Conclusive evidence for *OCT4* transcription in human cancer cell lines: possible role of a small OCT4-positive cancer cell population

Running head:
OCT4 transcription and translation in human cancer

Tomoyuki Miyamoto[1,2,*], Nobuhiko Mizuno[1*], Mitsuko Kosaka[1#], Yoko Fujitani[1], Eiji Ohno[2], Aiji Ohtsuka[1]

[1] Department of Human Morphology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, 2-5-1, Shikata, Kita, Okayama 700-8558, Japan

[2] Department of Medical Life Science, Faculty of Medical Bioscience, Kyushu University of Health and Welfare/Cancer Cell Institute of Kyushu University of Health and Welfare, 1714-1, Yoshino, Nobeoka, Miyazaki 882-8508, Japan

* These authors contributed equally to this study.

 # Corresponding author: Mitsuko Kosaka, PhD

Department of Human Morphology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, 2-5-1, Shikata-cho, Kita-ku, Okayama 700-8558, Japan.

Phone: 81-86-235-7092, ext. 7092

FAX: 81-86-235-7095

E-mail: m-kosaka@md.okayama-u.ac.jp

Author contributions

Tomoyuki Miyamoto: collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript

Nobuhiko Mizuno: conception and design, collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript

Mitsuko Kosaka: conception and design, collection and/or assembly of data, data analysis and interpretation, manuscript writing, provision of study material, financial support, administrative support, final approval of manuscript

Yoko Fujitani: collection and/or assembly of data, data analysis and interpretation, final approval of manuscript

Eiji Ohno: financial support, provision of study material, final approval of manuscript

Aiji Ohtsuka: financial support, administrative support, final approval of manuscript

Funding

This work was supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JP23592606 & JP15K15016 to M.K.) and the Translational Research Network Program from the Japan Agency for Medical Research and Development (to M.K.).

Keywords: OCT4, splicing, cancer, cancer stem cells, malignancy

ABSTRACT

The role of octamer-binding transcription factor 4 (OCT4) in human cancer is still debated. Although many studies have been published on human OCT4, determining which of the findings are accurate or which are false-positives is currently challenging. We thus developed the most reliable method to date for highly specific and comprehensive detection of genuine *OCT4*-transcript variants without false-positive results. Our results provided clear evidence that the transcripts of *OCT4A*, *OCT4B*, *OCT4B1* and other novel splicing variants are indeed present in many cancer cell lines, but are rarely detected in normal tissue-derived differentiated cells. Using the tagged genomic transgene, we then verified endogenous OCT4A translation in cancer cell subpopulations. Moreover, analysis of possible other protein isoforms by enforced expression of *OCT4B* variants showed that the B164 isoform, designated human OCT4C, is preferentially produced in a cap-dependent manner. We confirmed that the OCT4C isoform, similar to OCT4A, can transform non-tumorigenic fibroblasts *in vitro*. Finally, ablation of OCT4-positive cells using promoter-driven diphtheria toxin A (DTA) in high malignant cancer cells caused a significant decrease in migration and Matrigel invasion. These findings strongly suggest a significant contribution of OCT4 to the phenotype of human cancer cells.

Significance statement

Abundant information on human OCT4 expression has been provided by stem cell and cancer biology studies; however, this includes a large amount of unconvincing data owing to the existence of active *OCT4* pseudogenes. To overcome this problem, we developed an indisputable method for detecting genuine *OCT4* transcripts and translation products, which eliminates all false-positive results. Moreover, we show conclusive evidence for the presence of an OCT4-positive subpopulation and the correlation with migration and invasion in human cancer cells. Our methods and experimental data eliminate longstanding confusion and represent the first step toward uncovering the true role of OCT4 in human somatic cancer.

Introduction

Stem cells play a critical role in the generation of complex multicellular organisms and tumor development. Tumors contain a small subpopulation of cells, termed cancer stem (-like) cells (CSCs) or tumor-initiating cells (TICs), that exhibit self-renewal capacity and are responsible for tumor maintenance and metastasis (1,2). Accordingly, the ability to identify, target, and eliminate CSCs is critical for cancer diagnosis and therapy. Growing evidence indicates cross-talk and correlations among stemness pathways, tumor progression, and metastasis; however, the functional significance of overexpressed stem cell markers in cancer is largely unknown (3,4).

The transcription factor octamer-binding transcription factor 4A (OCT4A; also known as OCT3, OCT-3/4, or POU5F1) is a key regulator of pluripotency during the earliest stages of mammalian development (5,6), pluripotency maintenance, and embryonic stem cell self-renewal (7,8). In addition, it is an essential factor in cellular reprogramming and pluripotency acquisition (9-11). In adult male mice, OCT4A maintains the pluripotency of spermatogonial stem cells as well as their undifferentiated, self-renewing state (12,13). Moreover, *OCT4A* transcripts are consistently detected in human embryonic carcinomas and testicular germ-cell tumors with pluripotent potential, suggesting its critical role in embryonic or germ-cell tumorigenesis (14-17).

Numerous studies have focused on *OCT4A* as a candidate CSC marker; however, investigations of *OCT4* expression in somatic and/or cancer tissues have yielded controversial results, despite the

importance of the locus for stem cell and tumor biology. Differences among studies can be attributed to the presence of highly homologous transcribed pseudogenes (pgs) and transcript variants (18-20). Although *OCT4* expression was demonstrated at both the mRNA and protein levels in somatic and/or tumor cells (Fig. S1), some critical studies have highlighted the potential misinterpretation of *OCT4A*-expression results in somatic cancers (21,22) depending on the experimental design and the use of nonspecific or poorly characterized reagents, including antibodies and primers (20,23,24).

Human *OCT4* is alternatively spliced into at least three transcript variants, i.e. *OCT4A*, *OCT4B*, and *OCT4B1*, further complicating the interpretation of expression results (25,26). *OCT4A* is well-studied, whereas the functions of the other two variants are still under investigation (27-29). *OCT4B* and *OCT4B1* do not share the pluripotency characteristics of *OCT4A*, but are associated with anti-apoptotic effects and stress responses (30). Previous studies reported that *OCT4B* and *OCT4B1* encode the same protein, of which at least three isoforms (B265, B190, and B164) are produced by alternative translation initiation (28). Moreover, a single-nucleotide polymorphism (SNP; rs3130932) in *OCT4B*, first ATG → AGG, is expected to result in reduced expression in individuals carrying the AGG genotype, although the AGG genotype in rs3130932 is not associated with increased (or decreased) cancer risk (31). Therefore, the functions of OCT4B and OCT4B1 in cancer remain largely unknown.

Many oncogenes and tumor suppressors are differentially spliced in cancer cells, and many of these cancer-specific isoforms contribute to the transformed phenotype of cancer cells (32-34). An

undiscovered cancer-specific OCT4 isoform might contribute to CSC maintenance; however, further investigations of *OCT4* variants at the mRNA and protein levels are needed to determine relationships between OCT4 isoforms and oncogenesis and their potential as CSC markers.

In this study, we developed a simple reverse-transcription polymerase chain reaction (RT-PCR) method using specific primer sets and excluding amplification of active *OCT4* pgs and genomic DNA contamination. In addition, we comprehensively identified *OCT4* multiple transcripts, as well as their possible translation products, in human cancer cells. Our findings highlight the importance of OCT4A and OCT4C isoforms in tumorigenicity. Furthermore, we addressed the function of the OCT4A-positive subpopulation in a highly malignant tumor cell line.

Materials and methods

Isolation of human OCT4-pg1, -pg3, and -pg4 DNA

Human OCT4-pg1, -pg3, and -pg4 DNA fragments were isolated from human genomic DNA using the following primer sets. HOCT4-pg1-FO (5′-TCAGGCACTGTGTTCATTGCTAGTGAG-3′) and HOCT4-pg1-RV (5′-ACTGTGTCCCAGGCTTCTTTATTTAAG-3′) (product size: 1453 bp); HOCT4-pg3-FO (5′-AACGCTTCAACAAGAAGATACAGACATG-3′) and HOCT4-pg3-RV (5′-CAAGAGCATCATTGAACTTCACCTTC-3′) (product size: 1396 bp); and HOCT4-pg4-FO (5′-ATAAATGGTCAAGATGTCTCAAACTAC-3′) and HOCT4-pg4-RV (5′-

TCCTAAATTCTTATATACTGTTAGATC-3′) (product size: 1567 bp).

PrimeStar PCR enzyme (Takara, Tokyo, Japan) or EmeraldAmp PCR enzyme (Takara) was used for all PCRs. Two-step PCR conditions were as follows: 35 cycles at 96°C for 30 s and 68°C for 2 min. PCR products were isolated and ligated into the PCR-Blunt vector (Invitrogen, Carlsbad, CA, USA) and sequenced using an ABI-3130 sequencer (Applied Bioscience, Tokyo, Japan; Central Research Laboratory, Okayama University Medical School). All PCR primers are described in Table 1.

Cell culture

Cell lines were obtained from the JCRB (Osaka, Japan), RIKEN BRC (Tsukuba, Japan), or ECACC (Salisbury, UK) in 2016 and passaged for <6 months before experiments. MCF7 (JCRB0134), HeLa (RCB0007), Ishikawa (ECACC 99040201), HEC265 (JCRB1142), HEC1 (JCRB0042), and HEC50B (JCRB1145) cells were cultured in Eagle's minimum essential medium (MEM; #21442-25; Nacalai Tesque, Kyoto, Japan) supplemented with 10% fetal bovine serum (FBS). TTA1, kindly provided by Dr. Yoshida, Kanagawa Cancer Center (35), A549 (RCB0098), S2 (RCB2133), and PA1 (RCB1946) cells were cultured in RPMI1640 (# 30264-85; Nacalai Tesque) supplemented with 10% FBS. HEK293T, ARPE-19 (ATCC CRL-2302; ATCC, Manassas, VA, USA), HFF, HUVEC (#8000; CosmoBio, Tokyo, Japan), and HAoSMC (#6110; CosmoBio) cells were cultured in Dulbecco's MEM supplemented with 10% FBS. All cell types are summarized in Table S1. Cell lines were routinely tested for *Mycoplasma*

contamination in our laboratory.

Total RNA extraction and RT-PCR

Total RNA was extracted from each cell line using TRIzol reagent (#10296028; Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. First-strand cDNA was synthesized using oligo-dT primers and the SuperScript III first-strand synthesis system (#18080051; Life Technologies) according to the manufacturer's protocol. Reverse transcriptase-negative (RT−) control samples were obtained without reverse-transcriptase treatment. PCR was performed using a thermal cycler (Thermo Fisher Scientific, Massachusetts, USA) with the following conditions: 96°C for 1 min, followed by denaturation at 96°C for 30 s, annealing and extension at 68°C for 2 min (35 cycles), and final elongation at 68°C for 7 min. PrimeSTAR HS DNA polymerase was used according to manufacturer instructions (#R010B; TaKaRa). *Glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*) was used to test cDNA integrity. PCR products were separated by 1.5% agarose gel electrophoresis, stained with ethidium bromide, and visualized under an ultraviolet light.

Cloning and sequencing analysis

After electrophoretic separation, PCR amplicons were extracted using the QIAquick gel extraction kit (#28706; Qiagen, Hilden, Germany) and cloned into the pCR-Blunt vector (#K280040; Life

Technologies). Recombined constructs were transformed into TOP10 competent cells (#K280040; Life Technologies). Plasmids were isolated using the QIAGEN plasmid mini kit (#12125; Qiagen) and sequenced using an ABI-3130 sequencer. Plasmids extracted from randomly selected colonies were classified based on their sequences, which were analyzed using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) or genetic information processing software (Genetyx Corporation, Tokyo, Japan).

Plasmid construction and transfection

The transfection of plasmids expressing *OCT4* variants was performed using Lipofectamine 2000 or 3000 (Life Technologies) according to the manufacturer's instructions. DNA constructs are schematically depicted in Fig. S2. Transfection efficiencies were confirmed every time using red fluorescent protein (RFP)-expressing control vector. The study was conducted in accordance with guidelines issued by the Okayama University Safety Committee for Recombinant DNA Experiments.

Immunocytochemistry

For immunocytochemistry, non-transfected and transfected (*OCT4* genomic transgene tagged FLAG; pOCT4Gen-FLAG, Fig. S2) cells were fixed with 4% paraformaldehyde for 15 min at 25°C. The cells were then permeabilized in 0.2% Triton X-100 for 20 min. Mouse monoclonal anti-FLAG antibody (clone

M2; #F1804, Sigma) was used. Cells were incubated with the primary antibody at room temperature for 45 min, washed three times in phosphate-buffered saline (PBS), incubated with the secondary antibody (Alexa 488 conjugated goat anti-mouse IgG, #ab150113; Invitrogen), and washed three times in PBS. The cells were then counterstained with DAPI (#D1306; Invitrogen) and visualized under a fluorescence microscope. The primary antibody was replaced by PBS in negative controls.

In-gel detection of nanoluciferase fusion protein

At 24 h after transfection (*OCT4* genomic transgene tagged nanoluciferase; pOCT4Gen-Nluc, Fig. S2), cells were washed with PBS and lysed for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Each sample was separated by 10% SDS-PAGE. The gel was washed twice with 25% isopropanol for 15 min, and twice with water for 10 min. To detect Nluc fusion protein, the gel was soaked in Nano-Glo In-Gel Detection Reagent (#N3020; Promega, Madison, WI, USA) and the signal intensities were quantified directly with a chemiluminescence imager (Fusion FX7; M&S instruments Inc., Osaka, Japan).

Luciferase assay and bioluminescence imaging

At 24 h after transfection (pOCT4Gen-Nluc, *OCT4* 5′ upstream regulatory region (36,37) without the coding region to drive enhanced green fluorescent protein (EGFP) and diphtheria toxin fragment A

(DTA); pOCT4-GFP and pOCT4-DTA, Fig. S2), cells were washed with PBS and lysed with Passive Lysis Buffer (#E1941; Promega). Nluc activities were measured using the Nano-Glo Luciferase Assay System (#N1110; Promega) by a microplate reader (Flexstation 3; Molecular Devices, Central Research Laboratory, Okayama University Medical School). For bioluminescence imaging, the growth medium was replaced with Opti-MEM medium containing the substrate (Nano-Glo Luciferase Assay System, #N1110; Promega). Luminescence images were obtained using the LV200 system (Olympus Life Science, Tokyo, Japan), kindly supported by Drs. Matsui H and Yoshii T (Okayama University) and Mr. Yamada (Olympus Co., Ltd.); this system was equipped with a Hamamatsu ImagEM X2 CCD camera (#C9100-23B; Hamamatsu Photonics), a 20×/0.2 NA objective, and a temperature-controlled stage. Images were acquired with the acquisition feature of the Olympus CellSense software package. For image acquisition, exposure times/EM (electron-multiplying) gains were set to 100 msec/4 and 30 sec/1000 for the bright-field and luminescence channels, respectively.

5′-rapid amplification of cDNA ends (5′-RACE)

Total RNA was isolated from transfected COS7 cells. 5′-RACE was performed using the 5′-RACE system for rapid amplification of cDNA ends (v2.0; #18374058; Life Technologies) according to the manufacturer's instructions, and nested PCR was performed. The first PCRs were performed using the following primer sets: nFO1 primer 5′-GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3′ and

nRV1 primer 5′-TCACTTGTCGTCATCGTCCTTGTAATC-3′. The second PCRs were performed using nested primer sets: nFO2 primer 5′-GGCCACGCGTCGACTAGTAC-3′ and nRV2 primer 5′-GTTTGAATGCATGGGAGAGCCCAGAG-3′. PCR products were ligated into the pCR-Blunt vector and sequenced.

Western blot

At 32 h after transfection, cells were treated with the proteasome inhibitor MG132 (50 μM; #135-18453; Wako, Richmond, VA, USA) for 4 h, washed with PBS, and lysed for SDS-PAGE. Each sample was separated by 12% SDS-PAGE and the proteins were electrotransferred to a polyvinylidene fluoride membrane (#IPVH00010; Merck Millipore, Billerica, MA, USA), which was blocked with 2% skimmed milk in Tris-buffered saline with Tween-20 and probed with mouse monoclonal alkaline-phosphatase-conjugated anti-FLAG antibody (clone M2; #A9469; Sigma-Aldrich, St. Louis, MO, USA). The membranes were developed using the NBT/BCIP liquid substrate system (#B1911; Sigma-Aldrich).

Soft-agar colony formation assay

We established stable cell lines that expressed OCT4A, OCT4C, OCT4CΔNLS, and empty vector. The soft-agar colony formation assay was performed as previously described (38). After a 3-week culture, visible colonies were counted.

Matrigel invasion assay

Migration and invasion assays were performed according to the manufacturer's instructions (8 μm, #354480; Corning, NY, USA). A total of 2.5 x 10^4 transfected cells was seeded onto the top of control insert or Matrigel chamber (#354578; Corning), and normal growth medium was added to each well in the lower chamber. Following 3 days of incubation, non-invasive cells were removed from the upper chamber, and the cells attached to the lower chamber were fixed with methanol, stained with 10% Giemsa solution (#15003; Muto Pure Chemicals, Tokyo, Japan) and then counted under a light microscope for a whole field per well and six replicate wells per condition.

Results

Specific primers for detection of human OCT4A transcripts

Bioinformatics analysis identified at least six *OCT4* pgs in humans that are highly homologous to the true human *OCT4* gene (39). Notably, the transcripts from *OCT4-pg1*, *OCT4-pg3*, and *OCT4-pg4* can be translated into protein products, but OCT4A-like activity is lacking (18,21). Here, we re-analyzed the sequences of the human *OCT4* gene and its pgs by performing Ensembl BLAST searches of the human genome and designing specific primer sets (SetA1 and SetA2) expected to avoid false-positive detection of *OCT4A* transcripts (Fig. 1 and Table 1). Using *OCT4A* cDNA and *OCT4-pg1*, *-pg3*, and *-pg4* DNAs,

we demonstrated the specificity of the SetA1 and SetA2 primers relative to that of other promising primer sets used previously (18,20,26; Fig. 1A). As a control, we showed that primer Set1 containing nucleotide (nt) sequences shared between the genuine *OCT4A* gene and pgs could amplify genuine *OCT4A* cDNA, as well as *OCT4-pg1*, *-pg3*, and *-pg4* DNAs (Fig. 1Ba). The common forward primer of primer Set2 and Set3 featured a one-base mismatch at the 3′ end of *OCT4-pg1*, *-pg3*, and *-pg4* DNAs. The reverse primer in primer Set2 also contained a single-base mismatch at the 3′ end relative to *OCT4-pg1*, *-pg3*, and *-pg4* DNAs, whereas the reverse primer in primer Set3 contained only a single mismatched base at the 3′ end relative to the *OCT4-pg1* sequence. Although both primer Set2 and Set3 contained a mismatched base at a critical position for PCR primer specificity and sensitivity, they did not eliminate pg amplification (Fig. 1Bb and 1Bc). By contrast, when we used primer SetA1 and SetA2, we obtained only genuine *OCT4A* products and no DNA from any of the pgs (Fig. 1Bh and 1Bi).

The Liedtke-2 and Atlasi primer sets carry a polymorphism at the 3′ end, which is unique in *OCT4* and theoretically differentiates the genuine transcript and pgs. The reverse primer here is intron-spanning and designed to avoid amplification of genomic DNA. The forward primer in the Liedtke-1 and Suo primer sets targets regions featuring a sequence slightly different from the pg sequence (Fig 1C). The Liedtke-1 and -2 and Atlasi primer sets amplified both genuine *OCT4A* and certain pg DNAs (Fig. 1Be–1Bg). Conversely, the Suo primer set was highly specific and excluded pg amplification, but its sensitivity was lower than that of SetA1 and SetA2 (Fig. 1Bd, 1Bh, and 1Bi). Moreover, although the Suo primer set

was specific, it could not be used to distinguish genomic DNA contamination, because both the forward and reverse primers were designed within exon 1.

These results suggested that almost all of the primer sets used in previous *OCT4A* expression analyses had the potential to amplify pg transcripts and contaminating genomic DNA. We confirmed that our newly designed primer sets (SetA1 and SetA2) allowed the genuine *OCT4A* product to be distinguished from both pgs and genomic DNA.

Expression of* bona fide *OCT4 and its novel transcript variants in human cancer cells

Using Primer SetA1, we reinvestigated *OCT4A* transcript expression in a wide variety of human cancer cell lines by RT-PCR (Fig. 2A and 2B). The predicted *OCT4A* PCR product (1347 bp) was detected in PA1 cells (positive controls), as well as in many of the cancer cell lines examined (HeLa, Ishikawa, HEC265, HEC1, HEC50B, A549, and HEK293T), indicating the presence of *OCT4A* transcripts in these cancer cells. Next, we compared the *OCT4A* expression levels among these cell lines by semi-quantitative RT-PCR analysis (Fig. S3). As a result, the *OCT4A* mRNA levels in Ishikawa cells were much lower than those in PA1 cells (1/1000-10000) and the levels in HEC50B and A549 cells, established from high malignant cancer, were around 1/250-500 of the amount in PA1 cells. In the case of normal tissue-derived cell lines (ARPE-19, HFF, HUVEC, and HAoSMC) and three cancer cell lines (MCF7, TTA1, and S2), *OCT4A* transcripts were not detected. These sequencing results confirmed that all RT-PCR products were specifically amplified from

bona fide OCT4A transcripts (Table S2). These results indicate that the genuine *OCT4A* gene is undoubtedly transcribed at various levels in a variety of human cancer cell lines.

In addition to *OCT4A*, two novel *OCT4A* splicing variants carrying an additional exon were identified (Fig. 2C). Alignment with the human genomic sequence revealed the presence of additional exons in the human consensus genome. One novel transcript variant, designated *OCT4A1* (GenBank accession number LC006945), contained an additional exon (exon 1c, 118 bp), and another, designated *OCT4A2* (GenBank accession number LC006944), retained intron 2 (233 bp). *OCT4A1* transcripts were detected in HeLa, HEC265, HEC1, and HEC50B cells, and *OCT4A2* transcripts were detected in HEC1 and A549 cells (Table S2).

Another specific primer set, SetB, was designed to avoid detection of *OCT4A* and pgs while detecting possible *OCT4B* splice variants (Fig. 2A). In PA1 cells, four transcripts of distinct sizes were detected (Fig. 2B), suggesting the existence of novel splicing variants other than *OCT4B* and *OCT4B1*. Using BLAST, we determined that bands at 995 bp and 1228 bp corresponded to the previously detected *OCT4B* and *OCT4B1*, respectively. Sequencing analysis revealed three novel splicing variants, *OCT4B2*, *OCT4B3,* and *OCT4Bns* (GenBank accession numbers LC006946, LC006948, and LC006947; 1512 bp, 1279 bp, and 1774 bp, respectively) (Fig. 2C).

Similarly, more than three *OCT4B* splicing variants were detected in HeLa, Ishikawa, HEC265, HEC1, HEC50B, A549, and PA1 cells by gel electrophoresis. In other cell lines (MCF7, TTA1, S2,

HEK293T, ARPE-19, HUVEC, and HAoSMC), only *OCT4Bns* was detected by gel electrophoresis (Fig. 2B and Table S2). Based on these results, the expression of multiple *OCT4B* variants was correlated with *OCT4A* expression in human cancer cells, except for HEK293T cells transformed by expression of the large T antigen from the SV40 virus.

Verification of possible translation of OCT4 in tumor cells using the tagged genomic transgene

To confirm possible OCT4 translation in cancer cells without false-positive signals, the FLAG-tagged genomic transgene (pOCT4Gen-FLAG, Fig. S2) was introduced for detection of the protein products. We confirmed that transcripts derived from pOCT4Gen-FLAG also mimic the endogenous splicing variants (Fig. S4). Cells immunoreactive with the Flag antibody were detected in minor populations of HEC50B and A549 cells (Fig. 3A and 3B). The number of immunostained cells to RFP-positive cells was higher in HEC50B (126/1789, 7.0%) and A549 (48/1577, 3.0%) cells than in Ishikawa cells (4/1388, 0.3%) (Fig. 3A and 3B), which seemed to be correlated to endogenous transcription levels. To clarify the translated products with higher sensitivity, we performed direct detection by SDS-PAGE after transfection of Nluc-tagged genomic transgene into HEC50B, A549 and PA1 cells. A protein product of 531 amino-acids, which was the estimated size of the OCT4A-Nluc fusion protein, was recognized in HEC50B, A549, and positive control PA1 cells (Fig. 3C). These results indicate that the translation OCT4A protein occurs at least in a small population of human cancer cells. Products other than OCT4A were not detected in

HEC50B and A549 cells.

Identification of isoforms encoded by human OCT4 transcript variants induced by enforced expression

To investigate whether other transcript variants encode proteins, the coding sequences were cloned into an enforced expression vector with a FLAG tag located at its N- or C-terminus, followed by transfection into COS7 cells. Each DNA construct is shown in Fig. S2. To avoid problems associated with antibody specificity, we used an anti-FLAG antibody for detection of OCT4 isoforms by western blot (Fig. 4A, 4C and S5). In addition, 5′-RACE analysis confirmed the RNA transcript variant type detected within the transfected cells (Fig. 4B).

Irrespective of the FLAG-tag, transfection of both *OCT4A* and *OCT4A2* expression vectors resulted in the expression of ~49-kDa proteins, similar to the size of the OCT4A protein (A360) (Fig. 4A). In addition, sequencing results from 5′-RACE revealed a PCR product, consistent with *OCT4A* in the both case of *OCT4A* and *OCT4A2* over-expression (Fig.4B). It means that the sequence of intron 2 in *OCT4A2* mRNA was further spliced in cells overexpressing *OCT4A2* transcripts. By contrast, no OCT4A1 translation products were detected when the construct expressing the FLAG-tag at the C-terminus was used for transfection. Sequencing results from 5′-RACE revealed a longer PCR product, consistent with full-length *OCT4A1* sequences, and a shorter PCR product missing both exon 1 and exon 1C from *OCT4A1* transcripts (designated *OCT4ΔE1*). When FLAG was added to the *OCT4A1* N-terminus, a ~24-

kDa product was detected, suggesting that acquisition of a new in-frame UGA terminal codon within the novel exon 1c resulted in a 504-nt open reading frame (ORF) predicted to encode a truncated 168-amino-acid peptide (A168; Fig. 4Da). Amino acids 1 through 136 were identical between the OCT4A1 and OCT4A proteins, including a similar N-terminal domain (N-TD). However, OCT4A1 largely lacked the rest of the N-TD, the POU-specific domain, the POU-homeodomain, and the C-terminal transactivation domain (C-TD). Currently, it remains unknown whether OCT4A1 protein possesses some function. In *OCT4ΔE1*, the out-of-frame AUG located in exon 2 (E2-AUG) encoded the first methionine, and the terminal codon in exon 4 resulted in translation products predicted to contain 77 amino acids, lacking the FLAG-tag in-frame (Fig. 4Da). Moreover, overexpression of the *OCT4ΔE1* construct induced B164-protein production (Fig. 4C, *lane 19*). When E2-AUG was replaced with AGG, production levels increased substantially (Fig. 4C, *lane 20*).

Surprisingly, transient transfection of each *OCT4B* variant resulted in detection of two major PCR products (Fig. 4B). The long PCR product was the full-length *OCT4B* transcript, and the smaller product was an alternative transcript with a partial deletion (86 bp) of exon 1b, termed *OCT4BΔ86* (Fig. 4B). Our experiments confirmed that further splicing reactions occurred when the *OCT4A* and *OCT4B* expression vectors were expressed in COS7 cells.

Western blot analysis of *OCT4B* variant constructs in COS7 cells revealed major translation products of the B164 isoform from all transcript variants (Fig. 4C, *lanes 1–6, 9, 12, and 13*). The B265

translation product was clearly observed when *OCT4B* was expressed, but only very low levels were detected when other *OCT4B* variants (*OCT4B1* to *OCT4Bns*) were expressed (Fig. 4C, *lanes 1–6*). These results were consistent with the translation products expected based on 5′-RACE results (Fig. 4B). Overexpression of the *OCT4BΔ86* construct resulted in B164 translation product expression, and the levels did not increase, even when E2-AUG was replaced with an AGG codon (Fig. 4C, *lanes 7 and 8*, 4Db).

An SNP (ATG or AGG) was reported at the first AUG codon in *OCT4B* (25). This polymorphism putatively inhibits B265 translation from *OCT4B* transcripts. Here, we investigated *OCT4* polymorphisms in each cancer cell line (Table S1). In some cell lines, the AGG codon was confirmed instead of the AUG codon. To investigate predicted translation products from the *OCT4B* and *OCT4BΔ86* transcripts presenting the AGG codon, overexpression of these constructs was performed. Transfection of these constructs mainly resulted in B164 product expression (Fig. 4C, *lanes 14–16, 21, and 22*).

Alternative translation products from CUG codons, such as B190 and B201, in *OCT4B* variants were barely detected after *OCT4B* transfection (28). B190 and B201 translation product levels were lower than those of B265 or B164 products, whereas the products clearly appeared in addition to the B265 products along with disappearance of the B164 products when each CUG codon was converted to an AUG codon by site-directed mutagenesis (Fig. 4C, *lanes 10 and 11*). These observations strongly suggest that the major translation product from each *OCT4B* variant (*B1* to *Bns*) was OCT4B164, although the

protein levels were much lower than those of OCT4A (when comparing results shown in Fig. 4A and 4C).

Additional experimental data and our hypothesis for OCT4B translational control are shown in Fig. S5. The details are described in the Discussion.

Human OCT4C protein (B164) exhibits transformation activity in NIH-3T3 cells

We previously showed that mouse OCT-3/4C exhibits transformation activity in NIH-3T3 cells (38). *OCT4B* variant transcripts are translated into the B164 protein, and the human OCT4B164 isoform is highly homologous to the mouse OCT-3/4C isoform; therefore, the transforming activity of human OCT4B164, designated human OCT4C, was examined using normal NIH-3T3 fibroblasts. Localization of the human OCT4C protein was assessed in NIH-3T3 cells, revealing that the EGFP-OCT4C fusion protein was mainly located in the nucleus of NIH-3T3 cells, similar to EGFP-OCT4A (Fig. 5A), whereas EGFP-B265 was localized in the cytoplasm (data not shown). Localization of EGFP-OCT4C in the nucleus was confirmed by disruption of the nuclear localization signal by site-directed mutagenesis (EGFP-OCT4CΔNLS), which resulted in OCT4CΔNLS cytoplasmic localization (Fig. 5B). The stable expression of human OCT4C induced the transformation of NIH-3T3 cells (Fig. 5C). Soft-agar colony formation assays indicate that OCT4C and OCT4A overexpression had similar effects, whereas OCT4CΔNLS presented no transformation activity (Fig. 5C). These data indicate that OCT4C nuclear localization is necessary to induce the transformation of normal fibroblasts.

Effects of ablating OCT4A-positive cells in human cancer cells

To clarify the role of OCT4-positive cancer cells, we attempted to ablate OCT4-positive cancer cells using the pOCT4-DTA construct (Fig. 6A). In PA1 cells as a control, DTA-induced cell death was largely observed upon pOCT4-DTA transfection. In contrast, in HEC50B cells, cellular morphology and live cell numbers were not significantly changed after transfection of pOCT4-DTA compared with that after transfection of pOCT4-GFP (Fig. 6A). To confirm whether pOCT4-DTA can be effective in a small OCT4-positive population, we measured Nluc activity upon co-transfection of pOCT4Gen-Nluc and pOCT4-DTA (Fig. 6B). We detected a significant decrease of Nluc activity, suggesting that pOCT4-DTA specifically induced cell death in the OCT4-positive small population in HEC50B cells.

Finally, we investigated the effects of cell ablation of OCT4-positive cells using HEC50B cells, which are known as a highly malignant tumor cell line. OCT4-positive cell ablation by pOCT4-DTA caused a significant decrease in migration and Matrigel invasion (Fig. 6C). In the migration assay, the number of migrated cells was 586 (2.3%) and 67 (0.27%) for 2.5 x 10^4 seeded cells with pOCT4-GFP and pOCT4-DTA transfection, respectively. In the invasion assay, the number of invaded cells was 130 (0.52%) and 34 (0.13%) for 2.5 x 10^4 seeded cells with pOCT4-GFP and pOCT4-DTA transfection, respectively. These results indicate that the OCT4-positive cell population plays an important role in the cell migration and invasion of HEC50B malignant tumor cells.

Discussion

Although human OCT4 expression has been characterized in studies of stem cells and cancer biology, the data are controversial. Here, we developed a method to effectively analyze *OCT4* expression. Despite advising caution in OCT4 analysis, inconsistent results from the assessment of OCT4 expression continue to be reported (19,20,24). We verified the specificity of *OCT4* PCR primers, finding that almost all previously used primer sets contained nts common to or mismatched with *OCT4* pgs (Fig. S1). One primer set targeting the 3′ untranslated region (UTR) is frequently used to detect endogenous *OCT4* gene activation involved in the reprogramming of differentiated cells into induced pluripotent stem cells (40). These primer sequences are not specific, but match completely with *OCT4-pg1*, and despite potential false-positive amplification of *OCT4-pg1*, this primer set remains widely used for human *OCT4* expression analyses. Among the primer sets tested, only that previously used by Suo et al. (18) excluded pg amplification, although it could not eliminate amplification of contaminating genomic DNA. Total RNA is routinely pretreated with DNase I to exclude genomic DNA, but this does not completely eliminate genomic DNA contamination (41). Consequently, primers must be designed to discriminate specific PCR products from amplified genomic DNA and cDNA to accurately and reliably examine *OCT4* expression. To eliminate pg amplification, we designed a new forward PCR primer containing a unique and specific sequence at the 5′ UTR. By contrast, it was challenging to design unique and specific reverse

primers that anneal to the genuine human *OCT4* sequence. Ultimately, we designed a reverse primer not matching intron sequences, thereby eliminating amplification of contaminating genomic DNA. This primer can also be used to detect and isolate unknown *OCT4* splicing variants. We confirmed that our newly developed primer sets (Fig. 1 and Table 1) excluded false-positive amplifications.

Using specific primer SetA1, we provided evidence of the presence of multiple *OCT4* transcript variants, including two novel variants (*A1* and *A2*), in human cancer cells. Based on our recent and current data, we recognized the necessity for properly re-examining previous studies of human *OCT4* expression. The BLAST analysis used to confirm positive data cannot distinguish multiple variants and/or genomic DNA, especially for short sequences. By contrast, critical evidence of a lack of *OCT4* expression in somatic cancer cells (21,22) could be explained by dominant amplification of highly expressed pgs rather than genuine *OCT4* in PCR analyses using conventional primers. Our simple RT-PCR method allows the accurate detection of human *OCT4* transcript variants. Reinspection and additional rigorous data obtained using this method are needed to understand the role of *OCT4* in human CSCs.

Moreover, we carefully tested the *OCT4* expression levels in human cancer cells by semi-quantitative RT-PCR. *OCT4A* is highly expressed and regulates pluripotency and self-renewal in pluripotent ES or EC cells. In this study, the PA1 ovarian teratocarcinoma cell line was used as OCT4A-positive EC cells. When compared with PA1 cells, A549 and HEC50B cells totally contained 1/250-500 of the amount of *OCT4A* mRNA at most (Fig. S3). Thus, the total expression levels of *OCT4A* in cancer

cell lines were not so high but not extremely low because our analysis suggested that only a small subpopulation has significant expression of *OCT4A* (Fig. 3).

To address the possibility of OCT4 translation, we detected FLAG- or Nluc-tagged OCT4 proteins expressed by the OCT4 regulatory regions in several cancer cells. In HEC50B and A549 cells, which definitely express OCT4, cells positive for nuclear OCT4 staining were identified more often than in Ishikawa cells, which reflected the respective mRNA levels (Fig. 3 and S3). Moreover, we confirmed that the OCT4A isoform was translated by in-gel detection of Nluc in these cancer cell lines. As described above, we demonstrated the possibility of OCT4A translation using the transgene. Furthermore, to confirm endogenous OCT4A protein expression more directly, we utilized a modular Nluc reporter construct containing six concatenated repeats of the biding motif for OCT4A (PORE) (Fig. S6). Bioluminescence images were obtained from living cells after transient transfection of this reporter gene (pPORE-Nluc) under a LV200 bioluminescence microscope (Fig. S6A). Co-transfection of pOCT4-DTA with pPORE-Nluc caused a significant decrease of Nluc-positive cells and also luciferase activity according to the DNA amount (Fig. S6A and S6B). From these results, we concluded that OCT4A proteins are indeed translated in a subpopulation of HEC50B cells, which were established from poorly differentiated endometrial cancer, classified as a highly malignant cancer. These results raised again the possibility that the frequency of OCT4A-positive cells is related to the malignancy of cancer. Recent studies reported that overexpression of OCT4A enhanced migration and/or invasion capability in

medulloblastoma, oral squamous cell carcinoma and malignant melanoma cells (42-44). Similarly, we confirmed that overexpression of OCT4A in HEC50B cells caused a 2–3-fold increase in migration and invasion capability (data not shown).

Notably, it has been thought that CSCs may be intrinsically migratory and/or invasive (45-48). In this study, we developed a method to mimic the visualization of endogenous OCT4 protein translation in a minor population of cancer cells and ablated these cells. As a result, the migration and invasion activities, thought to be caused by CSCs, were definitively suppressed. Using this method for tracking and analysis of OCT4-positive cells, the involvement of OCT4 with CSCs will be clearer. Based on the data presented here, no normal tissue-derived cells expressed *OCT4A*. Therefore, OCT4 might truly play important roles in malignancy, especially in migration and invasion, at least in some types of human somatic cancer. Accumulation of convincing data by a correct method will clarify the function of OCT4 and its significance as a prognostic and predictive biomarker in human malignant tumors.

Novel splicing variants have been also identified for *OCT4B* transcripts. The existence of multiple *OCT4B* transcripts suggests that previous human *OCT4B*-expression results are insufficient based on the inability of the primers to identify or discriminate among multiple transcript variants. Although several studies reported *OCT4B1* expression in some tumor cells (29,49), new analyses are required because the designed primers containing sequences in intron 2 can exclude *OCT4B*, but not *OCT4B2* and *OCT4Bns*.

Based on our RT-PCR results, multiple *OCT4*-transcript variants are expressed in human cancer cells, whereas no or few transcript variants were detected in normal tissue-derived differentiated cells. These data suggest a correlation between the expression of *OCT4A* and *OCT4B* variants other than *OCT4Bns* in human cells. We did not observe a similar correlation previously reported in mouse postnatal somatic tissues (38). These findings suggest a substantial difference in the mechanism of transcriptional regulation of *Oct4* between humans and mice.

To identify other possible translated proteins, we examined FLAG-tagged protein products after their *in vitro* enforced expression. Unexpectedly, the expression of *OCT4A2* and *OCT4B* variants in COS7 cells revealed that *OCT4* transcripts underwent further splicing and became *OCT4A* and *OCT4B* or *OCT4BΔ86* transcripts. *OCT4A1* overexpression in COS7 cells also produced truncated translation products caused by the recognition of exon 1 and exon 1c as introns. A similar observation was reported for *OCT4B1* constructs transfected in human bladder cancer cells (50). In that case, exon 2b (Intron 2) of *OCT4B1* mRNA was further spliced into *OCT4B* mRNA. Therefore, *OCT4A* and *OCT4B* variant mRNAs might undergo further aberrant splicing upon overexpression because all tested *OCT4* variants retained introns as a cryptic exon. To the best of our knowledge, an *OCT4BΔ86* transcript variant isolated from overexpression experiments has not been previously identified from each original cell line. As for mouse *Oct-3/4B*, our previous study identified a transcript variant type similar to human *OCT4BΔ86* in newborn mouse ocular tissues (38). Some of the mouse *Oct-3/4B* transcript variants showed a variety of splicing

sequences in the upstream region flanking exon 2. It is necessary to further investigate whether an *OCT4BΔ86* transcript variant or the human OCT4C protein is expressed in human cancer cells.

The *OCT4B* transcript is thought to produce B265, B190, and B164 by alternative translation when overexpressed in cultured cells (28). B190 is reportedly translated from a non-AUG (CUG) codon. In this study, we identified B265 and B164 protein products, but barely detected the B190 protein (Fig. 4C). The context of the AUG codon of B265 (cagAUGc) does not show an optimal Kozak consensus sequence (A/GccAUGG). Upon conversion to a strong Kozak consensus sequence (gccaccAUGg), we observed an increase in B265 products, but no increase in B164 products (Fig. S5A, *lane B-Kozak*; Fig. S5Bc). Our theory for this is presented in Fig. S5.

When the initiation codon is AUG in *OCT4B* transcripts (Baug; Fig. S5Ba), ribosomes dominantly initiate protein synthesis from the first AUG codon to produce only the B265 protein (Fig. S5Ba, *arrow with B265*). In *OCT4BΔ86* transcripts, the original initiating AUG codon frame of B265 is terminated at exon 3 before reaching the B164 initiation codon (Fig. S5Bb, *arrow with BΔ86stop*). In this case, the use of the E2 AUG codon might be suppressed by the overlapping BΔ86stop ORF (Fig. S5Bb, *arrow with BaugΔ86*). Therefore, ribosome reinitiation from the AUG codon in exon 3 results in B164 synthesis (Fig. S5Bb, *arrow with B164*). When the context of the exon 2 AUG codon (ccgAUGt) was converted to a strong Kozak context sequence (accAUGg), B164 translation product expression was dramatically decreased (Fig. S5A, *lane B-E2Kozak*) due to suppression of ribosome re-initiation from the

B164 AUG codon (Fig. S5Bd; *bold and dotted arrows with B164*), resulting in two distinct (B265 and B164) products detected from expression of the *OCT4B* construct (Fig. S5A). Zhang et al. identified the minimal sequence nt 201-231 of OCT4B IRES using the *Renilla* luciferase/firefly luciferase bicistronic reporter system (51). However, without eliminating possible spurious splicing (52), it cannot be concluded whether *OCT4B* mRNAs undergo stringent IRES activity. In any case, we could not observe any IRES activities in the putative IRES sites of *OCT4B* mRNA (Fig. S7), refuting the IRES hypothesis and proposing a cap-dependent translation-control system in human *OCT4B* mRNA variants (Fig. 4, S5).

If an AGG codon exists at the position of the *OCT4B* start site (Bagg; i.e., the rs3130932 SNP), two short upstream ORFs appear in exon 1b (Fig. S5Be, *arrows with up1 and up2*) before reaching E2 AUG, which is also likely used for translation initiation. In this case, the AUG codon in exon 3 was suppressed by the overlapping ORF from the E2 AUG start site (Fig. S5Be, *broken arrow with B164*), resulting in low B164 protein levels. However, changing the E2-AUG codon to AGG resulted in an increase in B164 protein levels (Fig. 4C, *lanes 14–16*). Similarly, in the case of *OCT4BaggΔ86* transcripts, E2 AUG became the first AUG codon and suppressed overlapping downstream AUG codons (Fig. S5Bf, *broken arrow with B164*), resulting in low B164 protein levels produced from the *OCT4Bagg* construct (Fig. 4C, *lanes 14–16*). Examination of transcript/translation products is required to obtain accurate information from overexpression studies.

In mice, we reported a transcript variant transcribed from the intron-3 promoter and encoding

the mouse OCT-3/4C isoform (38). A similar transcript has not been identified in humans, but further investigation is warranted, as the human OCT4B164 isoform sequence is similar to that of mouse OCT-3/4C. Here, we named this variant human OCT4C. Human OCT4C also exhibited transformation activity upon overexpression in normal fibroblasts. Disruption of OCT4C nuclear localization abrogated this activity, suggesting that it localizes to the nucleus and is capable of transforming normal NIH-3T3 cells, as observed upon *OCT4A* expression. These data suggest that OCT4A and OCT4C, rather than B265 or B190, are expressed in human tumor cells. However, a limitation of this study included limiting the analysis to particular human cancer cell lines; therefore, it remains unclear whether the OCT4 isoform is produced in human somatic tumors due to the lack of a specific antibody. Further investigations are required to evaluate human OCT4 isoforms encoded by multiple transcript variants.

Most mammalian genes are believed to generate multiple transcript variants and protein isoforms by alternative transcription and/or splicing. Inherited and acquired changes in pre-mRNA splicing play a significant role in human disease development, and many cancer-associated genes are regulated by alternative splicing (53,54). Our data indicate that *OCT4* also generates multiple transcripts in human cancer cells. Further investigation of major human *OCT4* transcript variants in normal and tumor tissues might be of diagnostic importance and provide potential drug targets. In addition, analyses of the splicing process, accurate characterization of *OCT4* splice variants, and determination of the roles of OCT4A and/or OCT4C isoforms might improve the current understanding of malignant

transformation. It might be important to clarify whether OCT4C protein is actually translated and functional in human cancer cells, even if in a very small amount.

To elucidate the true role of the human OCT4 gene in somatic cancer, the results of previous studies must be re-inspected and additional accurate data from rigorous studies must be collected. Our current work represents the first single step in this direction.

Conclusion

The indisputable PCR primer sets of the present study allowed highly specific and comprehensive analysis of human *OCT4*, removing all false-positives. The *OCT4* multiple transcripts -A, -B, and -B1 and five novel variants were identified in many cancer cell lines but scarcely in non-tumor cells. We demonstrated authentic OCT4A translation in a human cancer cell subpopulation, which might exacerbate cell migration and invasion. The primary possible proteins from *OCT4B* variants, if any, might be OCT4C (B164), which has transformation activity, suggesting an important role for human tumorigenicity as well as OCT4A. These findings provide specific experimental information to support further accurate analysis of human OCT4A and offer new insights into the unique function of OCT4C as well as OCT4A in cancer stem cells.

Acknowledgments

The authors wish to thank our laboratory members for helpful support and Editage (www.editage.jp) for English language editing. We are grateful to Shinichi Takeshita for technical assistance with the bioluminescence imaging.

Conflicts of interest

The authors declare that they have no conflicts of interest related to the content of this manuscript.

References

1 Jordan CT, Guzman ML, Noble M. Cancer stem cells. N Engl J Med 2006;355:1253–1261.

2 Clevers H. The cancer stem cell: Premises, promises, and challenges. Nat Med 2011;17:313–319.

3 Gupta PB, Chaffer CL, Weinberg RA. Cancer stem cells: Mirage or reality? Nat Med 2009;15: 1010–1012.

4 Medema JP. Cancer stem cells: the challenges ahead. Nat Cell Biol 2013;15;338–344.

5 Brehm A, Ovitt CE, Schöler HR. Oct-4: more than just a POUerful marker of the mammalian germline? APMIS 1998;106:114–126.

6 Nichols J, Zevnik B, Anastassiadis K, et al. Formation of pluripotent stem cells in the

mammalian embryo depends on the POU transcription factor Oct4. Cell 1998;95:379–391.

7 Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, de-differentiation or self-renewal of ES cells. Nat Genet 2000;24:372–376.

8 Boyer LA, Lee TI, Cole MF, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 2005;122:947–956.

9 Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006;126:663–676.

10 Kim JB, Zaehres H, Wu G, et al. Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. Nature 2008;454:646–650.

11 Kim JB, Sebastiano V, Wu G, et al. Oct4-induced pluripotency in adult neural stem cells. Cell 2009;136:411–419.

12 Pesce M, Wang X, Wolgemuth D, et al. Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. Mech Dev 1998;71:89–98.

13 Pesce M, Schöler HR. Oct-4: Control of totipotency and germline determination. Mol Reprod Dev 2004;55:452–457.

14 Gidekel S, Pizov G, Bergman Y, et al. Oct-3/4 is a dose-dependent oncogenic fate determinant. Cancer Cell 2003;4:361–370.

15 Cheng L, Sung MT, Cossu-Rocca P, et al. OCT4: Biological functions and clinical applications

as a marker of germ cell neoplasia. J Pathol 2007;211:1–9.

16 Rijlaarsdam MA, van Herk HA, Gillis AJ, et al. Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours. Br J Cancer 2001;105;854–863.

17 Bustamante-Marín X, Garness JA, Capel B. Testicular teratomas: an intersection of pluripotency, differentiation and cancer biology. Int J Dev Biol 2013;57;201–210.

18 Suo G, Han J, Wang X, et al. Oct4 pseudogenes are transcribed in cancers. Biochem Biophys Res Commun 2005;337:1047–1051.

19 Liedtke S, Enczmann J, Waclawczyk S, et al. Oct4 and its pseudogenes confuse stem cell research. Cell Stem Cell 2007;1:364–366.

20 Liedtke S, Stephan M, Kögler G. Oct4 expression revisited: potential pitfalls for data misinterpretation in stem cell research. Biol Chem 2008;389:845–850.

21 Zhao S, Yuan Q, Hao H, et al. Expression of OCT4 pseudogenes in human tumours: Lessons from glioma and breast carcinoma. J Pathol 2011;223:672–682.

22 Wezel F, Pearson J, Kirkwood L A, et al. Differential expression of Oct4 variants and pseudogenes in normal urothelium and urothelial cancer. Am J Pathol 2013;183:1128–1136.

23 Cantz T, Key G, Bleidissel M, et al. Absence of OCT4 expression in somatic tumor cell lines. Stem Cells 2008;26;692–697.

24 Warthemann R, Eildermann K, Debowski K, et al. False-positive antibody signals for the pluripotency factor OCT4A (POU5F1) in testis-derived cells may lead to erroneous data and misinterpretations. Mol Hum Reprod 2012;18:605–612.

25 Takeda J, Seino S, Bell GI. Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues. Nucleic Acids Res. 1992;20:4613–4620.

26 Atlasi Y, Mowla SJ, Ziaee SA, et al. OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. Stem Cells 2008;26:3068–3074.

27 Gao Y, Wang X, Han J. The novel OCT4 spliced variant OCT4B1 can generate three protein isoforms by alternative splicing into OCT4B. J Genet Genomics 2010;37:461–465.

28 Wang X, Zhao Y, Xiao Z, et al. Alternative translation of OCT4 by an internal ribosome entry site and its novel function in stress response. Stem Cells 2009;27:1265–1275.

29 Asadi MH, Mowla SJ, Fathi F, et al. OCT4B1, a novel spliced variant of OCT4, is highly expressed in gastric cancer and acts as an antiapoptotic factor. Int J Cancer 2011;128:2645–2652.

30 Mirzaei MR, Najafi A, Arababadi MK, et al. Altered expression of apoptotic genes in response to OCT4B1 suppression in human tumor cell lines. Tumour Biol 2014;35:9999–10009.

31 Katafigiotis S, Papamichos SI, Katopodi R, et al. A case-control study on the rs3130932 single nucleotide polymorphism in the OCT4B translation initiation codon in association with cancer state. Eur J

Cancer Prev 2011;20:248–251.

32 Venables JP. Aberrant and alternative splicing in cancer. Cancer Res 2004;64:7647–7654.

33 Sumithra B, Saxena U, Das AB. Alternative splicing within the Wnt signaling pathway: Role in cancer development. Cell Oncol 2016;39: 1–13.

34 Ferraiuolo M, Di Agostino S, Blandino G, et al. Oncogenic intra-p53 family member interactions in human cancers. Front Oncol 2016;6:77.

35 Yoshida A, Asaga T, Masuzawa C, et al. Alteration of tumorigenicity in undifferentiated thyroid carcinoma cells by introduction of normal chromosome 11. J Surg Oncol 1994;55:170–174.

36 Nordhoff V, Hübner K, Bauer A, et al. Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences. Mamm Genome 2001;12:309–317.

37 Yang HM, Do HJ, Oh JH, et al. Characterization of putative cis-regulatory elements that control the transcriptional activity of the human Oct4 promoter. J Cell Biochem 2016;96:821–830.

38 Mizuno N, Kosaka M. Novel variants of Oct-3/4 gene expressed in mouse somatic cells. J Biol Chem 2008;283:30997–31004.

39 Pain D, Chirn GW, Strassel C, et al. Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. J Biol Chem 2005;280: 6265–6268.

40 Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human

fibroblasts by defined factors. Cell 2007;131:861–872.

41 Panagopoulos I, Möller E, Isaksson M, et al. A PCR/restriction digestion assay for the detection of the transcript variants 1 and 2 of POU5F1. Genes Chromosomes Cancer 2008;47:521–529.

42 da Silva PBG, Teixeira Dos Santos MC, Rodini CO, et al. High OCT4A levels drive tumorigenicity and metastatic potential of medulloblastoma cells. Oncotarget 2017;8:19192–19204.

43 Tsai LL, Hu FW, Lee SS, et al. Oct4 mediates tumor initiating properties in oral squamous cell carcinomas through the regulation of epithelial-mesenchymal transition. PLoS One 2014;9:e87207.

44 Borrull A, Ghislin S, Deshayes F, et al. Nanog and Oct4 overexpression increases motility and transmigration of melanoma cells. J Cancer Res Clin Oncol 2012;138:1145–1154.

45 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–674.

46 Perlikos F, Harrington KJ, Syrigos KN. Key molecular mechanisms in lung cancer invasion and metastasis: a comprehensive review. Crit Rev Oncol Hematol 2013;87:1–11.

47 Nieto MA. Epithelial plasticity: a common theme in embryonic and cancer cells. Science 2013;342:1234850.

48 Ye X, Tam WL, Shibue T, et al. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. Nature 2015;525:256–260.

49 Asadzadeh J, Asadi MH, Shakhssalim N, et al. A plausible anti-apoptotic role of upregulated OCT4B1 in bladder tumors. Urol J 2012;9:574–580.

50 Farashahi Yazd E, Rafiee MR, Soleimani M, et al. OCT4B1, a novel spliced variant of OCT4, generates a stable truncated protein with a potential role in stress response. Cancer Lett 2011;309:170–175.

51 Zhang W, Wang X, Xiao Z, et al. Mapping of the minimal internal ribosome entry site element in the human embryonic stem cell gene OCT4B mRNA. Biochem Biophys Res Commun 2010;394:750–754.

52 Holcik M, Graber T, Lewis SM, et al. Spurious splicing within the XIAP 5′ UTR occurs in the Rluc/Fluc but not the betagal/CAT bicistronic reporter system. RNA 2005;11:1605–1609.

53 Cieply B, Carstens RP. Functional roles of alternative splicing factors in human disease. Wiley Interdiscip Rev RNA 2015;6:311–326.

54 Biamonti G, Catillo M, Pignataro D, et al. The alternative splicing side of cancer. Semin Cell Dev Biol 2014;32:30–36.

Figure legends

Figure 1. Verification of human *OCT4A*-specific PCR primer sets.

(A) Schematic diagram of the human *OCT4A* mRNA structure (E1–E5: exon 1 to exon 5). Black triangle ATG: position of the OCT4A translation initiation codon. The black lines show the names of various primer sets and the predicted PCR product sizes. Primer sequences are listed in Table 1.

(B) Gel electrophoresis for PCR products. Each sample contained 10 fg or 10 ag of DNA template. DNA marker, 100-bp ladder; N, negative control.

(C) Schematic diagram of the forward PCR primer position utilized in this study (*bold arrow line*) and previously reported PCR primers (18,19) (*dotted arrow lines*). Shaded sequences indicate the alignment of most 5′ homologous regions between *pg1* and *OCT4* sequences.

Figure 2. Analysis of human *OCT4* expression and a schematic diagram of the multiple transcripts in various cell lines.

(A) Schematic representation of the position of PCR primer sets utilized in this study. (a) Human *OCT4* genomic structure is illustrated. Open boxes with numbers indicate authentic exons, and bold lines indicate introns. (b) Human *OCT4* transcript structure is shown. Shaded box indicates retained introns. Empty and solid triangles indicate the positions of the first AUG codon and termination codon,

respectively. To amplify *OCT4A* and *OCT4B* transcripts, the SetA1-F and SetA1-R primer set and SetB-F and SetB-R primer set were utilized, respectively.

(B) Expression of human *OCT4* transcripts in various cell lines. Several cancer cell lines express both *OCT4A* and *OCT4B* transcripts, whereas none of the normal cell lines expresses *OCT4A* transcripts. Various cancer cell lines express at least one to four distinct variants, as shown by RT-PCR using primer SetB. RT^+ and RT^- indicate the presence or absence of reverse transcriptase treatment, respectively. As DNA size markers, a 1-kb ladder was used to assess the size of *OCT4A* and *OCT4B* PCR products, and a 100-bp ladder was used for *GAPDH*.

(C) *OCT4A*: authentic human *OCT4A* transcript structure. *OCT4A1* and *OCT4A2*: novel *OCT4A* transcript splicing variants. The shaded boxes with numbers indicate retained introns. *OCT4B* and *OCT4B1* are known variants. *OCT4B2*, *OCT4B3*, and *OCT4Bns* are novel *OCT4B* transcript variants.

Figure 3. Identification of OCT4-positive cells using the tagged genomic transgene in human cancer cell lines.

FLAG or Nluc-tagged genomic transgene was transfected into Ishikawa, HEC50B, and A549 cells. The constructs contain upstream regulatory regions of the OCT4 gene, including the CR1, CR2, CR3 and CR4 regions conserved among mammals.

(A) Immunocytochemistry using anti-FLAG antibody in pOCT4Gen-FLAG-transfected cells. Anti-FLAG and RFP indicate FLAG-tagged OCT4 protein-positive cells and vector-transfected cells, respectively. Number of FLAG-positive cells/RFP-positive cells was 4/1388 (0.3%), 126/1789 (7.0%) and 48/1577 (3.0%) in Ishikawa, HEC50B, and A549 cells. Scale bar; 50 μm.

(B) Merged images of immunocytochemistry. Anti-FLAG signal was identified in the nucleus. Scale bar; 50 μm.

(C) Direct detection by SDS-PAGE by transfection of Nluc-tagged genomic transgene. The image represents a 120-min exposure after the addition of substrate reagent. A band of 531 amino-acids, which was the estimated size of the OCT4A-Nluc fusion protein, was clearly detected (OCT4A-Nluc). The gel stained with Coomassie Brilliant Blue is shown under the panel as a loading control.

Figure 4. Possible translation products from human *OCT4* transcript variants.

Western blot analysis of FLAG-tagged OCT4 variants expressed in COS7 cells. Molecular weight markers are indicated on the right side of the panels.

(A) A360 represents the full-length OCT4A protein with 360 amino acids. A168 represents the N-terminal region of OCT4A1 with 168 amino acids.

(B) 5′-RACE results from *OCT4* variants expressed in COS7 cells. A 1-kb ladder was used as a DNA size marker. Details of *OCT4ΔE1* and *OCT4BΔ86* are depicted in (D).

(C) Western blot analysis of FLAG-tagged human *OCT4* variants and genetically modified *OCT4* variants expressed in COS7 cells. B265 represents of the full-length OCT4B protein with 265 amino acids. B164 represents the protein translated from the in-frame AUG codon present in exon 3 to produce a protein with 164 amino acids. The * indicates the position of the CUG codons present in exon 2 (note: four CUG codons are present in exon 2, which are numbered 1st to 4th, in order, from the 5′ region). The acrylamide gel stained with Coomassie Brilliant Blue is shown under the panel as a loading control.

(D) Schematic diagram of the expected protein products from *OCT4* constructs expressed in COS7 cells. (a) The empty and solid triangles indicate the positions of the AUG codon and termination codon, respectively. The *OCT4ΔE1* arrow initiated from exon 2 indicates the predicted translation from the AUG codon present in exon 2 suppressing translation from the exon 3 AUG codon. (b) The upstream ORF (upORF) is indicated by a narrow arrow. Dotted arrows indicate the out-of-frame ORF from the exon 2 AUG codon repressed by the overlapping ORF.

Figure 5. Cellular localization and transformation activity of OCT4C in NIH-3T3 cells.

(A) Localization of EGFP-OCT4A and -OCT4C in NIH-3T3 cells. Scale bar; 50 μm.

(B) A nuclear localization signal (NLS) is necessary for localization of the OCT4C protein in the nucleus. Disruption of the NLS in OCT4C results in localization of the fusion proteins in the cytoplasm (EGFP-OCT4CΔNLS).

(C) Soft-agar colony formation assay. Data represent the mean ± standard deviation (S.D.). OCT4C exhibits transforming activities equivalent to those of OCT4A. This transformation activity disappeared when OCT4C NLS was disrupted (CΔNLS).

Figure 6. Effects of ablating OCT4-positive cells in human cancer cells.

(A) Phase-contrast microscopic image of pOCT4-GFP- or pOCT4-DTA-transfected PA1 or HEC50B cells. pOCT4-GFP, pOCT4-DTA; *OCT4* 5′ upstream regulatory region drives EGFP or DTA (Fig. S2). In PA1 cells as control, pOCT4-DTA-induced cell death was largely observed. In HEC50B cells, cellular morphology and number of living cells were not significantly changed after transfection of pOCT4-DTA compared with that after transfection of pOCT4-GFP. Scale bar; 100 μm.

(B) Effects of pOCT4-DTA transfection on pOCT4Gen-Nluc activity in HEC50B cells. The DNA amounts (μg) of each construct per well are indicated. The quantitative data are presented as the mean ± S.D. Luciferase intensity was decreased by pOCT4-DTA in a concentration-dependent manner. Three independent experiments were performed and reproducibility was confirmed. *; $p < 0.005$, **; $p < 0.001$, Student's *t*-test.

(C) Transwell migration and invasion assay of HEC50B cells. After 24 hours of pOCT4-DTA or pOCT4-GFP transfection, 2.5×10^4 cells were seeded onto the top of the insert. After 72 hours of incubation at 37°C in a CO_2 incubator, the membranes were collected and stained with Giemsa solution.

The quantitative data are presented as the mean ± S.D. of the total number of migrated or invaded cells for $2.5x10^4$ seeded cells from six independent wells. Scale bar; 300 μm. *; $p < 0.01$, **; $p < 0.001$, Student's *t*-test.

Tables

Table 1. Primer sets used for PCR.

Gene		Sequence
OCT4A		
Set1	Forward	5'-AAGGCGGCTTGGAGACCTCTCAGCCTG-3'
	Reverse	5'-GGTTACAGAACCACACTCGGACCACAT-3'
Set2	Forward	5'-CCTCCCCGGAGCCCTGCACCGTCA-3'
	Reverse	5'-CAAAGCGGCAGATGGTCGTTTGGCTGAAT-3'
Set3	Forward	5'-CCTCCCCGGAGCCCTGCACCGTCA-3
	Reverse	5'-TGCTGGGCGATGTGGCTGATCTGCTGC-3'
Suo (2005)	Forward	5'-TCCCTTCGCAAGCCCTCAT-3'
	Reverse	5'-TGACGGTGCAGGGCTCCGGGGAGGCCCCATC-3'
Liedtke-1 (2006)	Forward	5'-AGCCCTCATTTCACCAGGCC-3'
	Reverse	5'-CAAAACCCGGAGGAGTCCCA-3'
Liedtke-2 (2006)	Forward	5'-GATGGCGTACTGTGGGCCC-3'
	Reverse	5'-CAAAACCCGGAGGAGTCCCA-3
Atlasi (2008)	Forward	5'-CTTCTCGCCCCCTCCAGGT-3'
	Reverse	5'-AAATAGAACCCCCAGGGTGAGC-3'
SetA1*	Forward	5'-AGAGAGGGGTTGAGTAGTCCCTTCGCA-3'
	Reverse	5'-CAAGAGCATCATTGAACTTCACCTTC-3'
SetA2	Forward	5'-AGAGAGGGGTTGAGTAGTCCCTTCGCA-3'
	Reverse	5'-TTTCTGCAGAGCTTTGATGTCCTGGGA-3'
OCT4B		
SetB*	Forward	5'-AGGCAGATGCACTTCTACAGACTATTC-3'
	Reverse	5'-CAAGAGCATCATTGAACTTCACCTTC-3'
GAPDH		
	Forward	5'-GCTTGTCATCAATGGAAATCCC-3'
	Reverse	5'-TTCACACCCATGACGAACATG-3'

* SetA1 and SetB can detect human OCT4 transcripts specifically and comprehensively.

OCT4, octamer-binding transcription factor 4; GAPDH, glyceraldehyde 3-phosphate dehydrogenase.