SPANDIDOS PUBLICATIONS

The *POU5F1P1* pseudogene encodes a putative protein similar to POU5F1 isoform 1

IOANNIS PANAGOPOULOS, EMELY MÖLLER, ANNA COLLIN and FREDRIK MERTENS

Department of Clinical Genetics, Lund University Hospital, SE-221 85 Lund, Sweden

Received February 22, 2008; Accepted May 29, 2008

DOI: 10.3892/or_00000105

Abstract. *POU5F1*, which encodes a transcriptional factor, has two alternatively spliced transcripts, 1 and 2, as well as six pseudogenes. Transcript 1 is considered to be a key regulator of cellular pluripotency and self-renewal. The *POU5F1* pseudogene, *POU5F1P1* on 8q24, encodes a protein with 95% homology with the isoform 1 of POU5F1. It is located 15 kbp downstream of the SNP rs6983267, which is strongly associated with an increased risk of prostate and colon cancer, and within the amplified region in a variety of human malignancies. The previous finding of expressed sequence tags suggests that *POU5F1P1* can be expressed. We showed that a putative POU5F1P1 protein is localized in the nucleus, acts as a transcriptional activator and regulates the expression in a similar way to the POU5F1 isoform 1. However, POU5F1P1 was a weaker activator than isoform 1 of POU5F1, possibly due to the amino acid substitutions.

Introduction

POU5F1 (also known as *OCT-4*, *OCT-3*, *OTF-3* or *OTF-4*) encodes a transcriptional factor of the POU family which specifically binds the conserved octamer motive ATTTG CAT and regulates genes containing this octamer within their enhancer or promoter regions. *POU5F1* has two alternatively spliced transcripts: a long (variant 1, NM_002701) and a short (variant 2, NM_203289) transcript, coding for isoforms 1 (OCT4A) and 2 (OCT4B) of POU5F1. The isoforms differ in their N-terminal ends but have identical POU DNA binding and C-transactivation domains (1). Of the two, isoform 1 is considered to be a key regulator of pluripotency as well as for the propagation of the mammalian germline (2,3) whereas isoform 2, localized in the cytoplasm, does not have the ability to confer self-renewal (4,5). For variant 1 of *POU5F1*, six pseudogenes have been reported, while pseudogenes for variant 2 have not been identified in the human genome (3,6). Suo *et al* (7) reported that two of the pseudogenes, mapped to chromosomal bands 10q21 and 8q24, were expressed in cancer cell lines as well as cancer tissues. Theoretically, the transcript from the pseudogene on 10q21 should not lead to the formation of a protein, whereas the expression of the pseudogene in 8q24 (official symbol *POU5F1P1*) can lead to the production of a protein with 95% homology to isoform 1 of POU5F1 (Fig. 1).

In the present study we investigated whether a putative protein encoded by *POU5F1P1* would have transactivational features and cellular localization similar to POU5F1 isoform 1.

Materials and methods

Vectors. The BAC clone RPCI11-327N12 (accession number AC016883) was purchased from Invitrogen and was used as a template for subsequent cloning experiments. PCR amplifications, ligations and additional auxiliary methods have been described in detail previously (8). The pBIND vector of the CheckMate Mammalian Two-Hybrid System (Promega) was used to construct products encoding fusions with the GAL4 DNA-binding domain. This vector expresses the *Renilla reniformis* luciferase, which allows the user to normalize the transfection efficiency. The *POU5F1P1* fragment was amplified using DQ486513-256FSalI (5'-GGG ATC CGT CGA CCC ATG GCG GGA CAC CTG GCT-3') and DQ486513-1316REagI (5'-ACC TGC GGC CGC TCA GTT TGA ATG CAT GGG AG-3') primers and was inserted in frame between the *Sal*I and *Eag*I restriction sites. The pBIND vectors with inserts coding for the full-length POU5F1 variant 1 and 2 proteins have been described elsewhere (9). For transcriptional activation through various DNA-binding sites, the GAL4 DNA-binding domain was removed by double digestion by *Nhe*I and *Sal*I, after which the plasmids were filled in with T4 DNA polymerase and self-ligated. The pBIND constructs with inserts for the full-length POU5F1 variant 1 and 2 but without the GAL4 DNA-binding domains, pGL4.10 with the *NANOG* proximal promoter, pGL4.13 with the *POU5F1* distal enhancer and pGL4.13-SV40 promoter, have been described in detail previously (9).

POU5F1P1 was amplified for subcellular localization from the clone RPCI11-327N12 using DQ486513-256FSalI and POU5F1wt1-1117RBamHI (5'-CGG TGG ATC CCG GTT TGA ATG CAT GGG AGA-3') primers, and was then cloned between the *Sal*I and *Bam*HI sites of a pEGFP-N1 vector (Clontech), in frame with cDNA coding for enhanced

Correspondence to: Dr Ioannis Panagopoulos, Department of Clinical Genetics, University Hospital, SE-221 85 Lund, Sweden
E-mail: ioannis.panagopoulos@med.lu.se

Key words: POU5F1 gene, pseudogene, POU5F1P1, stem cell, transactivator

```
           1                                                          60
NP_002692.2 MAGHLASDFA FSPPPGGGGD GPGGPEPGWV DPRTWLSFQG PPGGPGIGPG VGPGSEVWGI
ABF29403.1  .......... .......... ..W.A..... ..L..... .......... ..........

           61                                                         120
NP_002692.2 PPCPPPYEFC GGMAYCGPQV GVGLVPQGGL ETSQPEGEAG VGVESNSDGA SPEPCTVTPG
ABF29403.1  ........L. .......... .......... ......S... .......N.. .......P..

          121                                                         180
NP_002692.2 AVKLEKEKLE QNPEESQDIK ALQKELEQFA KLLKQKRITL GYTQADVGLT LGVLFGKVFS
ABF29403.1  .......... ....K..... .......... .......... .........I ..........

          181                                                         240
NP_002692.2 QTTICRFEAL QLSFKNMCKL RPLLQKWVEE ADNNENLQEI CKAETLVQAR KRKRTSIENR
ABF29403.1  .K........ .......... .......... .......... ......M... ..........

          241                                                         300
NP_002692.2 VRGNLENLFL QCPKPTLQQI SHIAQQLGLE KDVVRVWFCN RRQKGKRSSS DYAQREDFEA
ABF29403.1  .......... .........- .......... .......... .......... ..........

          301                                                         360
NP_002692.2 AGSPFSGGPV SFPLAPGPHF GTPGYGSPHF TALYSSVPFP EGEAFPPVSV TTLGSPMHSN
ABF29403.1  .......... ...P...... .......... .......... ...V...... I.........
```

Figure 1. Alignment of the sequence with accession number NP_002692.2, which corresponds to isoform 1 of POU5F1 with the putative POU5F1P1 protein with accession number ABF29403.1, encoded by the sequence with accession number DQ486513.

green fluorescence protein (EGFP). The pDsRed2-ER and pDsRed-monomer-actin vectors were obtained from Clontech.

Cell lines and transfection experiments. HeLa cells (DSMZ, ACC57) were cultured in RPMI-1640 medium, supplemented with fetal bovine serum (10%), 2 mM L-glutamine, 100 IU/ml penicillin and 100 μg/ml streptomycin. For transfections, 5000 cells (in 100 μl medium) were seeded in 96-well plates, and 24 h later they were transfected using the FuGENE HD transfection reagent:DNA at a ratio of 3:2 according to the company's recommendations (Roche Applied Science). To study the transcriptional activity, 400 ng of pBIND plasmid DNA (empty or with the various inserts) together with 1600 ng of the pG5*luc* (Promega) vector were used. For experiments concerning transcriptional activation through various DNA-binding sites, 400 ng of pBIND plasmid DNA without the GAL4 DNA binding domain together with 1600 ng of the pGL4.10 with the *NANOG* proximal promoter, pGL4.13 with the *POU5F1* distal enhancer or pGL4.13-SV40 promoter vector were used.

Cells were lysed 24 h after transfection in 40 μl of 1X passive lysis buffer (Promega) and assayed for the firefly and *Renilla* luciferase activities using the dual luciferase assay system (Promega), according to the manufacturer's instructions. The results were normalized against the *Renilla* luciferase activities. Measurements were performed with the Veritas 96-microplate luminometer (Turner Biosystems) using 20 μl cell lysate and 96-well medium binding Lumitrac 200 plates (Greiner bio-one). Two independent transfection experiments were performed with eight replicas for each construct. The results are presented as the median together with the 25th and 75th percentiles. The Mann-Whitney two-tailed test was used for the statistical analysis using the statistiXL software (http://www.statistixl.com).

Subcellular localization. For subcellular localization, 400,000 HeLa cells were plated in 2 ml medium in a 35-mm culture dish and 24 h later they were transfected with a 100 μl transfection complex containing 3 μl FuGENE HD transfection reagent, 1 μg pEGFP-POU5F1P1 plasmid DNA and

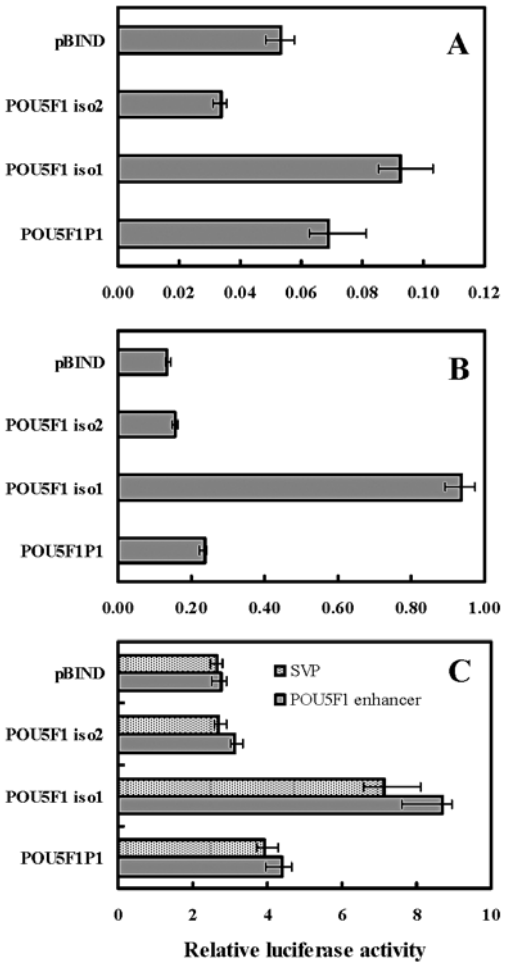

Figure 2. Transcriptional activation potential of *POU5F1P1* in HeLa cells. (A) cDNA fragments coding for POU5F1P1, POU5F1 isoforms 1 and 2, were inserted in frame into a pBIND vector and co-transfected with the pG5*luc* vector into HeLa cells. The fusions with the GAL4 DNA-binding domain activated the luciferase expression. (B) *POU5F1P1*, *POU5F1* transcripts 1 and 2, and empty vector (pBIND) were co-transfected with luciferase reporter constructs containing the *NANOG* proximal promoter. (C) The co-transfection of *POU5F1P1*, *POU5F1* transcripts 1 and 2, and empty vector (pBIND) with luciferase reporter constructs containing the SV40 promoter (SVP) or the *POU5F1* distal enhancer/SV40 promoter upstream of the reporter gene (*POU5F1* enhancer).

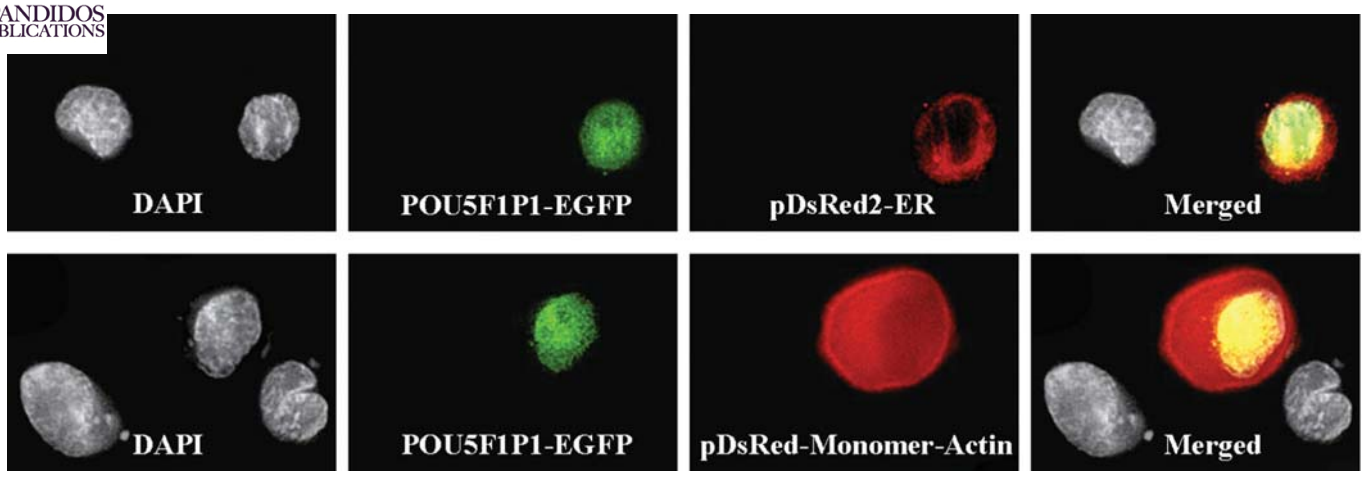

Figure 3. Images of epifluorescence microscopy showing the distribution of DAPI fluorescence (cell nucleus), EGFP fluorescence (POU5F1P1) and RFP fluorescence (pDsRed2-ER or pDsRed-monomer-actin vector). The protein encoded by *POU5F1P1* was localized to the nucleus.

1 μg pDsRed2-ER or pDsRed-monomer-actin vector. The cells were harvested 24 h after transfection, fixed in 4% formaldehyde solution, washed in 1X PBS and spread on a glass slide. The spreads were counterstained with 0.5 mg/ml DAPI (Sigma-Aldrich) and were visualized under an Axioplan 2 imaging microscope (Zeiss). Detailed information of the entire protocol is given elsewhere (8).

Results

To compare the transcriptional activation potentials of the putative POU5F1P1 and the two POU5F1 wt proteins, we constructed plasmids expressing GAL4 fusion proteins which were co-transfected with a luciferase reporter plasmid (pG5*luc*) containing GAL4 DNA-binding elements into HeLa cells. The transfected POU5F1P1 fused to GAL4 activated the reporter expression (Fig. 2A), indicating that it contains an activation domain. POU5F1P1 activated transcription less efficiently than POU5F1 isoform 1 but more strongly than POUF1 isoform 2 (P<0.001).

Subsequently, we assessed the effect of the putative POU5F1P1 protein, compared to the POU5F1 isoforms, on the expression of the luciferase reporter gene through regulatory elements. The pBIND-based constructs, with GAL4 removed, were co-transfected with constructs containing luciferase under the control of either the *NANOG* promoter or the *POU5F1* distal enhancer together with the SV40 promoter. In the two experiments, the constructs expressing POU5F1P1 or POU5F1 isoform 1 activated the luciferase expression compared to the control plasmid (empty pBIND) (Fig. 2B and C). The POU5F1 isoform 1 vector yielded the strongest activation of the three constructs, followed by POU5F1P1 and POU5F1 isoform 2 vectors. In the experiment with the *NANOG* promoter, POU5F1 isoform 1 activated the luciferase expression ~7 times more than the control (pBIND-vector) followed by POU5F1P1 (1.7 times stronger) and POU5F1 isoform 2 (1.2 times stronger). The co-transfection of SV40 promoter linked to the luciferase reporter gene with POU5F1 isoform 1 or POU5F1P1, resulting in a strong activation of SV40 promoter activity. Relative to the control plasmid pBIND, activation of the reporter expression was ~2.7 and 1.5 for POU5F1 isoform 1- and POU5F1P1-expressing contructs, respectively (P<0.001). No activation was found with the POU5F1 isoform 2. The insertion of the *POU5F1* distal enhancer upstream of the SV40 promoter further increased the activation. Relative to the control plasmid pGL4.13 (containing only the SV40 promoter), the increase was 1.2-fold for POU5F1 isoform 1 (P=0.002), 1.12 for POU5F1P1 (P=0.021) and 1.15 for POU5F1 isoform 2 (P=0.002), whereas no significant difference was found for the pBIND construct (P=0.254).

To investigate the subcellular localization of POU5F1P1, the pEGFP-N1 vector, containing the cloned transcript fused to the 5'-end of the EGFP, was co-transfected with pDsRed2-ER, and designed for fluorescent labeling of the ER in living cells, or the pDsRed-monomer-actin vector (RFP). The EGFP-signal was markedly detected within the nucleus whereas the RFP-signal was distributed outside the nucleus (Fig. 3). This finding indicated that the putative protein POU5F1P1 is located in the nucleus, similar to POU5F1 isoform 1, and not in the cytoplasm as POU5F1 isoform 2.

Discussion

Several aspects of the *POU5F1P1* pseudogene prompt further investigation. When comparing the DNA sequence of *POU5F1P1* and the mRNA sequence of *POU5F1* it can be concluded that *POU5F1P1* is a processed mRNA that has been inserted into the genome. Furthermore, *POU5F1P1* has a complete open reading frame and can, theoretically, encode a POU5F1 homolog protein. This protein contains the homeobox domain signature (PS00027) IAQQLGLEKDVV RVWFCNRRQKGK, the POU domain signature 1 (PS00035) KRITLGYTQADVG and the POU domain signature 2 (PS00465) SQTTICRFEALQLS, as well as the coiled-coil structures found in POU5F1 isoform 1. However, it has one amino acid residue less and 14 mismatches with isoform 1 of POU5F1 (Fig. 1). Moreover, its location in chromosome

band 8q24.21 makes *POU5F1P1* an attractive target for neoplasia-associated genetic aberrations. *POU5F1P1* is located ~15 kbp downstream of the SNP rs6983267 polymorphisms which are strongly associated with an increased risk of prostate (10-14) and colorectal cancer (13,15-17). In addition, *POU5F1P1* is found within the amplified region in a variety of human malignancies, including acute myeloid leukemia, ovarian cancer, medulloblastoma, osteosarcoma and chondrosarcoma (18-23). There is evidence that *POU5F1P1* may be expressed as expressed sequence tags (ESTs) from this pseudogene, of which five were found in colon tumors (GenBan accession Nos. BX108223, AA534908, AA515307, AA515288 and AA480280). Moreover, a sequence assigned as mRNA (accession No. DQ486513) was found in the breast cancer cell line MCF-7. Expression of *POU5F1P1* has been reported in cancers of the uterine cervix, lung, thyroid gland, esophagus, colon and urinary bladder, whereas no expression was found in normal skin, striated muscle or cultured fibroblasts, suggesting that *POU5F1P1* may play a role in oncogenesis (7). Furthermore, the expression of a mouse *pou5f1* pseudogene in mesenchymal stem cells was shown to promote cell proliferation and inhibit their osteochondral differentiation, reminiscent of the stem cell regulatory function of pou5f1 (24).

Our findings show that the putative POU5F1P1 protein displays features similar to isoform 1 of POU5F1. It localizes in the nucleus and is a transcriptional activator (Figs. 2 and 3). Moreover, the experiments with the *NANOG* proximal promoter and the *POU5F1* distal enhancer, both containing the conserved pou5f1/sox composite element known to bind POU5F1 (25,26), showed that POU5F1P1 regulates expression in a similar way to POU5F1 isoform 1. The SV40 promoter which drives the expression of the luciferase reporter gene has POU binding sites (27-29). Their presence explain the stronger promoter activity obtained with the co-transfection of POU5F1P1 or POU5F1 isoform 1. However, POU5F1P1 was a weaker activator than isoform 1 of POU5F1, perhaps due to the amino acid substitutions (Fig. 1).

POU5F1 reactivation has been found to be implicated in human tumors, and aberrant *POU5F1* expression was detected in various cancer and cancer cell lines (3,9,30). Bearing in mind the findings of the present study, and that non-specific primers (i.e, primers that do not distinguish transcripts from *POU5F1)* and pseudogenes were used in most studies (31), it is tempting to postulate that POU5F1P1, and not *POU5F1*, may be expressed in some tumors.

Acknowledgements

We thank Margareth Isaksson for her technical assistance. This work was supported by The Swedish Childhood Cancer Foundation, The Swedish Research Council and Gunnar Nilsson Cancer Foundation.

References

1. Takeda J, Seino S and 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 20: 4613-4620, 1992.
2. Pesce M and Scholer HR: Oct-4: control of totipotency and germline determination. Mol Reprod Dev 55: 452-457, 2000.
3. de Jong J and Looijenga LH: Stem cell marker OCT3/4 in tumor biology and germ cell tumor diagnostics: history and future. Crit Rev Oncog 12: 171-203, 2006.
4. Cauffman G, Liebaers I, Van Steirteghem A and Van de Velde H: POU5F1 isoforms show different expression patterns in human embryonic stem cells and preimplantation embryos. Stem Cells 24: 2685-2691, 2006.
5. Lee J, Kim HK, Rho JY, Han YM and Kim J: The human OCT-4 isoforms differ in their ability to confer self-renewal. J Biol Chem 281: 33554-33565, 2006.
6. Pain D, Chirn GW, Strassel C and Kemp DM: Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. J Biol Chem 280: 6265-6268, 2005.
7. Suo G, Han J, Wang X, Zhang J, Zhao Y and Dai J: Oct4 pseudogenes are transcribed in cancers. Biochem Biophys Res Commun 337: 1047-1051, 2005.
8. Panagopoulos I, Möller E, Isaksson M, Mandahl N, Vlamis-Gardikas A and Mertens F: Characterization of the native CREB3L2 transcription factor and the FUS/CREB3L2 chimera. Genes Chromosomes Cancer 46: 181-191, 2007.
9. Möller E, Stenman G, Mandahl N, *et al*: POU5F1, encoding a key regulator of stem cell pluripotency, is fused to EWSR1 in hidradenoma of the skin and mucoepidermoid carcinoma of the salivary glands. J Pathol 215: 78-86, 2008.
10. Savage SA and Greene MH: The evidence for prostate cancer risk loci at 8q24 grows stronger. J Natl Cancer Inst 99: 1499-1501, 2007.
11. Yeager M, Orr N, Hayes RB, *et al*: Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39: 645-649, 2007.
12. Zheng SL, Sun J, Cheng Y, *et al*: Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. J Natl Cancer Inst 99: 1525-1533, 2007.
13. Haiman CA, Le Marchand L, Yamamato J, *et al*: A common genetic risk factor for colorectal and prostate cancer. Nat Genet 39: 954-956, 2007.
14. Cheng I, Plummer SJ, Jorgenson E, *et al*: 8q24 and prostate cancer: association with advanced disease and meta-analysis. Eur J Hum Genet 16: 496-505, 2008.
15. Tomlinson I, Webb E, Carvajal-Carmona L, *et al*: A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 39: 984-988, 2007.
16. Tuupanen S, Niittymaki I, Nousiainen K, *et al*: Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. Cancer Res 68: 14-17, 2008.
17. Poynter JN, Figueiredo JC, Conti DV, *et al*: Variants on 9p24 and 8q24 are associated with risk of colorectal cancer: results from the Colon Cancer Family Registry. Cancer Res 67: 11128-11132, 2007.
18. Storlazzi CT, Fioretos T, Paulsson K, *et al*: Identification of a commonly amplified 4.3 Mb region with overexpression of C8FW, but not MYC in MYC-containing double minutes in myeloid malignancies. Hum Mol Genet 13: 1479-1485, 2004.
19. Storlazzi CT, Fioretos T, Surace C, *et al*: MYC-containing double minutes in hematologic malignancies: evidence in favor of the episome model and exclusion of MYC as the target gene. Hum Mol Genet 15: 933-942, 2006.
20. Staebler A, Karberg B, Behm J, *et al*: Chromosomal losses of regions on 5q and lack of high-level amplifications at 8q24 are associated with favorable prognosis for ovarian serous carcinoma. Genes Chromosomes Cancer 45: 905-917, 2006.
21. Rossi MR, Conroy J, McQuaid D, Nowak NJ, Rutka JT and Cowell JK: Array CGH analysis of pediatric medulloblastomas. Genes Chromosomes Cancer 45: 290-303, 2006.
22. Atiye J, Wolf M, Kaur S, *et al*: Gene amplifications in osteosarcoma-CGH microarray analysis. Genes Chromosomes Cancer 42: 158-163, 2005.
23. Morrison C, Radmacher M, Mohammed N, *et al*: MYC amplification and polysomy 8 in chondrosarcoma: array comparative genomic hybridization, fluorescent in situ hybridization, and association with outcome. J Clin Oncol 23: 9369-9376, 2005.
24. Lin H, Shabbir A, Molnar M and Lee T: Stem cell regulatory function mediated by expression of a novel mouse Oct4 pseudogene. Biochem Biophys Res Commun 355: 111-116, 2007.
25. Chew JL, Loh YH, Zhang W, *et al*: Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol Cell Biol 25: 6031-6046, 2005.

DJ, Chew JL, Lim LH, *et al*: Transcriptional regulation by OCT4 and SOX2. J Biol Chem 280: 24731-24737, 2005.
27. Gallo GJ, Gruda MC, Manuppello JR and Alwine JC: Activity of simian DNA-binding factors is altered in the presence of simian virus 40 (SV40) early proteins: characterization of factors binding to elements involved in activation of the SV40 late promoter. J Virol 64: 173-184, 1990.
28. Gruda MC and Alwine JC: Simian virus 40 (SV40) T-antigen transcriptional activation mediated through the Oct/SPH region of the SV40 late promoter. J Virol 65: 3553-3558, 1991.
29. Kilwinski J, Baack M, Heiland S and Knippers R: Transcription factor Oct1 binds to the AT-rich segment of the simian virus 40 replication origin. J Virol 69: 575-578, 1995.
30. Yamaguchi S, Yamazaki Y, Ishikawa Y, Kawaguchi N, Mukai H and Nakamura T: EWSR1 is fused to POU5F1 in a bone tumor with translocation t(6;22)(p21;q12). Genes Chromosomes Cancer 43: 217-222, 2005.
31. Panagopoulos I, Möller E, Isaksson M and Mertens F: A PCR/restriction digestion assay for the detection of the transcript variants 1 and 2 of POU5F1. Genes Chromosomes Cancer 47: 521-529, 2008.