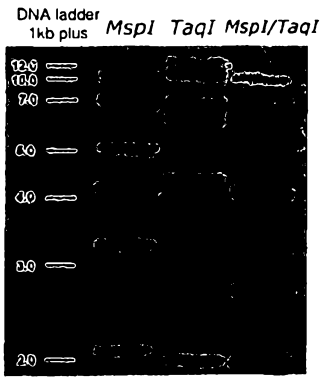


FIG. 3C

FIG. 4A

Name of siRNA	Target positions on cDNA (bp)	Target sequences
Plus-siZscan4 (J-064700-05: Dharmacon)	514-532 (exon II)	gtagcgatatgaggagatt
Plus-siZscan4 (J-064700-06: Dharmacon)	236-254 (exon II)	gaccaacaatttagagttt
Plus-siZscan4 (J-064700-07: Dharmacon)	304-322 (exon II)	caccaagtgctcagctaaa
Plus-siZscan4 (J-064700-08: Dharmacon)	362-380 (exon II)	gctgcaaagtctctggaag
siZscan4 (Zscan4_stealth508: Invitrogen)	508-532 (exon II)	ccagtggtagcgatatgaggagatt
shZscan4 (Genscript)	1463-1481 (exon IV)	gagtgaattgctttgtgtc

FIG. 4B

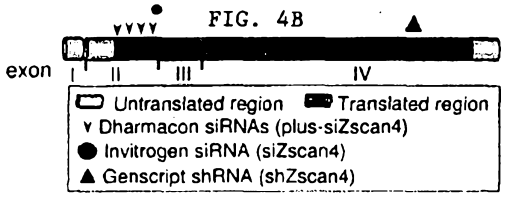
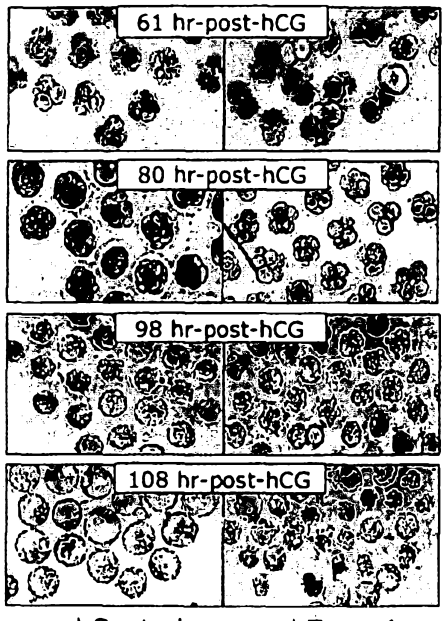


FIG. 4C



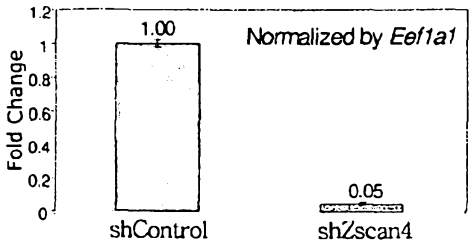


FIG. 4D

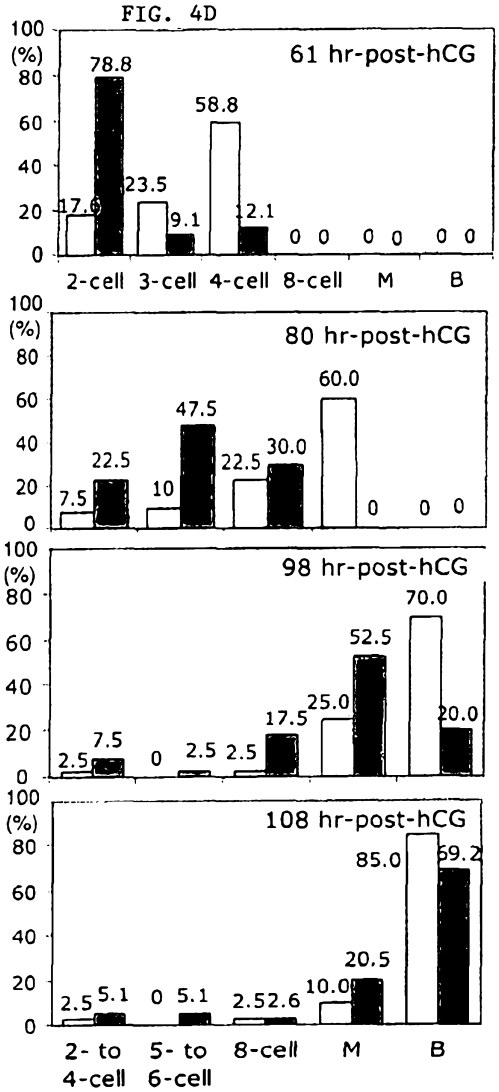


FIG. 4E

FIG. 5A

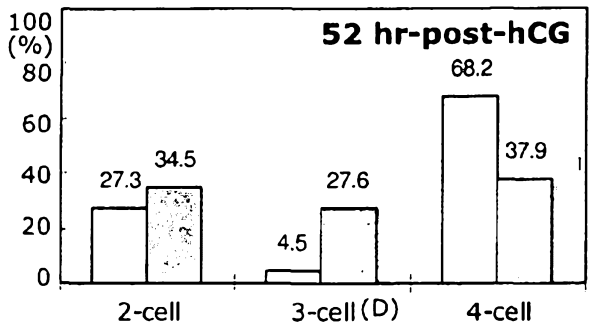


FIG. 5B

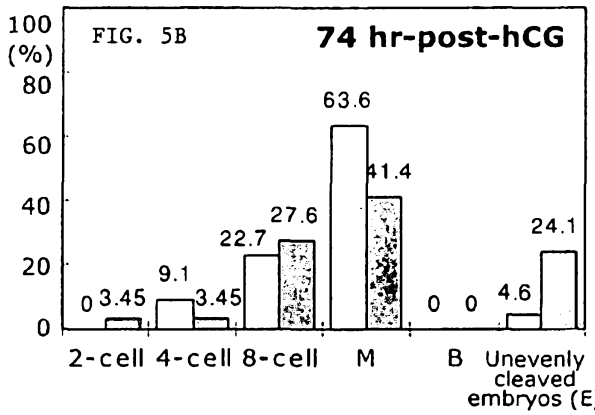


FIG. 5C

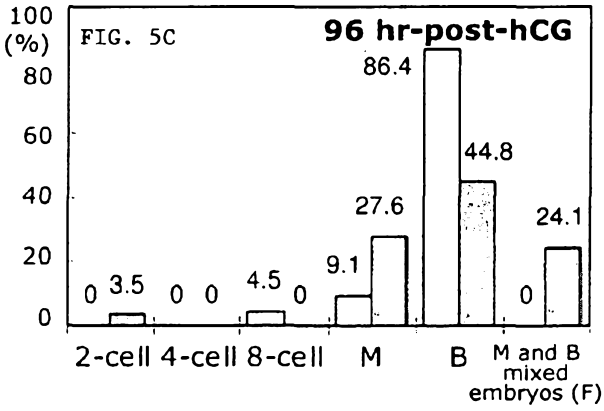


FIG. 5D

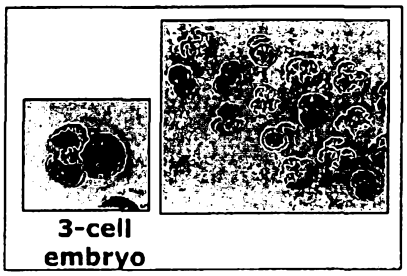


FIG. 5E

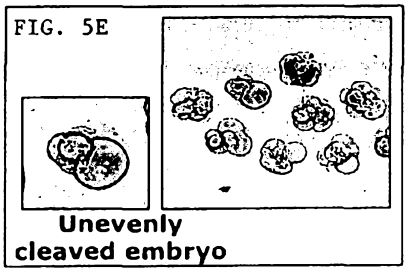
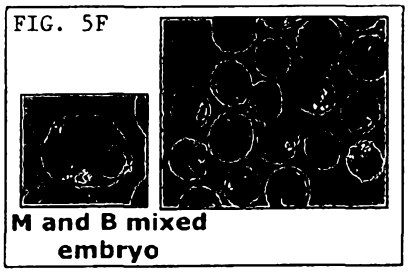


FIG. 5F



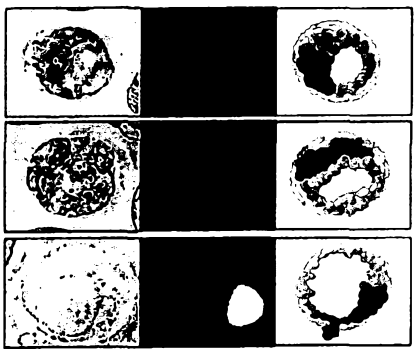
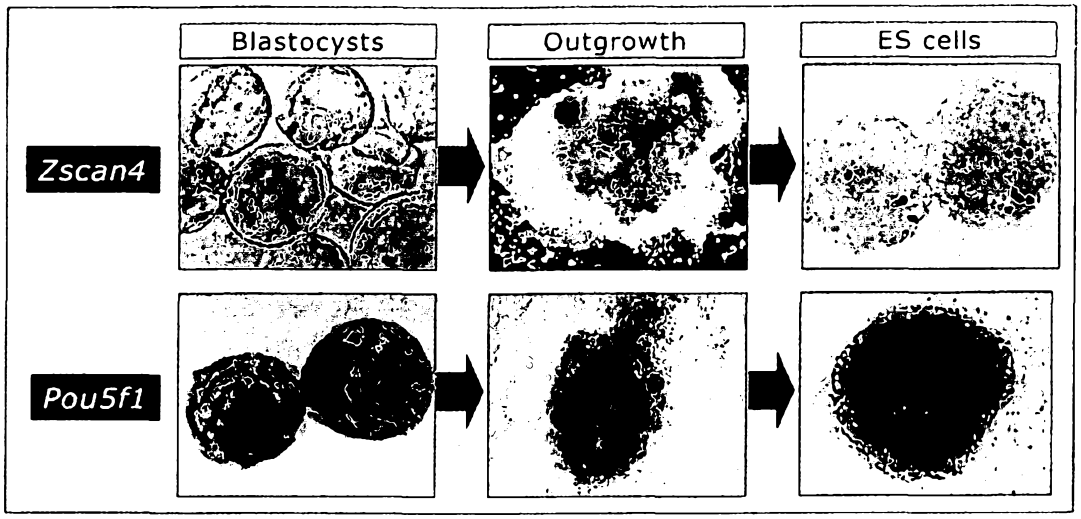


FIG. 5G

FIG. 6A


Blastocysts
Outgrowth
ES cells
Zscan4
Pou5f1

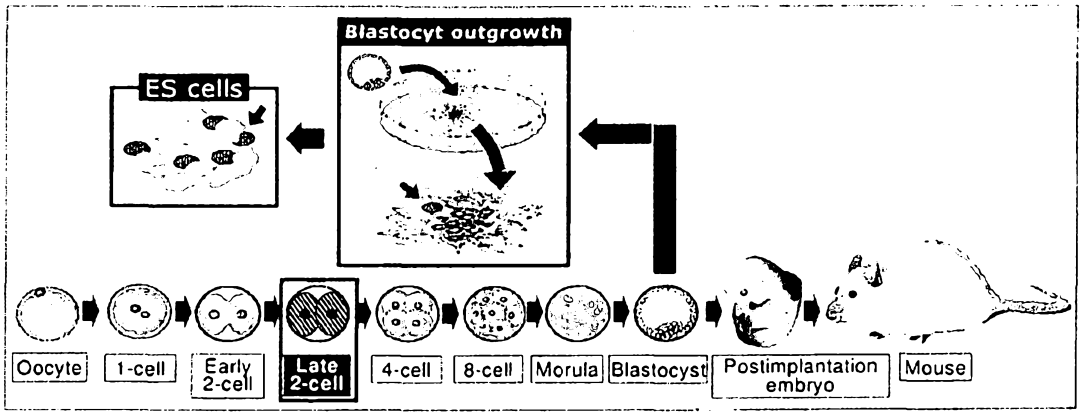

Blastocyt outgrowth
ES cells
Oocyte
1-cell
Early 2-cell
Late 2-cell
4-cell
8-cell
Morula
Blastocyst
Postimplantation embryo
Mouse

FIG. 6B

FIG. 7A

cDNA (length)	Human *ZSCAN4* (2230bp)	Mouse *Zscan4c* (2275bp)	Mouse *Zscan4d* (2268bp)	Mouse *Zscan4f* (2273bp)
ZSCAN4	-	54	55	54
Zscan4c		-	97	99
Zscan4d			-	97
Zscan4f				-

FIG. 7B

CDS (length)	Human *ZSCAN4* (1302bp)	Mouse *Zscan4c* (1518bp)	Mouse *Zscan4d* (1518bp)	Mouse *Zscan4f* (1518bp)
ZSCAN4	-	65	66	65
Zscan4c		-	98	99
Zscan4d			-	98
Zscan4f				-

FIG. 7C

Protein (length)	Human ZSCAN4 (433aa)	Mouse ZSCAN4C (506aa)	Mouse ZSCAN4D (506aa)	Mouse ZSCAN4F (506aa)
ZSCAN4	-	45	44	44
ZSCAN4C		-	95	99
ZSCAN4D			-	94
ZSCAN4F				-

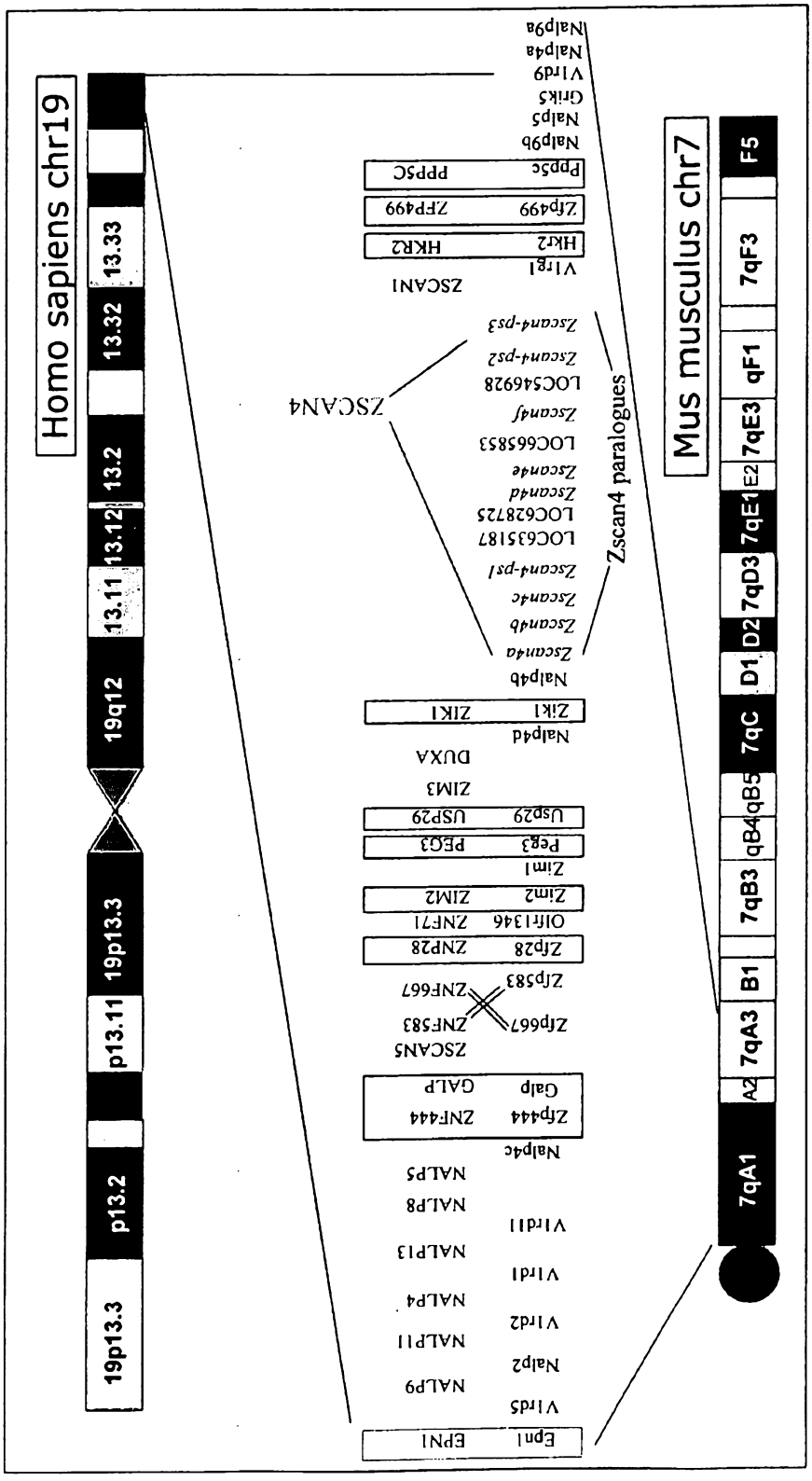
FIG. 7D

SCAN Domain (length)	Human ZSCAN4 (96aa)	Mouse ZSCAN4C (99aa)	Mouse ZSCAN4D (99aa)	Mouse ZSCAN4F (99aa)
ZSCAN4	-	50	50	50
ZSCAN4C		-	98	100
ZSCAN4D			-	98
ZSCAN4F				-

FIG. 7E

ZFP Domain (length)	Human ZSCAN4 (107aa)	Mouse ZSCAN4C (109aa)	Mouse ZSCAN4D (109aa)	Mouse ZSCAN4F (109aa)
ZSCAN4	-	59	58	59
ZSCAN4C		-	99	100
ZSCAN4D			-	99
ZSCAN4F				-

FIG. 8


Homo sapiens chr19
19p13.3
p13.2
p13.11
19p13.3
19q12
13.11
13.12
13.2
13.32
13.33
Mus musculus chr7
7qA1
A2
7qA3
B1
7qB3
qB4
qB5
7qC
D1
D2
7qD3
7qE1
E2
7qE3
qF1
7qF3
F5
Zscan4 paralogues
ZSCAN4
ZSCAN1
Epn1
EPN1
V1rd5
NALP9
Nalp2
NALP11
V1rd2
NALP4
V1rd1
NALP13
V1rd11
NALP8
NALP5
Nalp4c
Zfp444
ZNF444
Galp
GALP
ZSCAN5
Zfp667
ZNF583
ZNF667
Zfp583
Zfp28
ZNP28
Olfr1346
ZNF71
Zim2
ZIM2
Zim1
Peg3
PEG3
Usp29
USP29
ZIM3
DUXA
Nalp4d
Zik1
ZIK1
Nalp4b
Zscan4a
Zscan4b
Zscan4c
Zscan4-ps1
LOC635187
LOC628725
Zscan4d
Zscan4e
LOC665853
Zscan4f
LOC546928
Zscan4-ps2
Zscan4-ps3
Vlrg1
Hkr2
HKR2
Zfp499
ZFP499
Ppp5c
PPP5C
Nalp9b
Nalp5
Grik5
V1rd9
Nalp4a
Nalp9a

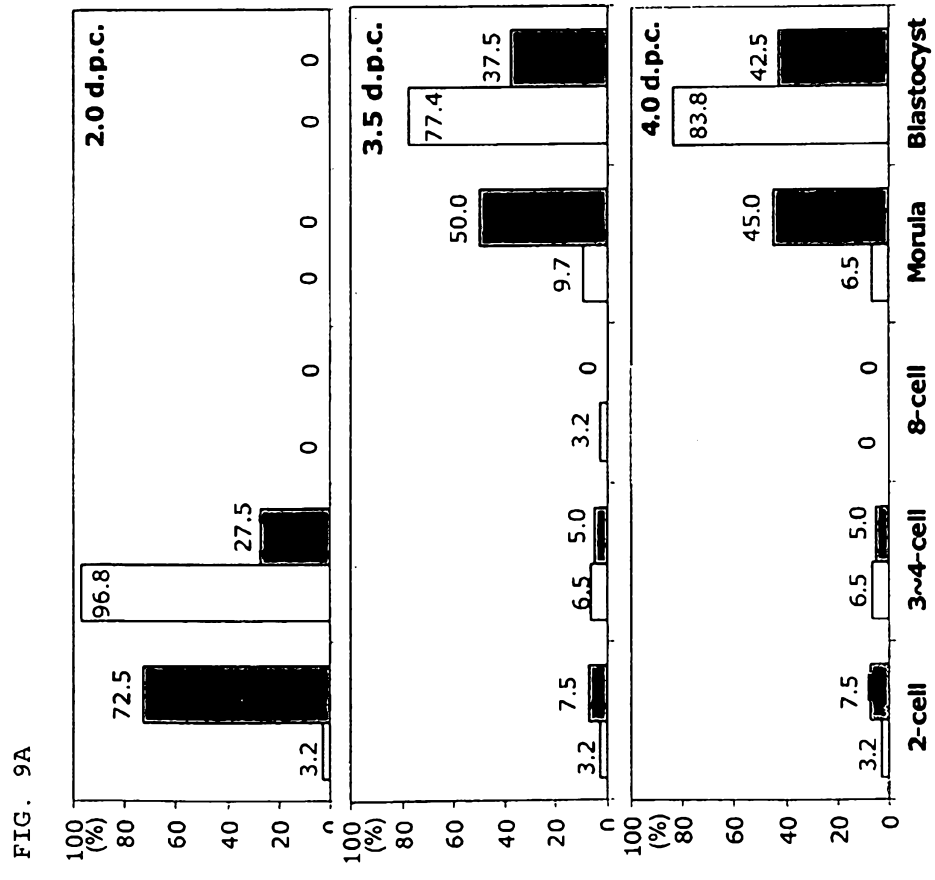

2.0 d.p.c.
3.5 d.p.c.
4.0 d.p.c.
100 (%) 80 60 40 20 0
3.2 72.5 96.8 27.5 0 0 0 0 0
3.2 7.5 6.5 5.0 0 3.2 9.7 50.0 77.4 37.5
3.2 7.5 6.5 5.0 0 0 6.5 45.0 83.8 42.5
2-cell 3~4-cell 8-cell Morula Blastocyst

FIG. 9A

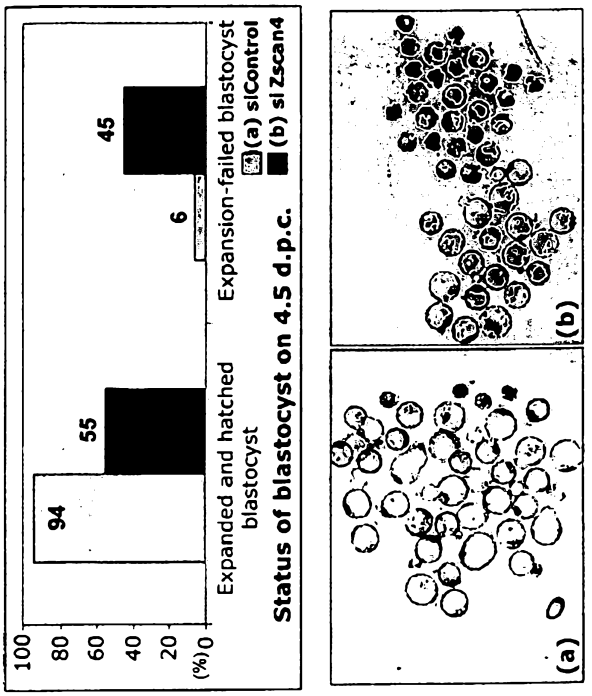

100 80 60 40 20 (%)0
94 55 6 45
Expanded and hatched blastocyst
Expansion-failed blastocyst
(a) siControl
(b) si Zscan4
Status of blastocyst on 4.5 d.p.c.
(a)
(b)

FIG. 9B

FIG. 10A

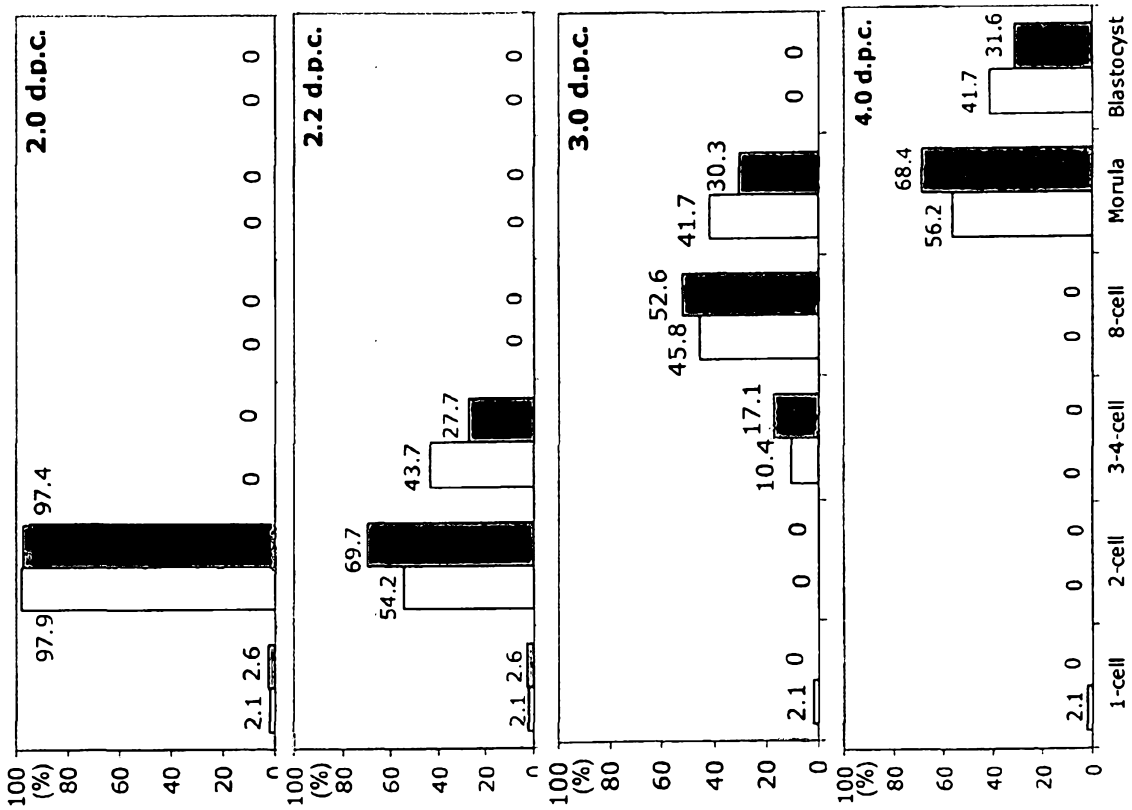


FIG. 10B

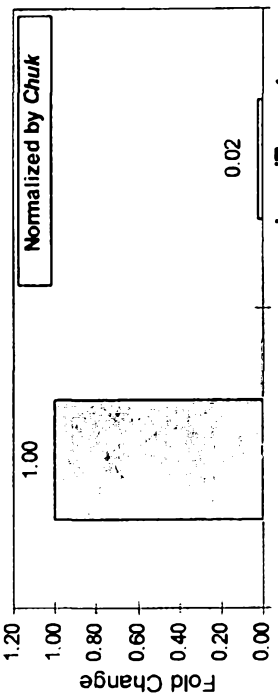


FIG. 10C

ΔΔ Ct'	Average	St Dev'
Scramble	1.00	0
plus-siZscan4	0.02	0.02

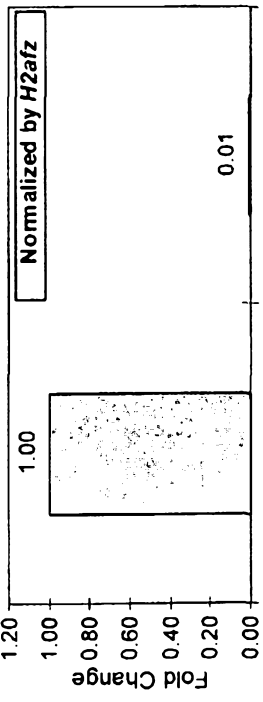


FIG. 10D

ΔΔCt'	Average	St Dev'
Scramble	1.00	0
plus-siZscan4	0.01	0.02

		Developmental delayed embryos of Scramble C1	Scramble C2	Zscan4 C1	Zscan4 C2	Zscan4 C3	Average of developmental delayed embryos of Scramble	Average of developmental delayed embryos of Zscan4
Zscan4	Ct'	24.00	220.5	30.91	27.48	30.61	23.02	29.67
	St Dev'	0.02	0.01	0.34	0.05	0.24	1.38	1.90
Chuk	Ct'	30.83	31.67	31.98	31.22	30.73	31.25	31.31
	St Dev'	0.83	0.48	0.30	0.86	0.17	0.59	0.63
H2afz	Ct'	20.91	20.89	20.29	20.47	19.04	20.90	19.93
	St Dev'	0.09	0.45	0.09	0.03	0.05	0.02	0.78

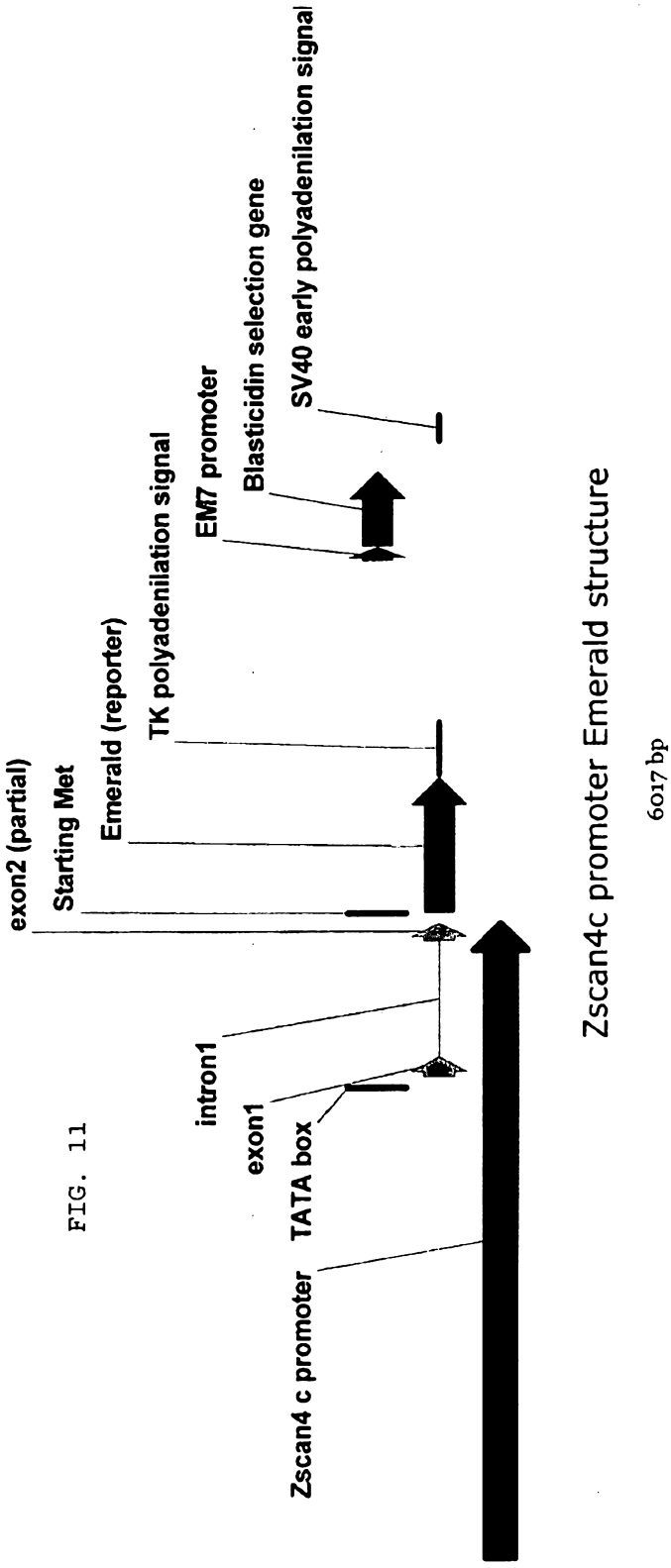

exon2 (partial)
Starting Met
Emerald (reporter)
TK polyadenilation signal
EM7 promoter
Blasticidin selection gene
SV40 early polyadenilation signal
intron1
exon1
TATA box
Zscan4 c promoter
Zscan4c promoter Emerald structure
6017 bp

FIG. 11

FIG. 12B

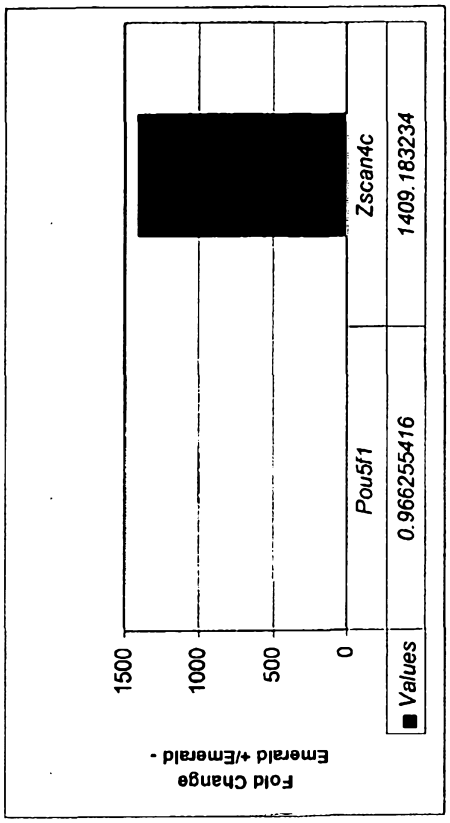
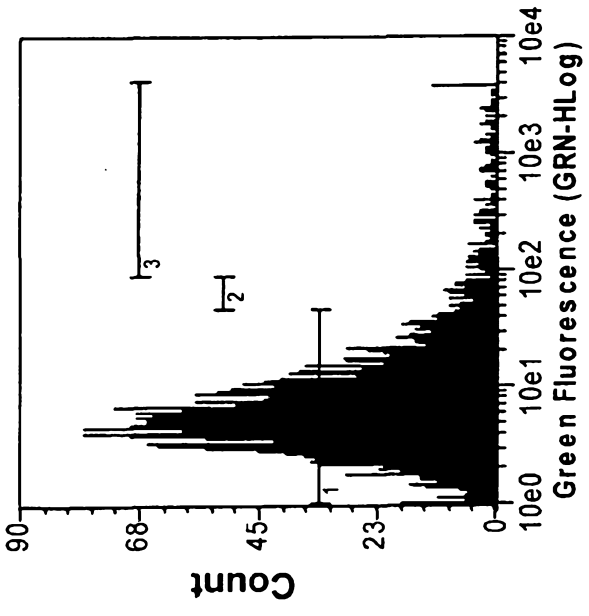


FIG. 12A



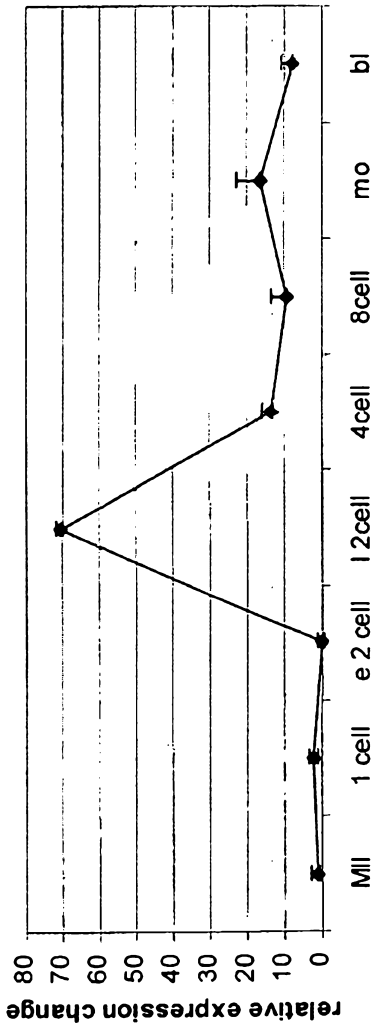


FIG. 13A

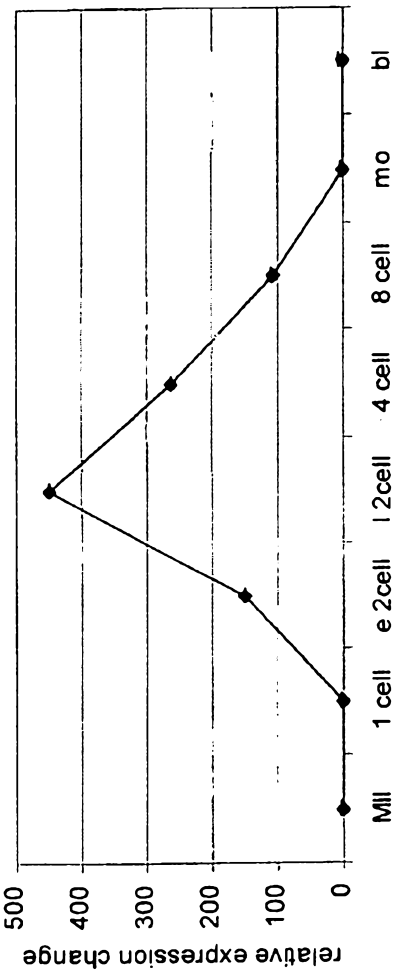


FIG. 13B

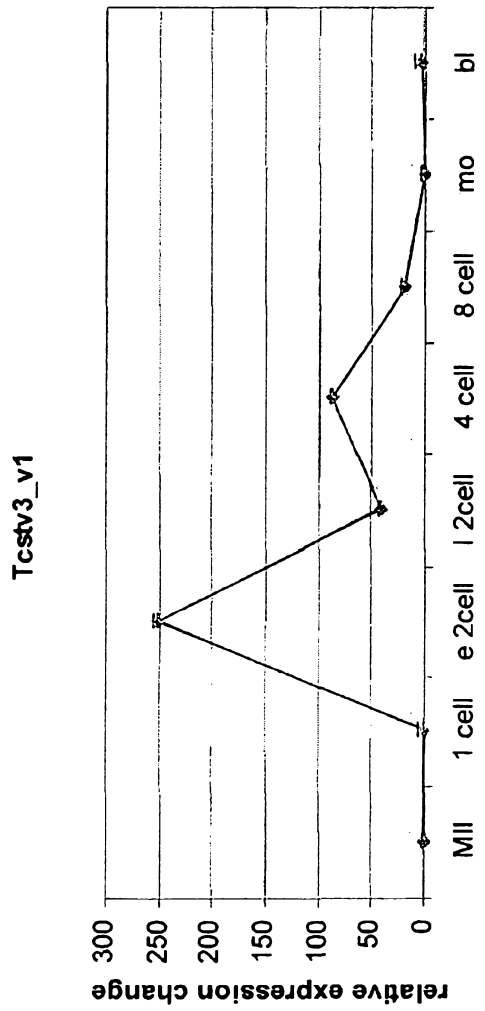


FIG. 13C

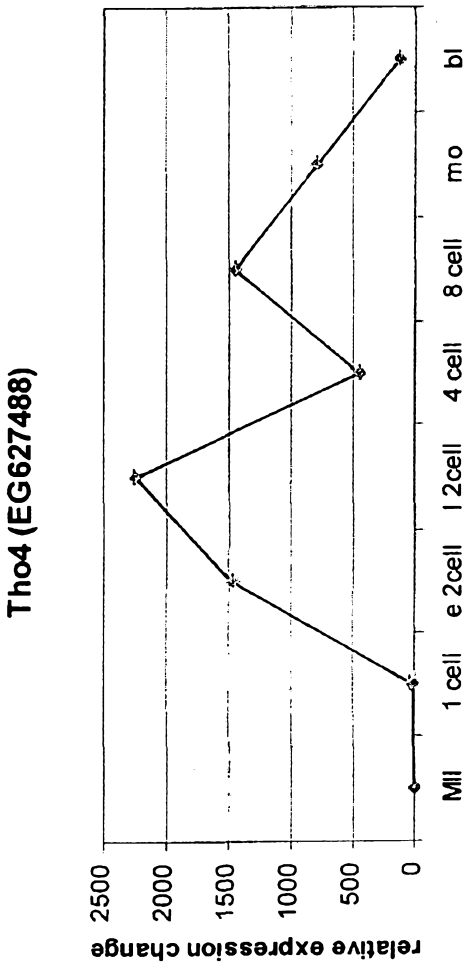


FIG. 13D

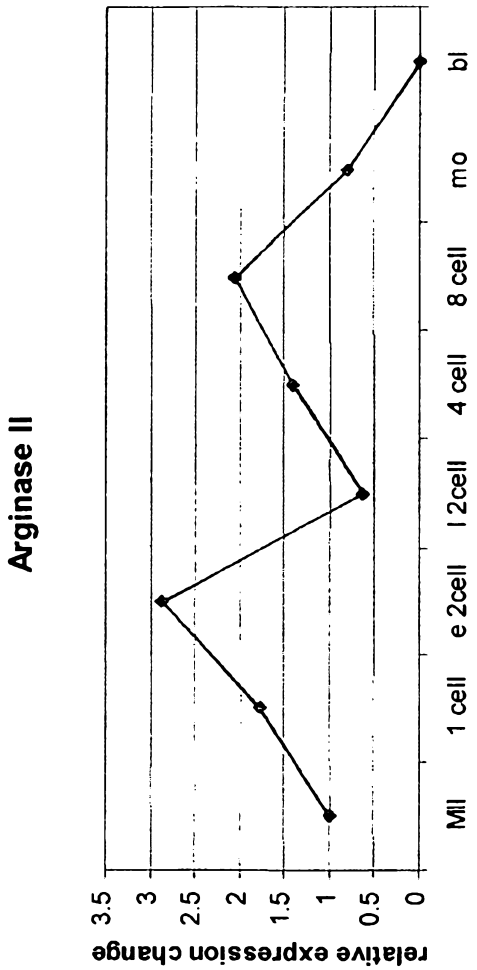


FIG. 13E

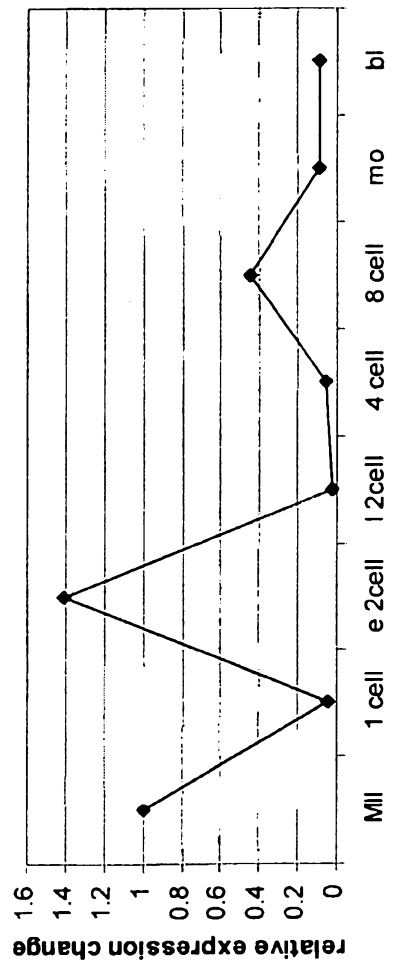


FIG. 13F

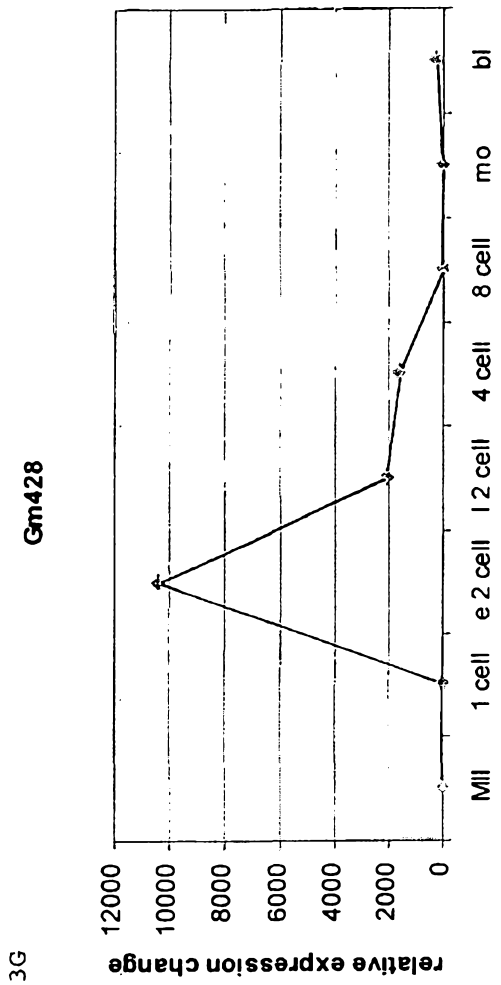

Gm428
relative expression change
12000
10000
8000
6000
4000
2000
0
MII
1 cell
e 2 cell
l 2 cell
4 cell
8 cell
mo
bl

FIG. 13G