Expression of NANOG and NANOGP8 in a variety of undifferentiated and differentiated human cells

SAKTHIKUMAR AMBADY[1], CHRISTOPHER MALCUIT[1,2], OLGA KASHPUR[1], DENIS KOLE[1],
WILLIAM F. HOLMES[1], EMMETT HEDBLOM[1], RAYMOND L. PAGE[1,2,3] and TANJA DOMINKO*[,1,2]

[1]Department of Biology and Biotechnology,[2]Bioengineering Institute and
[3]Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA, USA

ABSTRACT The transcription factor NANOG is essential for maintaining pluripotency in embryonic stem cells. We have previously reported the expression of NANOG in adult human fibroblasts; here we present a more thorough investigation into the expression of NANOG in a panel of both differentiated and undifferentiated human cells. We utilize RT-PCR, qRT-PCR, cloning and sequencing, sequence alignment, restriction digestion, immunocytochemistry, Western blotting, and EMSA to investigate expression of NANOG in a variety of somatic, transformed and stem cell phenotypes. RT-PCR and qRT-PCR analysis revealed the presence of NANOG transcripts in all the cell types examined, albeit at magnitudes lower than human embryonic stem cells. Further investigation by single nucleotide polymorphism analysis of expressed transcripts in several cell types detected a NANOG pseudogene, NANOGP8, one of only two NANOG pseudogenes with the potential of encoding a similar size protein to embryonic NANOG (eNANOG). Our analysis demonstrates that although the NANOG protein is detected in nearly all cells examined, expression of the eNANOG and/or NANOGP8 transcript as well as the sub-cellular localization of the protein is cell type-specific. Additionally, smooth muscle cells, which express exclusively NANOGP8, display nuclear localization of NANOG protein, indicating that NANOGP8 is a protein coding gene possibly functioning as a transcription factor. Lastly, all cell types expressing eNANOG and/or NANOGP8 were found to be capable of binding a NANOG consensus sequence in vitro. We conclude that eNANOG is not exclusively expressed in undifferentiated cells and that both eNANOG and NANOGP8 may function as transcription factors in a cell type-specific manner.

KEY WORDS: *expression, differentiated cell, human, NANOGP8, NANOG*

Introduction

NANOG is a 305 amino acid protein with a conserved homeodomain belonging to the homeobox gene family. It was first described as a key transcription regulator defining human embryonic stem cell (hESC) identity and self-renewal by both activating repressors of, and suppressing activators of differentiation (Mitsui *et al.*, 2003; Mullin *et al.*, 2008). Down-regulation of NANOG induces differentiation (Hatano *et al.*, 2005; Hyslop *et al.*, 2005), and over-expression of NANOG induces pluripotency (Silva *et al.*, 2006; Yu *et al.*, 2007). NANOG expression has been reported in embryonic stem cells (Chambers *et al.*, 2003), primordial germ cells (Yamaguchi *et al.*, 2005), fetal testis (Kerr *et al.*, 2008), human germ cell tumors (Hart *et al.*, 2005), seminomas and

breast carcinomas (Ezeh *et al.*, 2005), malignant cervical epithelial cells (Ye *et al.*, 2008), just recently in adult human fibroblasts (Page *et al.*, 2009) and in various mouse tissues *in vivo* (Hart *et al.*, 2004). NANOG contains an N-terminal domain, a DNA-binding homeodomain, and a C-terminal region consisting of two transactivation domains (Pan and Pei, 2003). NANOG maintains equilibrium between monomeric and active dimeric forms that helps maintain ES cell self-renewal (Mullin *et al.*, 2008; Wang *et*

***Address correspondence to:** Tanja Dominko. Department of Biology and Biotechnology,Worcester Polytechnic Institute, Worcester, MA, USA. Fax: +1-508-831-4120. e-mail: tdominko@wpi.edu

al., 2008) by regulating expression of downstream targets alone and in cooperation with OCT4 and SOX2 (Boyer *et al.*, 2005).

While NANOG pseudogenes have been described in cancer cells (Zhang *et al.*, 2006), it is believed that true embryonic NANOG (eNANOG) is expressed only in pluripotent cells where it is required for maintenance of cells' self renewal and pluripotency (Silva *et al.*, 2009). However, precise transcriptional analysis of NANOG is precluded due to the presence of 10 processed pseudogenes and one tandem duplication (Booth and Holland, 2004). Among the pseudogenes, NANOG P1, P2, P4, P7, P8, P9 and P10 exhibit more than 90% homology, and NANOGP5 exhibits about 85% homology to the parent gene in their open reading frames (ORFs; Table 2). Along with eNANOG, NANOGP1 (NANOG2) is expressed in human ES cells and encodes a protein of 232 amino acids (Hart *et al.*, 2004; Reubinoff *et al.*, 2000). NANOGP2, P4, P5, P9 and P10 contain premature stop codons with the potential for shorter proteins if translated in the same reading frame as eNANOG. NANOGP7 and P8 do not contain premature stop codons and have the potential to encode full-

A

B

Fig. 1. **Expression of NANOG. (A)** *Fold difference in the expression pattern of total NANOG (eNANOG and NANOGP8 combined) in different cell types compared to hESCs, expressed in logarithmic scale. HeLa cells and heart tissue show more than 10,000 fold lower levels of total NANOG transcript compared to hESCs.* **(B)** *The same data from (A) presented in a linear scale without hESCs to underscore varying levels of NANOG expression across non-ES cell types.*

TABLE 1

**HUMAN CELL TYPES/RNA USED IN
NANOG TRANSCRIPTION ANALYSIS**

Cell type	Source	Catalog number
Adult fibroblasts	ATCC	CRL-2352
Embryonic fibroblasts	ATCC	CRL-2097
HeLa cells	ATCC	CCL-2
SH-SY5Y Neuroblastoma cells	ATCC	CRL-2266
Smooth muscle cells (hSMCs)	Dr. B. Blackman, U. Virginia	NA
Umbilical vein endothelial cells (HUVECs)	Lonza	CC-2517
Mesenchymal stem cells (hMSCs)	Lonza	AA-2501
Total Heart (RNA)	ClonTech	636532
H9 Embryonic stem cells (hESCs)	WiCell	WA09
Human teratocarcinoma cells	ATCC	CRL-2073

length proteins. Based on GenBank sequences, NANOGP8 shares 99.5% homology to eNANOG to its putative ORF and can potentially code for a 305 amino acid protein that differs from the parent NANOG by only 3 amino acids (Table 2). Expression of NANOGP8 transcripts has been detected in various cancer cell lines and tumors, and over-expressed recombinant NANOGP8 under the control of a ubiquitous promoter exhibited nuclear localization and demonstrated an ability to promote entry of cells into S-phase (Zhang *et al.*, 2006) suggesting a possible role for NANOGP8 in cellular function.

Previously we reported the presence of NANOG transcripts in an adult human fibroblast culture (Page *et al.*, 2009). The purpose of the current study was to identify and further characterize the NANOG transcripts and the sub-cellular localization of translated products in a robust panel of cell types. We compared these characteristics between human primary differentiated cells, transformed cell lines, and embryonic stem cells. Here, for the first time, we report expression of both eNANOG and NANOGP8 transcript and protein in differentiated human cell types. In addition, we report DNA binding of NANOGP8 and present distinct cellular localization patterns between NANOGP8 and eNANOG. The implications of expressed NANOGP8 pseudogene in differentiated cells compared to human embryonic stem cells (hESCs) are discussed.

Results

Identification of NANOG transcripts in both differentiated and undifferentiated human cells

Based upon our previously published work in which we detected the expression of NANOG in a primary adult human fibroblast culture (Page *et al.*, 2009) we sought to examine NANOG expression across a panel of differentiated and undifferentiated cell types: human adult fibroblasts, human neonatal fibroblasts, HeLa cells, SH-SY5Y neuroblastoma cells, human smooth muscle cells (hSMCs), human vascular endothelial cells (HUVECs), mesenchymal stem cells (hMSCs), total heart tissue, and H9 human embryonic stem cells (hESCs) (Table 1).

We analyzed the presence of NANOG transcript by quantitative PCR (qPCR, primers listed in Table 3). The $\Delta\Delta$Ct method (Liss, 2002) was used to calculate fold change in expression levels of NANOG between cell types after normalizing to the control hESCs and represented on a logarithmic scale (Fig. 1A). Human ESCs showed the highest expression for NANOG, with all other cell types
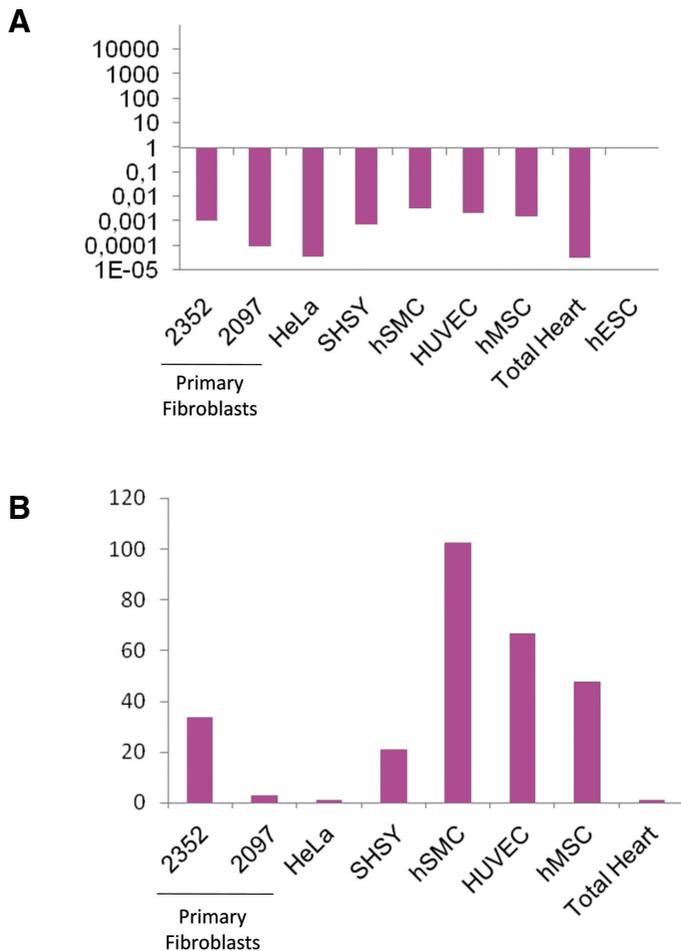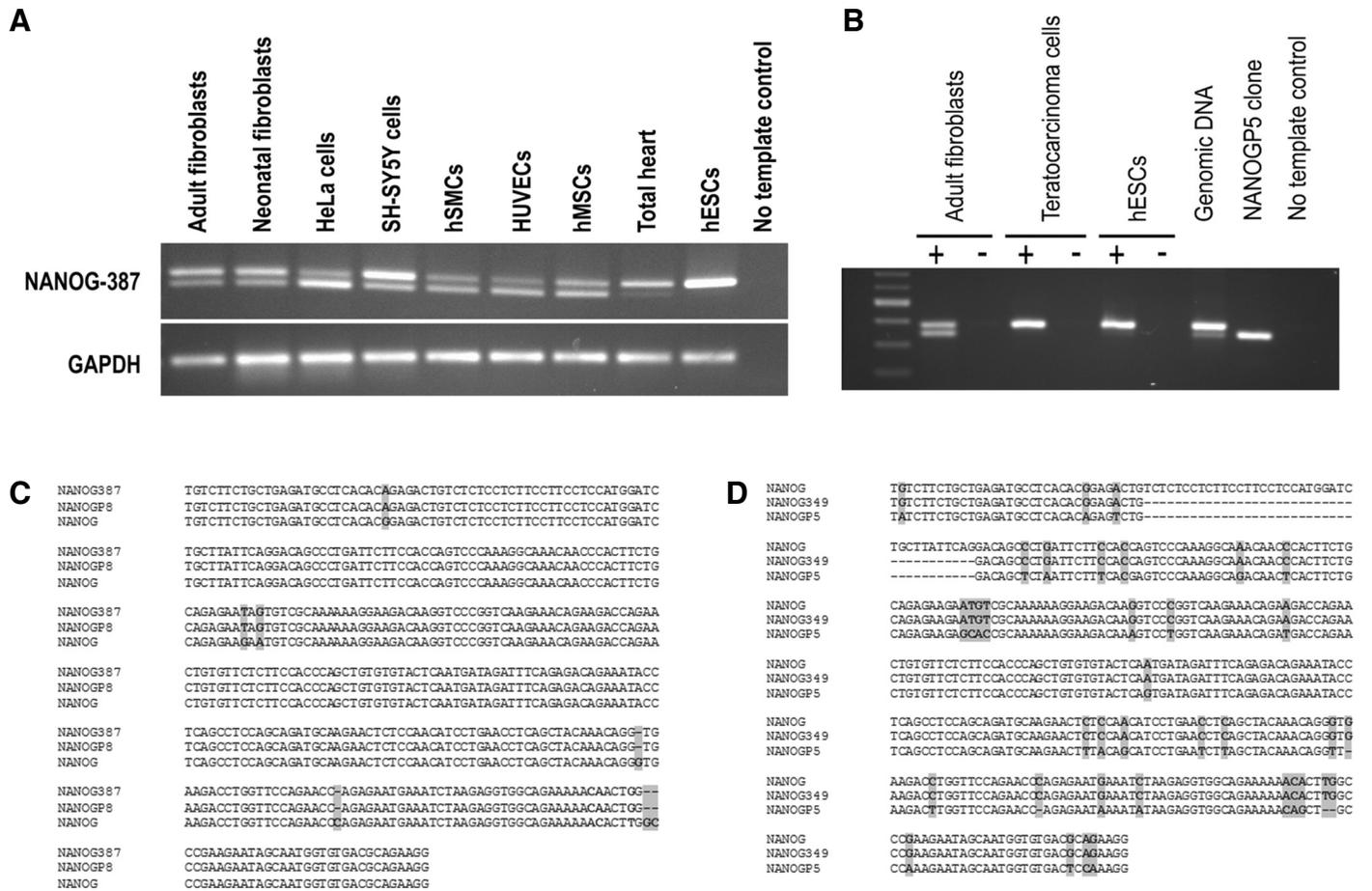
A

B

C

D

Fig. 2. Differential expression of NANOG in different cell types. (A) *RT-PCR. Note the differential amplification intensity of the two bands for NANOG within and between cell lines. Heart tissue sample shows a very faint lower band, while only the upper band is visible in hESCs. See Table 1 for details on cell types.* **(B)** *RT-PCR using Nanog-387 primers from adult human fibroblasts, human teratocarcinoma cells, human ES cells, human genomic DNA, and a plasmid clone containing 1260 bp NANOGP5 genomic fragment. Symbols + and – indicate presence or absence respectively, of reverse transcriptase (RT) in the reactions. Adult fibroblasts show the 387 bp and 349 bp amplicons. Teratocarcinoma and ES cells amplify only the 387 bp fragment. The primers amplify a predominant 387 bp fragment and a very weak 349 bp fragment from genomic DNA indicating relative abundance of the respective pseudogenes in genomic DNA. Control plasmid containing genomic NANOGP5 amplifies only the 349 bp fragment. Relevant band sizes in base pairs are indicated.* **(C)** *DNA alignment between 387 bp PCR fragment (upper band) amplified from fibroblasts and its corresponding region from eNANOG (NM_024865) and NANOGP8 (NC_000015). Note that the 387 bp fragment is 100% homologous to NANOGP8 (shaded regions).* **(D)** *DNA alignment between 349 bp PCR fragment (lower band) amplified from adult fibroblasts and its corresponding region from eNANOG (NM_024865) and NANOGP5 (NG_004101). The amplified fragment shows the same 38 bp deletion as in NANOGP5. The sequence shows 100% identity to eNANOG except for the 38 bp deletion (shaded regions).*

expressing approximately 500 to >10,000 fold less NANOG transcript. Among the non-ES cell types, hSMCs, had the highest transcript level (at 500 fold less than hESC); followed by HUVECs and hMSCs (about 750 fold less); adult human fibroblasts and SH-SY5Y cells (about 1000 fold less); neonatal fibroblasts (10,000 fold less); followed by HeLa cells and total heart (>10,000 fold less) (Fig. 1A).

To more closely analyze the differences between non-hESC samples, we reconfigured the dataset without hESC and present the same results normalized to the lowest sample, total heart (Fig. 1B). Quite interestingly, human smooth muscle cells exhibit close to 100-times the NANOG transcript expression in comparison to even HeLa cells, an immortalized cancer cell line. hMSCs show nearly 45-times the NANOG levels of HeLa cells, and the human adult dermal fibroblast cell line CRL-2352, in which we previously

reported NANOG expression, displays approximately 30-fold higher expression than HeLa. This data extends our previously published observations (Page *et al.*, 2009) to other cell types and demonstrates that human NANOG expression, although at lower levels of magnitude, is not restricted to hESCs.

Identification of NANOG and NANOGP8

To further refine which NANOG transcripts were represented in these preliminary findings we screened the panel of various cell types presented in Table 1. Reverse Transcriptase polymerase chain reaction (RT-PCR) analysis using Nanog-387 primers (Table 3) in adult human fibroblasts amplified two bands, one at the expected 387 bp and an additional lower band. The expected 387 bp fragment was amplified at varying intensities from all of the cell types tested and exclusively in hESCs. The lower 349 bp fragment

was amplified from most of the cell types excluding hESCs (Fig. 2A).

We addressed the possibility of genomic contamination of the RT-PCR reactions using adult fibroblasts, human teratocarcinoma cells and H9 human embryonic stem cells. Adult fibroblasts amplified two bands while the human teratocarcinoma cells and H9 human embryonic stem cells amplified only a single band at the expected 387 bp size, representing the parent gene, in the presence of reverse transcriptase. The primers amplified two bands from human genomic DNA, a strong upper band representing amplifications from multiple NANOG pseudogenes (not the parent gene) and a weak lower band representing a single copy of NANOGP5 in the genome. This was confirmed when the same primers amplified only a 349 bp fragment from a control plasmid clone containing NANOGP5 insert derived by PCR from human genomic DNA (Fig. 2B). The two bands from fibroblasts and the single band from teratocarcinoma cells were isolated separately,

cloned and sequenced to confirm their identities. BLAST analysis revealed that the 387 bp upper fragment (Nanog-387) from adult fibroblasts contained clones that were 100% identical to NANOGP8 (Fig. 2C). The 387 bp fragment from teratocarcinoma cells exhibited 100% homology to parent NANOG (not shown). The lower band from fibroblasts had a 38 bp deletion resulting in a fragment size of 349 bp (Nanog-349). Alignment of Nanog-349 sequence with NANOG and all of its pseudogenes revealed that only NANOGP5 exhibited the same 38bp deletion. Nanog-349 showed 100% homology to eNANOG except for the 38 bp deletion, while its overall homology to NANOGP5 was only 89.5% (Fig. 2D). Analysis is underway to determine whether Nanog-349 represents an unidentified isoform or pseudogene of NANOG with a similar deletion as NANOGP5.

Differentiating between eNANOG and NANOGP8 transcripts

Since Nanog-387 primers have a high degree of homology to

TABLE 2

HOMOLOGY BETWEEN NANOG AND SELECTED NANOG PSEUDOGENES IN THEIR ORFS

Gene	ORF (bp)	Premature stop codon position	Protein size (aa)	Percent homology								
				Nanog	P1	P2	P4	P5	P7	P8	P9	P10
NANOG	**918**	**None**	**305**	**100**	-	-	-	-	-	-	-	-
P1	699	None	232	92.4	100	-	-	-	-	-	-	-
P2	897	126	42	91.5	85.9	100	-	-	-	-	-	-
P4	915	66	22	95.2	88.5	89.3	100	-	-	-	-	-
P5	**879**	**159**	**53**	**85.4**	83.0	82.0	82.6	100	-	-	-	-
P7	918	None	305	94.8	89.1	89.2	92.2	83.5	100	-	-	-
P8	**918**	**None**	**305**	**99.5**	92.1	91.5	95.1	85.3	94.4	100	-	-
P9	895	447	149	91.5	84.9	88.3	88.9	81.8	90.8	91.4	100	-
P10	903	456	152	91.8	85.9	87.7	88.2	82.9	90.3	91.7	90.8	100

Protein size reflects the number of amino acids in the protein resulting from translation occurring in the same reading frame as parent NANOG. Homologies for NANOG, NANOGP5 and NANOGP8 are highlighted. Homology was determined using the LALIGN program.

TABLE 3

RT-PCR, Q-PCR, SEQUENCING PRIMERS AND EMSA PROBES

Primer Name	PCR type	Forward sequence (5' to 3')	Reverse sequence (5' to 3')	Amplicon size (bp)	$T^o{}_A$	Extension (min)	GenBank accession
Nanog-387	RT-PCR	TGTCTTCTGCTGAGATGCCTCACA	CCTTCTGCGTCACACCATTGCTAT	387	60	1:30	NM_024865
Nanog-1860	RT-PCR	TGAGTGTGGATCCAGCTTGT	TCATCGAAACACTCGGTGAA	1860	53	2:00	NM_024865
NanogP5	Genomic	GCTCATATTCCTCTGACATCTTCC	GCCAGAGACAGCTTCTATCATGTT	1260	57	2:00	NG_004101
GAPDH	RT-PCR	ATCACCATCTTCCAGGAGCGA	TTCTCCATGGTGGTGAAGACG	101	53	1:00	NM_002046
β-actin	RT-PCR	TCTGGCACCACACCTTCTACAA	CTTCTCCTTAATGTCACGCACG	391	58	1:30	NM_001101
Nanog	qPCR	GCTGGTTGCCTCATGTTATTATGC	CCATGGAGGAAGGAAGAGGAGAGA	345	55	1:00	NM_024865
β-actin	qPCR	AGAGCTACGAGCTGCCTGAC	GGATGCCACAGGACTCCA	111	60	1:00	NM_001101
Nanog-2031	RT-PCR	GCTGGTTGCCTCATGTTATTATGC	TCATCGAAACACTCGGTGAA	2031	53	2:30	NM_024865 NC_000015
Nanog-ORF	PCR	GCTGGTTGCCTCATGTTATTATGC	GC**GATATC**CACGTCTTCAGGTTGCATGTTCAT	1093	55	1:30	NM_024865 NC_000015
Nanog-F	Sequencing	CCTCCAGCAGATGCAAGAAC					NM_024865 NC_000015
Nanog-R	Sequencing	CGCTGATTAGGCTCCAACCATACT					NM_024865 NC_000015
EMSA-sense strand		GTGTGATATCGAATTCCGCCTCGGTACC**CTTCGCCG** **ATTAA**GTACTTAAGGTACC_GAGGCGGAATTCGATAT CACAC_		77	NA	NA	NA
EMSA complementary strand		GTGTGATATCGAATTCCGCCTCGGTACC**TTAAGTAC** **TTAAT**CGGCGAAGGGTACC_GAGGCGGAATTCGATA TCACAC_		77	NA	NA	NA
EMSA primer	PCR	**Biotin**-GTGTGATATCGAATTCCGCCTC		77	57	1:00	NA

The 21 base NANOG binding sequence for EMSA probe is shown in bold and the core binding region (**ATTAA** and **TTAAT** on sense and complementary strands respectively) are depicted in bold red. PCR priming sequences flanking the binding sequence are underlined. The oligonucleotides contain two Kpn I sites (GGTACC) for restriction digestion to test the efficiency of oligonucleotide annealing.

most of NANOG pseudogenes, primers specific for eNANOG were designed to discriminate eNANOG and NANOGP8 from the rest of the known pseudogenes (Nanog-1860, Table 3, Fig. 3A). The 1860 bp amplicon encompassing the entire open reading frame (ORF) and most of the 3' UTR from eNANOG and NANOGP8 contains a restriction enzyme site polymorphism for Sma I allowing distinction between eNANOG and NANOGP8. The Sma I restriction site is present in eNANOG 3' UTR but absent in NANOGP8 due to a C to G conversion (Fig. 3B). The 1860 bp amplicon from eNANOG can be digested into two fragments of 1236 bp and 624 bp while the NANOGP8 amplicon remains resistant to Sma I digestion. Sma I digested NANOG1860 PCR products from all cell types showed variable expression of eNANOG and NANOGP8 when resolved on 2% agarose gels (Fig. 3C). Human ES cells, neonatal fibroblasts, HeLa cells, SH-SY5Y cells and human mesenchymal stem cells (hMSCs) expressed the parent (eNANOG) transcript exclusively, whereas adult fibroblasts, HUVECs and total heart tissue contained both transcripts in varying proportions, and human smooth muscle cells (hSMCs) expressed only NANOGP8 (Fig. 3B).

Sequencing full-length NANOG transcripts in adult fibroblasts

To confirm that the 1860bp amplicon, above, encompassed the ORF for either eNANOG or NANOGP8 we cloned this amplicon into the pCR 4 TOPO vector (Invitrogen). Twenty five random clones were cultivated and resulting plasmid DNA was sequenced from both 5' and 3' ends using T7 and T3 primers. To ensure a full contiguous read, primers internal to the ORF of eNANOG/NANOGP8 were also used for sequencing (Table 3). All recombinant clones contained inserts that showed at least 97% homology to eNANOG and/or NANOGP8. Sequence derived from control embryonal carcinoma cells showed 100% homology to eNANOG (not shown). None of the sequenced clones exhibited splice variants. eNANOG and NANOGP8 differ at 5 positions in their ORF (P47, P144, P246,
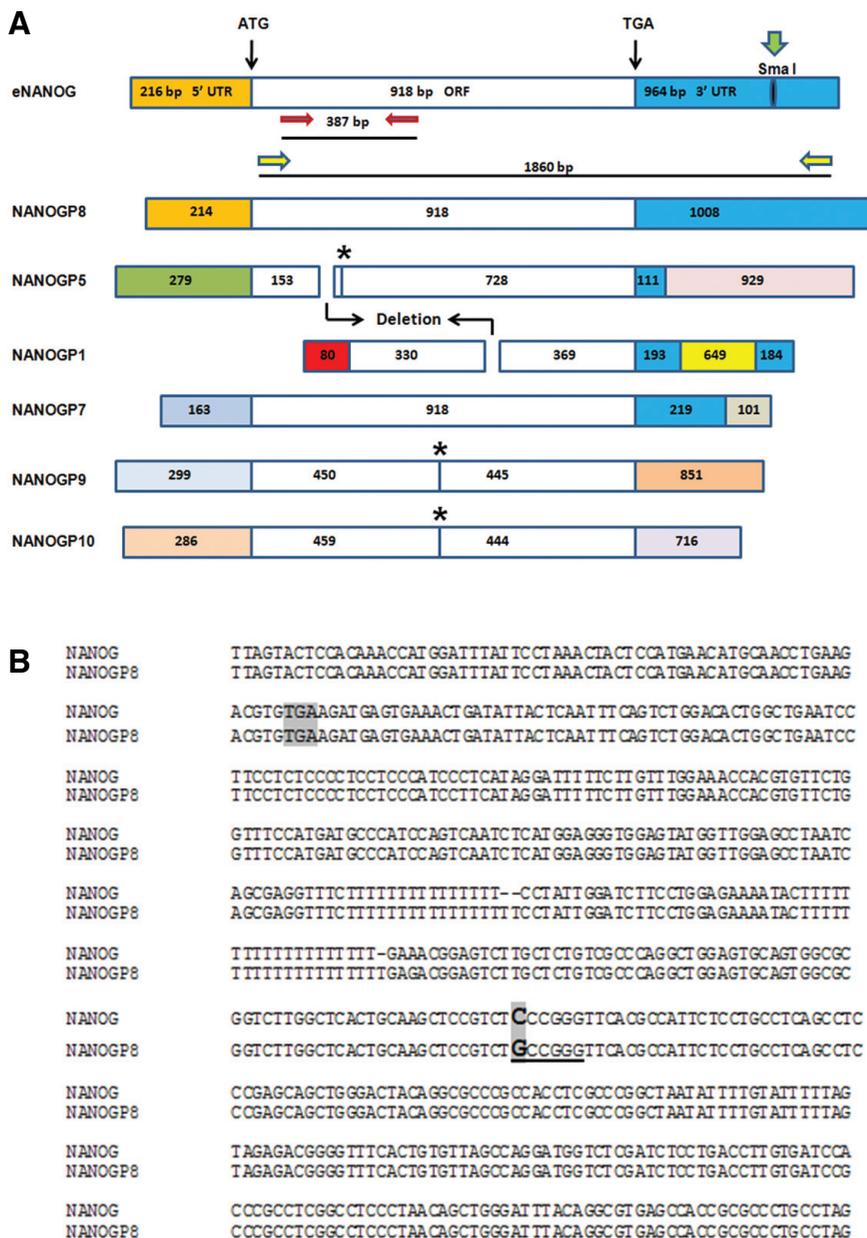

Fig. 3. **Characterization of NANOG and NANOGP8. (A)** *Cartoon of all documented NANOG isoforms and pseudogenes. Regions of homology are indicated as well as all primer locations used for isotype-specific analysis. Open areas represent possible open reading frames and asterisks denote premature stop codons. The Sma I restriction site lost by the SNP in NANOGP8 is indicated in eNANOG, top. Sequences are not to scale.* **(B)** *Alignment of 3' UTR regions (partial sequence) of eNANOG (NM_024865) and NANOGP8 (NC_000015). The stop codon TGA (shaded) is shown. The restriction site for Sma I (CCCGGG) in eNANOG is absent in NANOGP8 due to a C to G transversion (shaded region).* **(C)** *Sma I digested restriction pattern of NANOG1860. Sma I digests the 1860 bp amplicon derived from eNANOG into 1236 bp and 624 bp sized fragments while the amplicon derived from NANOGP8 is resistant to Sma I digestion.*
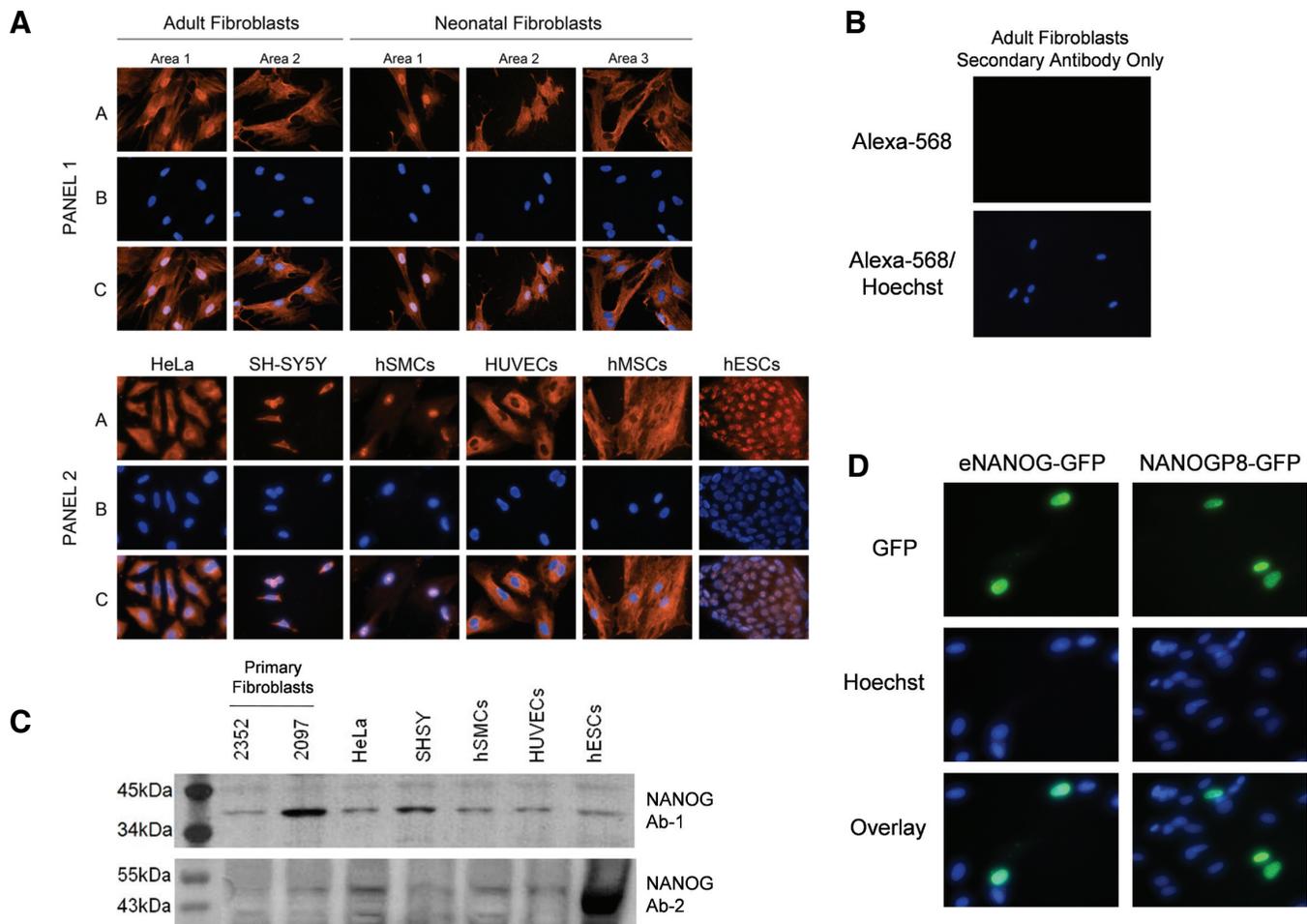
Fig. 4. **NANOG/NANOGP8 native protein expression and localization. (A)** *Adult and neonatal fibroblasts exhibit differential localization of NANOG/ NANOGP8 in subpopulation of cells in different regions of the culture dish (Panel 1). Note the strong nuclear localization (Area 1) and uniform cytoplasmic and nuclear localization (Area 2) in both adult and neonatal fibroblasts. Exclusion of protein from the nucleus (Area 3) is observed only in neonatal fibroblasts. Other cell types are shown in Panel 2. Rows A, B and C represent NANOG/NANOGP8 Alexafluor-568 (red), Hoechst stained nuclei (blue) and merged images respectively. All images were acquired at 400X magnification. **(B)** Immunofluorescence controls for antibody specificity. Top: Image depicts adult fibroblasts labeled only with secondary antibody used in A, above (Alexa-568 conjugated). Bottom: Image shows double staining of secondary antibody alone and Hoechst labeling of nuclei. **(C)** Western blot analysis in various cell types using two commercially available antibodies (Abnova, Ab-1, top; Abcam, Ab-2, bottom) raised against NANOG. A protein band of approximately 37 kDa corresponding to the predicted size of NANOG is detected in all cell types in addition to a larger, ~48 kDa band. **(D)** Overexpression of NANOG/NANOGP8-GFP. Neonatal fibroblasts were transfected with a plasmid containing the CMV promoter driving either eNANOG-eGFP or NANOGP8-eGFP. Live cells were labeled with Hoechst 33342 and imaged with FITC/GFP fluorescence and DAP/Hoechst fluorescence. Images were acquired at 400X magnification.*

P531 and P759) relative to the ATG start site (Supplementary Fig. S1). This information formed the basis for comparison of NANOG transcripts sequenced from adult fibroblasts and hSMCs in this study. Fibroblast sequences showed higher homology to NANOGP8 than to eNANOG. Additionally, fibroblasts contained a unique A to G point mutation at two positions (P368 and P488, Supplementary Fig. S2). The single clone sequenced from hSMCs exhibited very high homology to NANOGP8 except for a single base change at P47 where it carried the same cytosine base as eNANOG (Supplementary Fig. S3).

Detection and sub-cellular localization of eNANOG and NANOGP8 protein

Immunocytochemistry (ICC) was next conducted using a monoclonal antibody raised against NANOG to determine whether

eNANOG/NANOG P8 were localized properly to the nucleus. Expression of eNANOG/NANOGP8 protein was detected in all cells but showed differential cellular localization (Fig. 4A). Among cell types that exclusively expressed eNANOG transcript, hESCs displayed nuclear localization, while HeLa cells and SH-SY5Y cells showed predominantly (although not exclusively) cytoplasmic localization, and the majority of hMSCs appeared to exclude NANOG protein from the nucleus entirely. Neonatal fibroblasts that expressed only eNANOG transcript exhibited protein localization in either, nucleus, cytoplasm, or uniformly in both compartments in discrete sub-populations of cells (Fig. 4A, Panel 1, right). In cell types expressing both transcripts (HUVECs and adult fibroblasts), distinctly different localization of the proteins was observed. A majority of HUVECs showed nuclear exclusion while subpopulations of adult fibroblasts exhibited either predominantly

nuclear or uniform localization in nucleus and cytoplasm (Fig. 4A, Panel 1, left). Human smooth muscle cells that expressed only NANOGP8 transcript exhibited predominant nuclear localization of the protein, indicating that this previously classified pseudogene may play a functional role in certain cell types, and suggests a potential functional role for this pseudogene. The specificity of these ICC results was confirmed using the secondary antibody and did not exhibit non-specific labeling (Fig. 4B).

We next carried out Western blotting on the majority of cell types analyzed in the prior assays to further characterize specificity of the antibody used for ICC experiments. Small cell pellets from above experiments were used for protein samples, and thus protein concentrations were estimated and not quantitatively calculated. Using a mouse monoclonal antibody raised against NANOG (Abnova, Fig. 4C, top) as well as a rabbit polyclonal antibody raised against NANOG (Abcam, Fig. 4C, bottom) we detected NANOG protein in all the cell types with immunoreactive bands migrating at approximately 37 kDa and ~48 kDa (Fig. 4C). While 37 kDa is the reported molecular weight for NANOG resolved by SDS-PAGE, the unexpected 48 kDa band may be indicative of a post-translational modification such as sumoylation. Most interesting in these experiments is the detection of immunoreactive NANOG protein in hSMCs, a cell type that only detectably expresses the NANOGP8 transcript, indicating that NANOGP8 is actively translated into protein.

To confirm the validity of these immunolocalization results, we next transfected cells with a plasmid containing the CMV promoter driving expression of either eNANOG or NANOGP8, both of which were fused to eGFP. Overexpression of eNANOG and NANOGP8, respectively, in human neonatal fibroblasts resulted in active protein translocation to the nucleus as determined by fluorescence microscopy (Fig. 4D). These data are in accord with our previously published work and strengthen the data in this manuscript indicating that NANOGP8 is indeed, an actively produced gene product.

NANOG and NANOGP8 bind to its cognate DNA consensus sequence

Since the peptide sequence corresponding to the homeodomain of NANOGP8 is 100% homologous to that of eNANOG (Fig. 5A), we next sought to determine whether the NANOG protein expressed by a panel of differentiated and undifferentiated cells can bind to a consensus NANOG binding DNA motif (Mitsui *et al.*, 2003). Accordingly, we performed electrophoretic mobility shift assay (EMSA) with nuclear extracts from human teratocarcinoma cells and hESCs (positive controls), HeLa cells, neonatal fibroblasts, hSMCs, adult fibroblasts and *Drosophila* S3 cells (negative control). Nuclear extracts from all human cells produced a perceptible shift of the biotin labeled EMSA probe, while the *Drosophila* S3 cell nuclear extract, which does not contain an ortholog of NANOG, did not show probe binding (Fig. 5B). These results demonstrate for the first time that human somatic cells, even those expressing exclusively NANOGP8, produce a functional protein capable of binding one of NANOG's target sequences.

Differentiated hESCs do not express NANOG

In order to compare eNANOG/NANOGP8 expression in differentiated cells to eNANOG expression in differentiated hESCs, RT-PCR was performed on undifferentiated hESCs and hESCs taken through four passages of differentiation over a period of
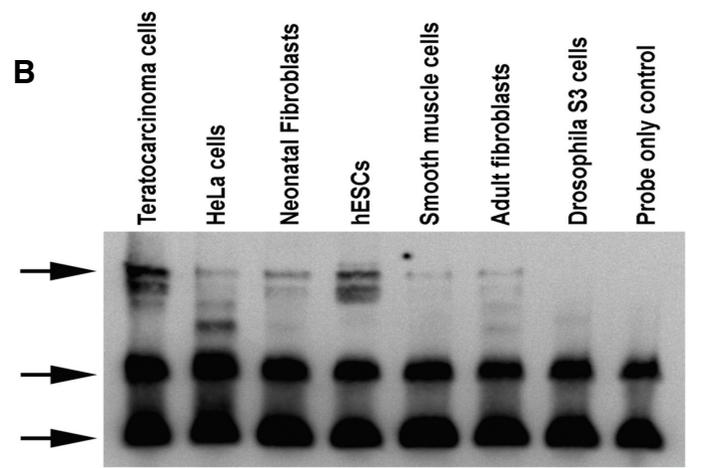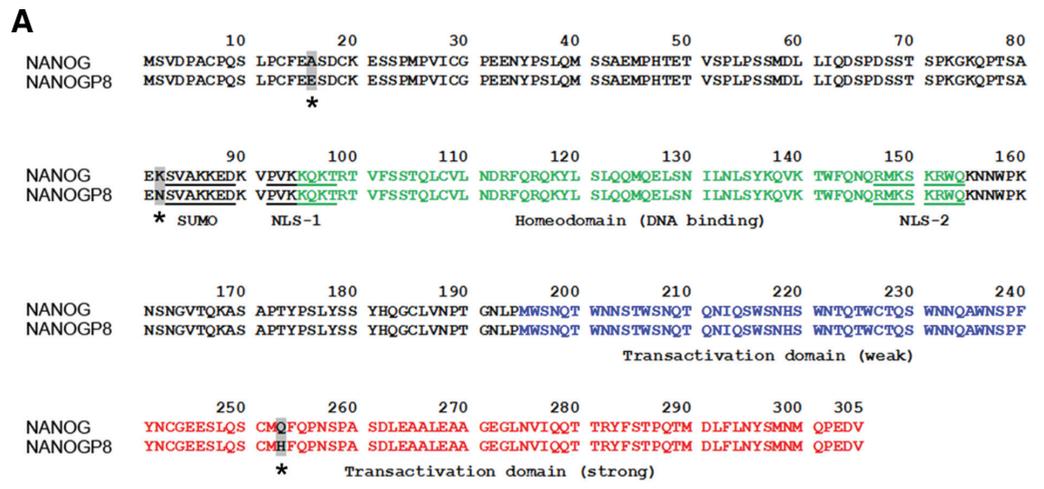

Fig. 5. NANOG and NANOGP8 binding to DNA. (A) *The three amino acid (aa) changes are denoted by the stars. Two aa changes are in the N-terminal region (1-94) at positions 16 and 82, and the third one in the second transactivation domain at position 253 in the second transactivation domain. Note the aa change from lysine to asparagine at position 82 adjacent to the predicted sumoylation site (SVAKKED). The two nuclear localization signals NLS-1 (PVKKQKT) and NLS-2 (RMKSKRWQ) are indicated (underlined). The homeodomain DNA binding region (95-154) and the two transactivation domains (155-240 and 241-305) are also indicated.* **(B)** *Electrophoretic Mobility Shift Assay (EMSA) using nuclear extracts from 7 different cell lines. The bottom and middle arrows indicate the unincorporated biotin labeled PCR primer and double stranded EMSA probe respectively. The top arrow indicates the shifted probe.*

about 40 days. The first two passages were performed by scraping the ES clumps using a sterile cell scraper and plating on gelatinized tissue culture plates, while the last two passages were done by trypsinization. Unexpectedly, NANOG transcripts were significantly down regulated by day 7 (passage 1) and disappeared completely by passage 2 and beyond (Supplementary Fig. S4). In consideration of the data presented above, this suggests that NANOG downregulation upon hESC differentiation may be a phenomenon unique to this *in vitro* process.

Discussion

It is becoming more evident that expression of pluripotent genes can be detected in a variety of cell types and at various stages of differentiation. The human genome contains approximately 21,000 protein coding genes (Clamp *et al.*, 2007) and approximately the same number (~19,000) of pseudogenes (Torrents *et al.*, 2003). Pseudogenes themselves may be expressed at higher levels than previously thought and may have a major role in controlling the expression of their respective parent genes (Chen *et al.*, 2004). Conversely, pseudogenes whose protein products exhibit high homology to their parent counterpart may complement the function of the parent gene. However, the fact that at least two of these genes, namely OCT4 and NANOG have a number of pseudogenes, complicates the interpretation of such observations. High sequence homology between parent genes and pseudogenes inhibits routine procedures for identification at the level of mRNA. The minimal differences in amino acid sequence results in antibodies that cannot distinguish between parent and pseudogenes. Nevertheless, the results described herein show that both eNANOG and NANOGP8 are transcribed and translated in a variety of human differentiated and undifferentiated cells.

Based on GenBank sequences, comparison of eNANOG and NANOGP8 shows 5 base pair changes at positions P47, P144, P246, P531 and P759 relative to the ATG start site (Supplementary Fig. S1) with a resultant 3 amino acid change at the protein level (Fig. 5A). Because of high degree of homology among eNANOG and many of its pseudogenes, sequencing the transcripts is the only way to discriminate NANOG isoforms. We had determined, based on Sma I restriction analysis, that adult fibroblasts expressed higher level of NANOGP8 relative to eNANOG while hSMCs expressed only NANOGP8 (Fig. 3). However, the high homology of eNANOG and its pseudogenes was a concern. Sequencing was therefore undertaken to determine whether the amplified products from fibroblasts represented other unknown isoforms of eNANOG. All sequences showed a higher degree of homology to NANOGP8 than to eNANOG. Of the 5 base changes between published sequences of eNANOG and NANOGP8, adult fibroblast derived clones exhibited NANOGP8 specific bases at 4 out of 5 positions (Supplementary Fig. S2). Additionally, all the clones showed two base changes at positions P368 and P488 that were unique to the fibroblast cell line used. These changes do not introduce premature stop codons. Therefore, the NANOG transcripts present in adult fibroblasts are capable of producing full-length proteins. Their ability to bind NANOG cognate sequences was verified by EMSA (Fig. 5).

One of the major chromatin modifications regulating gene expression is the methylation status of the CpG islands in gene promoters (Klose and Bird, 2006). CpG methylated promoters are generally associated with silenced genes, and interestingly there are conflicting reports on the methylation status of NANOG promoter in human fibroblasts. Some laboratories have observed high levels of NANOG promoter methylation in both neonatal (Yu *et al.*, 2009) and adult dermal fibroblasts (Park *et al.*, 2008; Takahashi *et al.*, 2007), while virtually no methylation of NANOG promoter has been reported by others (Barrand and Collas, 2009). Despite promoter demethylation, no NANOG transcript was detected in these latest studies and the authors suggest that hypomethylated state may constitute a permissive state enabling gene reactivation (Barrand and Collas, 2009). Alternatively, distinct subpopulations of fibroblasts, such as those expressing SSEA-3 (Byrne *et al.*, 2009) could be responsible for the bulk of the transcript. The subpopulations of adult fibroblasts with distinct localization of NANOG protein observed in the present study may similarly be differentially contributing to the majority of observed transcript. In addition, detection of the transcript may depend on the sensitivity of amplification conditions. Exclusive expression of eNANOG in neonatal fibroblasts and a transition to a combination of eNANOG and NANOGP8 in adult fibroblasts may indicate that either NANOGP8 expression results in differentiation or that its expression is the result of differentiation. The mechanisms of transcriptional regulation of NANOGP8, however, remain unknown. Our data also suggest that rapidly dividing cells (neonatal fibroblasts, HeLa, SH-SY5Y, hMSCs and hESCs) express eNANOG, while relatively slow dividing cells (adult fibroblasts, HUVECs, heart tissue) express both eNANOG and NANOGP8. Lastly, terminally differentiated smooth muscle cells (the slowest cycling of all) expressed only NANOGP8. It is not clear whether the dynamics of the cell cycle and differentiation state of the cells are correlated and further studies are needed to elucidate this relationship.

The presence of pseudogenes can potentially compromise RT-PCR and qPCR results (Liedtke *et al.*, 2007; Liedtke *et al.*, 2008a, Pain *et al.*, 2005). Complications arise due to either the failure to adequately eliminate genomic DNA contamination from RNA preparations or due to the presence of genuine pseudogene transcripts. All our experiments were carried out on RNA preparations treated with RNase free DNase. Additionally, first strand cDNA synthesis for qPCR involves an additional DNase step. Proper negative controls including RT-minus and no template controls in qPCR ruled out genomic contamination and primer self amplification, respectively. The commonly reported lack of NANOG detection in somatic cell types does not necessarily indicate the absence of the transcript. Many factors such as rarity of the message, enzyme sensitivity/processivity and buffer conditions can contribute to the differences in amplification efficiencies of polymerases. We observed that 4 out of 8 different polymerases tested failed to amplify the expected product in fibroblasts with Nanog-387 primers. All 8 polymerases amplified the control GAPDH and β-actin primers with equal efficiency (data not shown). However, Ex-Taq polymerase (used in all the experiments), rTaq polymerase and AccuStart Taq DNA polymerase showed the highest amplification efficiency.

Pseudogenes have been associated with several other stem-cell associated genes, such as OCT4 (Panagopoulos *et al.*, 2008; Suo *et al.*, 2005), and DPPA3 (Elliman *et al.*, 2006), and are expressed in various cancer cell lines. Recent studies that have

reported basal expression of not only NANOG, but also other ESC-specific genes OCT4 (Liedtke *et al.*, 2008b), SOX2 (Que *et al.*, 2009), LIN28 (Polesskaya *et al.*, 2007) and REX1 (Page *et al.*, 2009) in differentiated cells, call for their in depth characterization. While functionality of these transcription factors and/or their pseudogene-coded counterparts in differentiated cells remains unclear, it is unlikely that their presence alone assures a stem cell phenotype.

eNANOG and NANOGP8 proteins differ by three amino acids based on their published sequence (Fig. 5A). Two of the three amino acid changes occur in the N-terminal region (alanine to glutamic acid at positions 16; lysine to asparagine at position 82). The third change is observed in the C-terminal transactivation domain (glutamine to histidine at position 253). The amino acid change at position 82 lies adjacent to a predicted sumoylation site (Ren *et al.*, 2009), a mechanism reported to be involved in activation of several stem cell transcription factors (Tsuruzoe *et al.*, 2006; Wei *et al.*, 2007). Due to predicted identical size of NANOG and NANOGP8 protein and high amino acid homology (the two proteins differ by only 3 amino acids), NANOG antibodies cannot distinguish between the two. Therefore, the presence of the NANOG protein could be due to translation of eNANOG, NANOGP8, or both. Translation of NANOG transcripts was confirmed in all cell types examined by western blotting where the NANOG antibody detected a single 37 kDa immunoreactive band. Migration rate of this protein remained identical in all samples suggesting either stable post-translational modifications at predicted sumoylation and/or phosphorylation sites or their absence (Yates and Chambers, 2005) in all cell types. Detection of the immunoreactive NANOG band in smooth muscle cells, which express NANOGP8 exclusively, however, demonstrates that NANOGP8 is a protein coding gene.

Transcription factors must localize to the nucleus to actively regulate transcription. It has been shown that human NANOG contains two nuclear localization signals (NLS) in the homeodomain (Chang *et al.*, 2009; Do *et al.*, 2007), allowing its translocation into the nucleus. The putative NLS sequences are identical between eNANOG and NANOGP8 and would indicate that nuclear translocation of both proteins can be expected. More recently, however, a nuclear export signal (NES) has been described for human NANOG (Chang *et al.*, 2009) which suggests a possible cellular shuttling behavior. In our study, differential sub-cellular localization of NANOG protein was observed in fibroblasts. The protein localized exclusively to either the nucleus, cytoplasm, or was present uniformly throughout both compartments. The pattern of sub-cellular localization could not be predicted based on the presence of one or both NANOG or NANOGP8 transcripts. Absence of any obvious difference in the migration of the protein, similarly, would suggest that sub-cellular localization was not a consequence of protein modification, but rather a consequence of cellular shuttling. These observations certainly fit with the predicted NANOG shuttling behavior, but have not been experimentally verified.

In this study, in addition to HUVECs, exclusion of NANOG from nuclei was observed in human mesenchymal stem cells and agrees with observations of Zuk (Zuk, 2009) who described nuclear exclusion of NANOG in mesenchymal and other human adult stem cells, but localization detection depended on the antibody used. Diffuse localization of NANOG throughout the cell was observed in transformed cells (HeLa and SH-SY5Y), similar to its localization in malignant cervical epithelial cells (Ye *et al.*, 2008). Since it was not possible to differentiate between NANOG and NANOGP8 protein within the same cell, it is unclear whether one or the other show preferential compartmentalization in all cell types examined. Here, we observed that transgenic neonatal fibroblasts expressing a GFP-fusion of NANOG or NANOGP8 both show nuclear translocations of the respective proteins. This is in agreement with our ICC data in which human smooth muscle cells, which expressed NANOGP8 exclusively at the transcript level, showed predominant nuclear localization of NANOG. The finding that nuclear extracts from this same cell type possess the ability to bind a known NANOG DNA binding element in our EMSA study suggests that the endogenous nuclear NANOGP8 we observed may be binding similar elements in differentiated cells.

Heterogeneity in levels of NANOG expression has been reported for human ESCs. Low NANOG expressing hESCs show a higher propensity to differentiate compared to high NANOG expressing hESCs (Kalmar *et al.*, 2009). NANOG also exhibits different functions in different types of stem cells as reported by Ji *et al.* (Ji *et al.*, 2009). Interestingly, and unexpectedly, in this study the only cell type in which NANOG transcript was not found was differentiated hESCs. One interpretation of this result to which we have alluded above is that the NANOG transcript we detect in primary human cells is only from a small population of – perhaps – stem or precursor cells in these cultures. However, it should be noted that neonatal human fibroblasts, which express only eNANOG, continued to maintain transcriptional levels of this gene even after culturing to senescence, 26 passages later (data not shown). This argues that the developmental program initiated upon the *in vitro* differentiation of hESCs may be markedly different than the one functioning during *in vivo* (fetal) development. It would appear, from the data presented here, that NANOG expression from either the parent locus or the P8 pseudogene is conserved in a cell-type specific manner into adulthood. Future studies, then, should address the correlation between this loss of eNANOG and the full developmental potential of hESCs.

Taken together, our study suggests that even though eNANOG and/or NANOGP8 transcripts are expressed at low levels, they may function as transcription factors in differentiated cells. Rapidly growing cells express eNANOG while NANOGP8 appears later as cells differentiate, with terminally differentiated smooth muscle cells expressing only NANOGP8. It is possible that some of these factors assert specific functions in a concentration dependent manner (Rodriguez *et al.*, 2007) rather than on/off status. Their specific function may also depend on cell and tissue type. The observation that various cell types may express eNANOG, NANOGP8 or both, supports the notion that low levels of eNANOG or NANOGP8 may be required for normal cell function even in differentiated cells.

Materials and Methods

Cell culture

All cell types used to perform transcriptional analysis for NANOG (Table 1) were grown at 37° C as follows: HeLa cells were grown in DMEM/Ham's F12 medium (50:50) supplemented with 10% fetal bovine serum; human smooth muscle cells (hSMCs) in M199 medium supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin; human mesenchymal stem cells (hMSCs) in MSCGM (Lonza) and human

umbilical vein endothelial cells (HUVECs) in EGM medium (Lonza) under atmospheric O_2 and 5% CO_2. Human embryonic stem cells (hESCs; W09, WiCell) were grown on mitomycin inactivated mouse embryonic fibroblasts in knock out DMEM medium (Gibco) supplemented with 20% knockout serum replacement supplement (Invitrogen), 1X non-essential amino acids, 1X β-mercaptoethanol, 10 ng/ml FGF2 and 2 mM L-glutamine medium under atmospheric O_2 and 5% CO_2 and passaged every 3 days. Adult fibroblasts (CRL-2352), neonatal fibroblasts (CRL-2097), SH-SY5Y neuroblastoma cells, and teratocarcinoma cells (CRL-2073) were grown in DMEM/Ham's F12 medium (50:50) supplemented with 2 mM L-Glutamine, 10% fetal clone III (Hyclone) in 5% O_2 and 5% CO_2 without antibiotic supplementation. *Drosophila* S3 cells were a gift from Dr. Joseph Duffy of WPI, Worcester, MA, USA.

Reverse Transcriptase PCR (RT-PCR)

Total RNA was isolated from cells using TRIZOL reagent (Invitrogen) following manufacturer's protocol. RNA was resuspended in RNase/DNase free water and stored at −80°C. Total RNA samples were digested with RNase-free DNase (rDNAse I, Ambion) following manufacturer's protocol to eliminate possible genomic DNA contamination. Amounts of RNA were quantified by spectrophotometry. For human heart tissue, total heart RNA (ClonTech) was used. First strand cDNA synthesis was performed using Superscript III first strand cDNA synthesis kit (Invitrogen) with oligo-dT primers from 5 μg total RNA following the manufacturer's protocol. Approximately 250 ng of first strand cDNA, 100 ng of human genomic DNA (ClonTech) or 10 ng of plasmid DNA was used as a template for respective PCR amplifications. All primers were custom designed and synthesized (Operon Biotechnologies, Table 3). Amplifications were performed in 25 μl reactions using ExTaq polymerase buffer (Takara Bio, USA) supplemented with 1.5 mM $MgCl_2$, 200 nM dNTP mix, 1 μM of each primer and 0.75 U polymerase as follows - initial denaturation at 95°C for 2 minutes, followed by 30 cycles of denaturation at 94°C for 15 sec; primer-specific annealing for 30 sec; extension for 1.5 to 2 minutes based on amplicon length (Table 3) and a final extension at 72°C for 10 min. PCR products were resolved on 2% agarose gels containing 0.5 mg/ml ethidium bromide. Images were acquired using a Kodak 4000MM image analyzer.

Quantitative real-time PCR (qPCR)

Two micrograms of total RNA was subjected to DNase I digestion, followed by a reverse transcription using QuantiTect Reverse transcription kit (Qiagen) with a mixture of oligo-dT and random hexamers primers. Fifty nanograms of cDNA was used per well as template in qPCR reactions with NANOG-specific primers (Table 3) using 200 nM of each primer per reaction in 20 μl volumes. Non-template controls and RT-minus samples were used to control for potential contaminating DNA and primer self amplification, respectively. All qPCR reactions were performed in triplicates with the resultant values being combined into an average threshold cycle. The efficiency of qPCR was calculated from the slope of a relative standard curve using β-actin primers. Relative quantification was determined using a 7500 Real Time PCR system (Applied Biosystems) measuring SYBR green fluorescence (PerfeCTa™ SYBR Green FastMix, Low ROX, Quanta Biosciences). NANOG expression was normalized to β-actin for each cell type. The ΔΔCt values obtained for each cell type relative to NANOG expression in hESCs were used to determine fold change in RNA expression patterns.

Immunocytochemistry

Cells were seeded into 24-well plates (BD Falcon) at 10,000 cells per well. At about 70% confluence, cells were washed with DPBS w/ Ca++/Mg++ (Mediatech) and fixed in cold methanol for 10 min, washed with phosphate buffered saline (PBS) and stored in PBS at 4°C until use. Cells were incubated with 1.5N HCl for 20 min at room temperature, washed with PBS and blocked for 30 min at room temperature with PBS and 5% fetal bovine serum (FBS). A mouse monoclonal NANOG antibody (Abnova,

Cat # H00079923-M08) was added at 2.5 μg/ml in PBS containing Tween-20 (0.05%) for 30 min at room temperature. Cells were washed four times in PBS/Tween-20. Alexafluor-568 labeled goat anti-mouse IgG (Invitrogen) secondary antibody in blocking solution (4 μg/ml) was added for 30 min. Cells were washed four times in PBS and stored in PBS at 4°C until image analysis. Cells were counterstained with Hoechst 33342 at 0.5 μg/ml before imaging and visualized using an Olympus IX81 inverted microscope with epi-fluorescence using appropriate filters (Semrock) and phase contrast. Images were collected using a 12 bit Hamamatzu CCD camera and processed using Slidebook imaging software (Intelligent Imaging Innovations, Inc).

Western blotting

Total protein was isolated from subconfluent cells with RIPA cell lysis buffer (Santa Cruz), supplemented with complete protease inhibitor cocktail (PIC, Santa Cruz) and 1mM DTT. Lysates were incubated on ice for 30 min and vortexed every 10 min. Protein concentration was determined with Quant-iT protein assay kit (Invitrogen). Four volumes of protein supernatant was mixed with one volume of denaturing 5X sample buffer (BioRad) and heated to 95°C for 3 min. Proteins were separated on 4–20% gradient SDS-PAGE gels and transferred to PVDF membranes (BioRad) using Towbins transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, and 0.037% SDS). The membranes were blocked with Tween-20-Tris-buffered saline (TTBS: 25 mM Tris, 137 mM NaCl, 2.7 mM KCl, 0.2% Tween-20), 5% dry milk (Santa Cruz) and 5% FBS. The same buffer was used for primary and secondary antibody incubations. Mouse monoclonal (Abnova, Cat # H00079923-M08) and a rabbit polyclonal (Abcam, Cat # ab21624) antibodies were used as the primary antibodies (for respective replicates) and an alkaline phosphatase-conjugated rabbit polyclonal to mouse IgG or goat polyclonal to rabbit IgG (both Abcam) were used as the respective secondary antibody. Membranes were washed three times with TTBS and allowed to develop color over 10 to 15 minutes using Western blue substrate (Promega). Images were scanned using a CanoScan LiDE 200 scanner (Cannon).

Cloning of NANOG amplified bands and NANOGP5

RT-PCR amplification products obtained using Nanog-387 primers were separated on 2% agarose gels. Nanog-387 and Nanog-349 bands from adult fibroblasts and Nanog-387 band from teratocarcinoma cells were isolated and extracted using PrepEase gel extraction kit (USB). A 1260 bp NANOGP5 fragment was amplified from human genomic DNA using NANOGP5 primers (Table 3). The fragments were cloned into pCR II cloning vector (Invitrogen). Recombinant plasmids were isolated using Qiagen plasmid miniprep kit (Qiagen) and sequenced by automated sequencing. DNA homology search was performed using Basic Local Alignment Search Tool (BLAST) and sequence alignments were done using LALIGN. Sumoylation sites in the NANOG and NANOGP8 proteins were predicted using the SUMOsp 2.0 program (Ren *et al.*, 2009) (http://sumosp.biocuckoo.org/prediction.php).

Sequencing of Nanog-1860 amplicon from adult fibroblasts

Total RNA from adult fibroblasts (CRL-2352, ATCC) was prepared as described earlier. The 1860 bp amplicon was derived by PCR amplification using Nanog-1860 primers (Table 3). The amplified product was gel purified, re-amplified using the same primers and cloned into pCR 4 TOPO vector (Invitrogen). Plasmids were isolated from twenty five recombinant clones selected at random and subjected to automated sequencing. Restriction analysis was used to identify the orientation of inserts in relation to the sequencing primers on the TOPO vector. Sequence information for the ORF for each clone was obtained using appropriate primer from 5' ends of the insert and two NANOG-specific internal primers (Table 3). Similarly, control eNANOG and NANOGP8 clones derived from human teratocarcinoma cells and human smooth muscle cells respectively (cloned in pDRIVE vector, Qiagen) were also sequenced. Sequences were aligned for comparison using CLUSTALW

software (http://align.genome.jp/).

NANOG-GFP and NANOGP8-GFP vector construction and transfection

Approximately 20 ng of first strand cDNA synthesized from human teratocarcinoma cells (expressing eNANOG exclusively) and human smooth muscle cells (expressing NANOGP8 exclusively) was used to amplify a 2031 bp NANOG and NANOGP8 amplicons respectively. The amplified fragment contained the ORF of each transcript and most of its 3' UTR. The amplified bands were cloned into pCR2.1 TA cloning vector (Invitrogen). A second round of PCR was performed on the respective pCR2.1 clones to amplify the 1093 bp ORF of each gene minus the stop codon and cloned into pCR2.1 TA cloning vector. The primer sequences are presented in Table 3. The reverse primer used for the second PCR contained an EcoRV restriction site. The respective plasmids were digested with EcoRI and EcoRV, inserts gel purified (PrepEase, USB) and cloned into pAcGFP-N2 vector (ClonTech) to develop NANOG-GFP and NANOGP8-GFP in-frame fusions. Plasmid DNA isolated from fusion genes were sequence verified, and transiently expressed in neonatal human fibroblast cells using Lipofectamine 2000 (Invitrogen). Cells were counterstained with Hoechst 33342 at 0.5 μg/ml before imaging, and fluorescent images were captured using an Olympus IX81 inverted microscope with epi-fluorescence using appropriate filters (Semrock). Images were collected using a 12 bit Hamamatzu CCD camera and processed using Slidebook software (Intelligent Imaging Innovations, Inc).

Isolation of cytoplasmic and nuclear protein extracts

All manipulations were performed at 4ºC and all buffers and solutions were maintained at 4ºC. All buffers were supplemented with protease inhibitor (Roche) and 7 mM β mercaptoethanol just before use. Cells were trypsinized and rinsed twice in cold PBS and pelleted at 200 x *g* for 5 minutes. Cell pellets were resuspended in 200 μl of buffer 1 (50 mM NaCl, 10 mM HEPES pH 8.0, 500 mM sucrose, 1 mM EDTA, 0.5 mM spermidine, 0.15 mM spermine, 0.2 % TritonX-100) and vortexed. Nuclei were pelleted by centrifugation at 5,000 x g for 2 minutes. The supernatant was collected as cytoplasmic extract. The nuclear pellet was rinsed in 500 μl of buffer 2 (50 mM NaCl, 10 mM HEPES pH 8.0, 25 % glycerol, 0.1 mM EDTA, 0.5 mM spermidine, 0.15 mM spermine), centrifuged as before and the supernatant discarded. The nuclear pellet was resuspended in 50 μl buffer 3 (350 mM NaCl, 10 mM HEPES pH 8.0, 25 % glycerol, 0.1 mM EDTA, 0.5 mM spermidine, 0.15 mM spermine), vortexed for 30 minutes at 4ºC and centrifuged at 21,000 x *g* for 15 minutes. The supernatant was collected and saved as nuclear extract. Protein concentrations were determined by 280 nm absorbance using the NanoDrop 2000C (Thermoscientific).

Electrophoretic Mobility Shift Assay (EMSA)

Sense and complementary EMSA oligos designed based on consensus NANOG binding sequence (Mitsui *et al.*, 2003), and flanked by PCR priming sites were custom synthesized (Eurofins MWG Operon). A 5' biotin labeled PCR primer was used to amplify the EMSA probe, precipitated and resuspended in appropriate volume of DNase-free water. Sequences for EMSA probe and PCR primer are presented in Table 3. EMSA was carried out using the LightShift Chemiluminescent EMSA Kit (Piercenet) following the manufacturer's protocol. Briefly, 20 fmoles of 5' biotinylated EMSA probe was mixed with 5 μg of nuclear extract, in 1X binding buffer [5 mM MgCl2, 50 ng/μl Poly (dI «dC) and 0.05% NP-40] in a final volume of 20 μl and incubated at room temperature for 20 minutes. After the binding reaction, 5 μl of 5X loading buffer was added and samples loaded on an 8% native polyacrylamide gel in 0.5X TBE (30 minutes of pre-run followed by 60 minutes of sample separation). The DNA was transferred onto positively charged nylon membrane at 380 mA for 30 minutes and cross-linked with a hand held UV lamp (260 nm) for 10 minutes. Biotin labeled probes on the membrane were detected using the

Chemiluminescent Nucleic Acid Detection Module (Piercenet) following the manufacturer's recommendation. Images were captured using Kodak 4000MM image analyzer.

Differentiation of hESCs

Human ESCs were cultured in hESC medium as described earlier. Human ESC clumps were isolated using a sterile cell scraper and transferred to a gelatinized cell culture plate and cultured for 7 days (passage P1) in a medium containing equal amounts DMEM and Hams-F12 (Mediatech) supplemented with 2 mM L-glutamine and 15% FBS. The same subculturing procedure was repeated and cells grown in the same medium for an additional 7 days (passage P2). Passages 3 and 4 were done by subculturing after trypsin treatment of cells in the same medium for additional 11 days and 15 days, respectively. Total RNA from differentiated cells from each passage was isolated using TRIzol (Invitrogen) following the manufacturer's protocol. First strand cDNA was synthesized using poly-dT primers (Superscript III, Invitrogen) using 4 μg of total RNA. RT-PCR was performed using approximately 400 ng of first strand cDNA using Nanog-1860 primers and β-actin primers (Table 3) as described earlier.

Restriction Digestion

The 1860 bp fragment encompassing NANOG open reading frame and most of its 3'UTR was amplified using first strand cDNA templates synthesized from all cell types. PCR products were purified using PrepEase gel extraction kit (USB) and digested with Sma I (New England Biolabs). Digested products were resolved on 2% agarose gels containing 0.5 μg/ml ethidium bromide and images were acquired using a Kodak 4000MM image analyzer.

References

BARRAND, S. and COLLAS, P. (2009). Chromatin states of core pluripotency-associated genes in pluripotent, multipotent and differentiated cells. *Biochem Biophys Res Commun* 391: 762-767.

BOOTH, H. and HOLLAND, P. (2004). Eleven daughters of NANOG. *Genomics* 84: 229-238.

BOYER, L., LEE, T., COLE, M., JOHNSTONE, S., LEVINE, S., ZUCKER, J., GUENTHER, M., KUMAR, R., MURRAY, H., JENNER, R. *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947-956.

BYRNE, J., NGUYEN, H. and REIJO PERA, R. (2009). Enhanced generation of induced pluripotent stem cells from a subpopulation of human fibroblasts. *PLoS One* 4: e7118.

CHAMBERS, I., COLBY, D., ROBERTSON, M., NICHOLS, J., LEE, S., TWEEDIE, S. and SMITH, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113: 643-655.

CHANG, D.F., TSAI, S.C., WANG, X.C., XIA, P., SENADHEERA, D. and LUTZKO, C. (2009). Molecular characterization of the human NANOG protein. *Stem Cells* 27: 812-821.

CHEN, J., SUN, M., KENT, W., HUANG, X., XIE, H., WANG, W., ZHOU, G., SHI, R. and ROWLEY, J. (2004). Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res* 32: 4812-4820.

CLAMP, M., FRY, B., KAMAL, M., XIE, X., CUFF, J., LIN, M., KELLIS, M., LINDBLAD-TOH, K. and LANDER, E. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci USA* 104: 19428-19433.

DO, H., LIM, H., KIM, J., SONG, H. and CHUNG, H. (2007). An intact homeobox

domain is required for complete nuclear localization of human Nanog. *Biochem Biophys Res Commun* 353: 770-775.

ELLIMAN, S., WU, I. and KEMP, D. (2006). Adult tissue-specific expression of a Dppa3-derived retrogene represents a postnatal transcript of pluripotent cell origin. *J Biol Chem* 281: 16-19.

EZEH, U., TUREK, P., REIJO, R. and CLARK, A. (2005). Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma. *Cancer* 104: 2255-2265.

HART, A., HARTLEY, L., IBRAHIM, M. and ROBB, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Dev Dyn* 230: 187-198.

HART, A., HARTLEY, L., PARKER, K., IBRAHIM, M., LOOIJENGA, L., PAUCHNIK, M., CHOW, C. and ROBB, L. (2005). The pluripotency homeobox gene NANOG is expressed in human germ cell tumors. *Cancer* 104: 2092-2098.

HATANO, S.Y., TADA, M., KIMURA, H., YAMAGUCHI, S., KONO, T., NAKANO, T., SUEMORI, H., NAKATSUJI, N. and TADA, T. (2005). Pluripotential competence of cells associated with Nanog activity. *Mech Dev* 122: 67-79.

HYSLOP, L., STOJKOVIC, M., ARMSTRONG, L., WALTER, T., STOJKOVIC, P., PRZYBORSKI, S., HERBERT, M., MURDOCH, A., STRACHAN, T. and LAKO, M. (2005). Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem Cells* 23: 1035-1043.

JI, J., WERBOWETSKI-OGILVIE, T., ZHONG, B., HONG, S. and BHATIA, M. (2009). Pluripotent transcription factors possess distinct roles in normal versus transformed human stem cells. *PLoS One* 4: e8065.

KALMAR, T., LIM, C., HAYWARD, P., MUÑOZ-DESCALZO, S., NICHOLS, J., GARCIA-OJALVO, J. and MARTINEZ ARIAS, A. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* 7: e1000149.

KERR, C., HILL, C., BLUMENTHAL, P. and GEARHART, J. (2008). Expression of pluripotent stem cell markers in the human fetal testis. *Stem Cells* 26: 412-421.

KLOSE, R. and BIRD, A. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31: 89-97.

LIEDTKE, S., ENCZMANN, J., WACLAWCZYK, S., WERNET, P. and KOGLER, G. (2007). Oct4 and its pseudogenes confuse stem cell research. *Cell Stem Cell* 1: 364-366.

LIEDTKE, S., STEPHAN, M. and KÖGLER, G. (2008b). Oct4 expression revisited: potential pitfalls for data misinterpretation in stem cell research. *Biol Chem* 389: 845-850.

LISS, B. (2002). Improved quantitative real-time RT-PCR for expression profiling of individual cells. *Nucleic Acids Res* 30: e89.

MITSUI, K., TOKUZAWA, Y., ITOH, H., SEGAWA, K., MURAKAMI, M., TAKAHASHI, K., MARUYAMA, M., MAEDA, M. and YAMANAKA, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113: 631-642.

MULLIN, N., YATES, A., ROWE, A., NIJMEIJER, B., COLBY, D., BARLOW, P., WALKINSHAW, M. and CHAMBERS, I. (2008). The pluripotency rheostat Nanog functions as a dimer. *Biochem J* 411: 227-231.

PAGE, R.L., AMBADY, S., HOLMES, W.F., VILNER, L., KOLE, D., KASHPUR, O., HUNTRESS, V., VOJTIC, I., WHITTON, H. and DOMINKO, T. (2009). Induction of stem cell gene expression in adult human fibroblasts without transgenes. *Cloning Stem Cells* 11: 417-426.

PAIN, D., CHIRN, G., STRASSEL, C. and KEMP, D. (2005). Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. *J Biol Chem* 280: 6265-6268.

PAN, G. and PEI, D. (2003). Identification of two distinct transactivation domains in the pluripotency sustaining factor nanog. *Cell Res* 13: 499-502.

PANAGOPOULOS, I., MOLLER, E., COLLIN, A. and MERTENS, F. (2008). The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1. *Oncol Rep* 20: 1029-1033.

PARK, I.H., ZHAO, R., WEST, J.A., YABUUCHI, A., HUO, H., INCE, T.A., LEROU, P.H., LENSCH, M.W. and DALEY, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 451: 141-146.

POLESSKAYA, A., CUVELLIER, S., NAGUIBNEVA, I., DUQUET, A., MOSS, E. and HAREL-BELLAN, A. (2007). Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev* 21: 1125-1138.

QUE, J., LUO, X., SCHWARTZ, R. and HOGAN, B. (2009). Multiple roles for Sox2 in the developing and adult mouse trachea. *Development* 136: 1899-1907.

REN, J., GAO, X., JIN, C., ZHU, M., WANG, X., SHAW, A., WEN, L., YAO, X. and XUE, Y. (2009). Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. *Proteomics* 9: 3409-3412.

REUBINOFF, B., PERA, M., FONG, C., TROUNSON, A. and BONGSO, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nat Biotechnol* 18: 399-404.

RODRIGUEZ, R., VELKEY, J., LUTZKO, C., SEERKE, R., KOHN, D., O'SHEA, K. and FIRPO, M. (2007). Manipulation of OCT4 levels in human embryonic stem cells results in induction of differential cell types. *Exp Biol Med (Maywood)* 232: 1368-80.

SILVA, J., CHAMBERS, I., POLLARD, S. and SMITH, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. *Nature* 441: 997-1001.

SILVA, J., NICHOLS, J., THEUNISSEN, T., GUO, G., VAN OOSTEN, A., BARRANDON, O., WRAY, J., YAMANAKA, S., CHAMBERS, I. and SMITH, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* 138: 722-737.

SUO, G., HAN, J., WANG, X., ZHANG, J., ZHAO, Y. and DAI, J. (2005). Oct4 pseudogenes are transcribed in cancers. *Biochem Biophys Res Commun* 337: 1047-1051.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. and YAMANAKA, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131: 861-872.

TORRENTS, D., SUYAMA, M., ZDOBNOV, E. and BORK, P. (2003). A genome-wide survey of human pseudogenes. *Genome Res* 13: 2559-2567.

TSURUZOE, S., ISHIHARA, K., UCHIMURA, Y., WATANABE, S., SEKITA, Y., AOTO, T., SAITOH, H., YUASA, Y., NIWA, H., KAWASUJI, M. *et al.* (2006). Inhibition of DNA binding of Sox2 by the SUMO conjugation. *Biochem Biophys Res Commun* 351: 920-926.

WANG, J., LEVASSEUR, D. and ORKIN, S. (2008). Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. *Proc Natl Acad Sci USA* 105: 6326-6331.

WEI, F., SCHOLER, H.R. and ATCHISON, M.L. (2007). Sumoylation of Oct4 enhances its stability, DNA binding, and transactivation. *J Biol Chem* 282: 21551-21560.

YAMAGUCHI, S., KIMURA, H., TADA, M., NAKATSUJI, N. and TADA, T. (2005). Nanog expression in mouse germ cell development. *Gene Expr Patterns* 5: 639-646.

YATES, A. and CHAMBERS, I. (2005). The homeodomain protein Nanog and pluripotency in mouse embryonic stem cells. *Biochem Soc Trans* 33: 1518-1521.

YE, F., ZHOU, C., CHENG, Q., SHEN, J. and CHEN, H. (2008). Stem-cell-abundant proteins Nanog, Nucleostemin and Musashi1 are highly expressed in malignant cervical epithelial cells. *BMC Cancer* 8: 108.

YU, J., HU, K., SMUGA-OTTO, K., TIAN, S., STEWART, R., SLUKVIN, I. and THOMSON, J. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. *Science* 324: 797-801.

YU, J., VODYANIK, M.A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J.L., TIAN, S., NIE, J., JONSDOTTIR, G.A., RUOTTI, V., STEWART, R. *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318: 1917-1920.

ZHANG, J., WANG, X., LI, M., HAN, J., CHEN, B., WANG, B. and DAI, J. (2006). NANOGP8 is a retrogene expressed in cancers. *FEBS J* 273: 1723-1730.

ZUK, P. (2009). The intracellular distribution of the ES cell totipotent markers OCT4 and Sox2 in adult stem cells differs dramatically according to commercial antibody used. *J Cell Biochem* 106: 867-877.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

See our recent Special Issue *Placenta* edited by Joan S. Hunt and Kent L. Thornburg at:
http://www.ijdb.ehu.es/web/contents.php?vol=54&issue=2-3

Feeder- and serum-free establishment and expansion of human induced pluripotent stem cells
Mehdi Totonchi, Adeleh Taei, Ali Seifinejad, Mohammadsharif Tabebordbar, Hassan Rassouli, Ali Farrokhi, Hamid Gourabi, Nasser Aghdami,
Ghasem Hosseini-Salekdeh and Hossein Baharvand
Int. J. Dev. Biol. (2010) 54: 877-886

Chromatin states of developmentally-regulated genes revealed by DNA and histone methylation patterns in zebrafish embryos
Leif C. Lindeman, Cecilia L. Winata, Håvard Aanes, Sinnakaruppan Mathavan, Peter Aleström and Philippe Collas
Int. J. Dev. Biol. (2010) 54: 803-813

Pluripotency of bank vole embryonic cells depends on FGF2 and activin A signaling pathways
Aneta Suwinska, Andrzej K. Tarkowski and Maria A. Ciemerych
Int. J. Dev. Biol. (2010) 54: 113-124

A stem cell-derived gene *(Sddr)* negatively regulates differentiation of embryonic stem cells
Miwako Miura, Atsushi Ueda, Yukinari Takao, Emi K. Nishimura, Hiroshi Koide and Takashi Yokota
Int. J. Dev. Biol. (2010) 54: 33-39

FGF signalling and blastema growth during amphibian tail regeneration
P Ferretti, F Zhang, L Santos-Ruiz, JDW Clarke
Int. J. Dev. Biol. (2001) 45: S127-S128

5 yr ISI Impact Factor (2008) = 3.271