Comparative integromics on JMJD2A, JMJD2B and JMJD2C: Preferential expression of *JMJD2C* in undifferentiated ES cells

YURIKO KATOH[1] and MASARU KATOH[2]

[1]M&M Medical BioInformatics, Hongo 113-0033; [2]Genetics and Cell Biology Section, National Cancer Center Research Institute, Tokyo 104-0045, Japan

Received April 11, 2007; Accepted May 21, 2007

Abstract. Fertilized egg or totipotent zygote undergoes cleavage divisions to form a blastocyst, consisting of outer trophoectoderm cells and inner cell mass with pluripotent primitive ectoderm cells. Epigenetic reprogramming, erasure and maintenance of epigenetic modification, occurs during early embryogenesis. In 2004, we identified and characterized JMJD2A/JHDM3A, JMJD2B, JMJD2C, JMJD2D, JMJD2E and JMJD2F. JMJD2A, JMJD2B and JMJD2C share the common domain architecture with JmjN, JmjC, two PHD, and two TUDOR domains. In 2006, other groups characterized JMJD2 family members as the H3K9 and/or H3K36 histone demethylases. Here, comparative integromics analyses on JMJD2A, JMJD2B and JMJD2C were carried out. Mouse *Jmjd2a* was expressed in fertilized egg and 2-cell embryos, while human *JMJD2A* was expressed in undifferentiated and differentiated ES cells. AP1-binding site and six bHLH-binding sites within intron 13 of human *JMJD2A* gene were conserved in mouse *Jmjd2a* gene. Mouse *Jmjd2b* was expressed in 8-cell embryos and undifferentiated ES cells, while human *JMJD2B* was expressed in undifferentiated and differentiated ES cells. Two GATA-binding sites within intron 6 of human *JMJD2B* gene were conserved in mouse *Jmjd2b* gene. Mouse *Jmjd2c* and human *JMJD2C* were preferentially expressed in undifferentiated ES cells. Four NANOG-binding sites, one TCF/ LEF-binding site, and one bHLH-binding site were located within evolutionary conserved region at the 3'-flanking region of human *JMJD2C* gene. NANOG-TCF/LEF-, and bHLH-binding sites within the 3'-flanking region of human *JMJD2C* gene were conserved in chimpanzee, cow, mouse and rat *JMJD2C* othologs. Together these facts indicate that JMJD2C is the evolutionarily conserved target of Homeo-domain transcription factor NANOG, and that JMJD2C is the histone demethylase implicated in the epigenetic reprogramming during the early embryogenesis.

Correspondence to: Dr Masaru Katoh, Genetics and Cell Biology Section, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
E-mail: mkatoh-kkr@umin.ac.jp

Key words: epigenetics, histone demethylase, JmjC domain, blastocyst, primitive ectoderm, embryonic stem cells, reprogramming, NANOG

Introduction

Fertilized egg or totipotent zygote undergoes about six cleavage divisions to form a blastocyst in the cell autonomous manner (1). Blastocyst consists of inner cell mass (ICM) with pluripotent primitive ectoderm cells, and outer trophoectoderm cells. After the implantation of a blastocyst, primitive ectoderm cells give rise to epiblast cells differentiating into three germ layers. Epigenetic reprogramming, erasure and maintenance of epigenetic modification, occurs during the preimplantation development (1).

Epigenetics is to study the genome information carried by chromatin rather than coding sequence (2). Nucleosome, consisting of 147 bp genomic DNA wrapped around the core histone octamer is the fundamental unit of chromatin (3-5). Core histones, such as H3, H4, H2A and H2B, are predominantly globular except for their N-terminal tail. Histone tail is modified through methylation, acetylation, phosphorylation, ubiquitylation, sumoylation, and ADP-ribosylation. Histone methylation at H3K4, H3K36 and H3K79 are implicated in transcriptional activation, while those at H3K9, H3K27 and H4K20 in transcriptional repression.

In 1995, Takeuchi *et al* cloned and characterized the mouse *Jumonji* (*Jmj*) gene encoding a transcriptional repressor with JmjN and JmjC domains, which is the founding member of the Jumonji family (6). Because JmjC domain is implicated in histone demethylase activity, Jumonji family proteins play a key role during embryogenesis and carcinogenesis through the regulation of chromatin structure and gene expression (7).

In 2003, we identified and characterized *JMJD1C/TRIP8* gene encoding JmjC domain protein (8), which share the common domain architecture with JMJD1A/TSGA (9) and JMJD1B/5qNCA (10). In 2006, Yamane *et al* reported JMJD1A, using their in-house name JHDM2A, as a histone H3K9 demethylase (11). JMJD1 family members are histone demethylase implicated in the transcriptional regulation of target genes for nuclear receptors.

In 2004, we identified and characterized *JMJD2* family genes (12). *JMJD2A* gene at human chromosome 1p34.2, *JMJD2B* gene at 19p13.3 and *JMJD2C* gene at 9p24.1, consisting of multiple exons, encode Jumonji (Jmj) domain

A

Mouse early embryo and ES cells	*Jmjd2a*	*Jmjd2b*	*Jmjd2c*
Fertilized egg	1	0	0
2-cell embryo	2	0	0
4-cell embryo	0	0	0
8-cell embryo	0	5	0
Morula	0	0	0
Blastocyst	0	0	0
ES cells	0	1	14

B

Human ES cells and derivatives	*JMJD2A*	*JMJD2B*	*JMJD2C*
Undifferentiated ES cells	3	5	2
Embryoid body	5	6	0
Hepatocyte progenitors	4	2	0
Neural progenitors	4	1	0

C

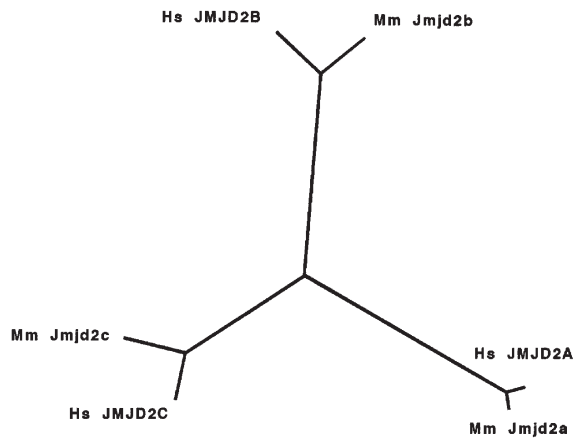

Figure 1. Comparative transcriptomics and proteomics on JMJD2 family members. (A), Expression profile of *Jmjd2a, Jmjd2b,* and *Jmjd2c* mRNAs during mouse early embryogenesis. (B), Expression profile of *JMJD2A, JMJD2B,* and *JMJD2C* in human ES cells and their derivatives. (C), Phylogenetic analysis on human and mouse JMJD2A, JMJD2B, and JMJD2C orthologs.

proteins with JmjN, JmjC, two PHD, and two TUDOR domains (12). *JMJD2D, JMJD2E* and *JMJD2F* genes, consisting of a single exon, are clustered at human chromosome 11q21, and encode proteins with JmjN and JmjC domains (12). In 2006, Chen *et al* characterized JMJD2A as histone demethylase for tri-methylated Lysine 9 and 36 on histone H3 (H3K9me3 and H3K36me3) (13), and Cloos *et al* characterized JMJD2C as histone demethylase for H3K9me3 and H3K9me2 (14). Here, comparative integromics analyses on JMJD2A, JMJD2B and JMJD2C were carried out to elucidate the JMJD2 family member implicated in the epigenetic reprogramming during early embryogenesis.

Materials and methods

In silico expression analyses. Expressed sequence tags (ESTs) derived from human and mouse *JMJD2* family genes were searched for using the BLAST programs as described previously (15,16). Human JMJD2A RefSeq (NM_014663.2), JMJD2B RefSeq (NM_015015.1), JMJD2C RefSeq (NM_015061.2), mouse Jmjd2a cDNA (AK129187.1), Jmjd2b RefSeq (NM_172132.1), and Jmjd2c RefSeq (NM_144787.1) were used as query sequences for the BLAST programs. The sources of human ESTs were listed up for *in silico* expression analyses.

Comparative proteomics analyses. Phylogenetic analyses on human and mouse JMJD2 family members were carried out using CLUSTALW program. Amino-acid sequences for human JMJD2A (NP_055478.2), JMJD2B (NP_055830.1), JMJD2C (NP_055876.2), mouse Jmjd2a (BAC97997.1), Jmjd2b (NP_742144.1), and Jmjd2c (NP_659036.1) were used for the phylogenetic analyses.

Comparative genomics analyses. Human genome sequences corresponding to *JMJD2* family genes were searched for using the BLAST programs as described previously (17,18). Exon-intron boundaries were determined based on the consensus sequence of exon-intron junctions ('gt ... ag' rule of intronic sequence) and codon usage within the coding region. Human genome sequences around the *JMJD2* family genes were then compared with chimpanzee, cow, mouse, and rat genome sequences to identify evolutionarily conserved regions. Binding sites for transcription factors, such as TCF/LEF, POU5F1, SOX2 and NANOG were then searched for as described previously (19,20).

Results

Expression profile of JMJD2 family members during early embryogenesis. Expression profiles of mouse *Jmjd2a, Jmjd2b, Jmjd2c* mRNAs (Fig. 1A) and those of human *JMJD2A, JMJD2B, JMJD2C* mRNAs (Fig. 1B) were investigated using *in silico* expression analyses. Mouse *Jmjd2a* mRNA was expressed in fertilized egg and 2-cell embryos, while human *JMJD2A* mRNA was expressed in undifferentiated and differentiated ES cells. Mouse *Jmjd2b* mRNA was expressed in 8-cell embryos and undifferentiated ES cells, while human *JMJD2B* mRNA was expressed in undifferentiated and differentiated ES cells. Mouse *Jmjd2c* and human *JMJD2C* mRNAs were preferentially expressed in undifferentiated ES cells.

Comparative proteomics on JMJD2 orthologs. Preliminary alignment of human and mouse RefSeq amino-acid sequences of JMJD2A, JMJD2B and JMJD2C orthologs revealed that mouse NM_172382.1 RefSeq encodes C-terminally truncated Jmjd2a protein. Because AK129187.1 cDNA was the representative mouse Jmjd2a cDNA, BAC97997.1 amino-acid sequence translated from AK129187.1 cDNA was used as the mouse Jmjd2a amino-acid sequence for the comparative proteomics analysis. Phylogenetic analysis revealed that human and mouse JMJD2A orthologs were relatively better conserved than JMJD2B and JMJD2C orthologs (Fig. 1C).

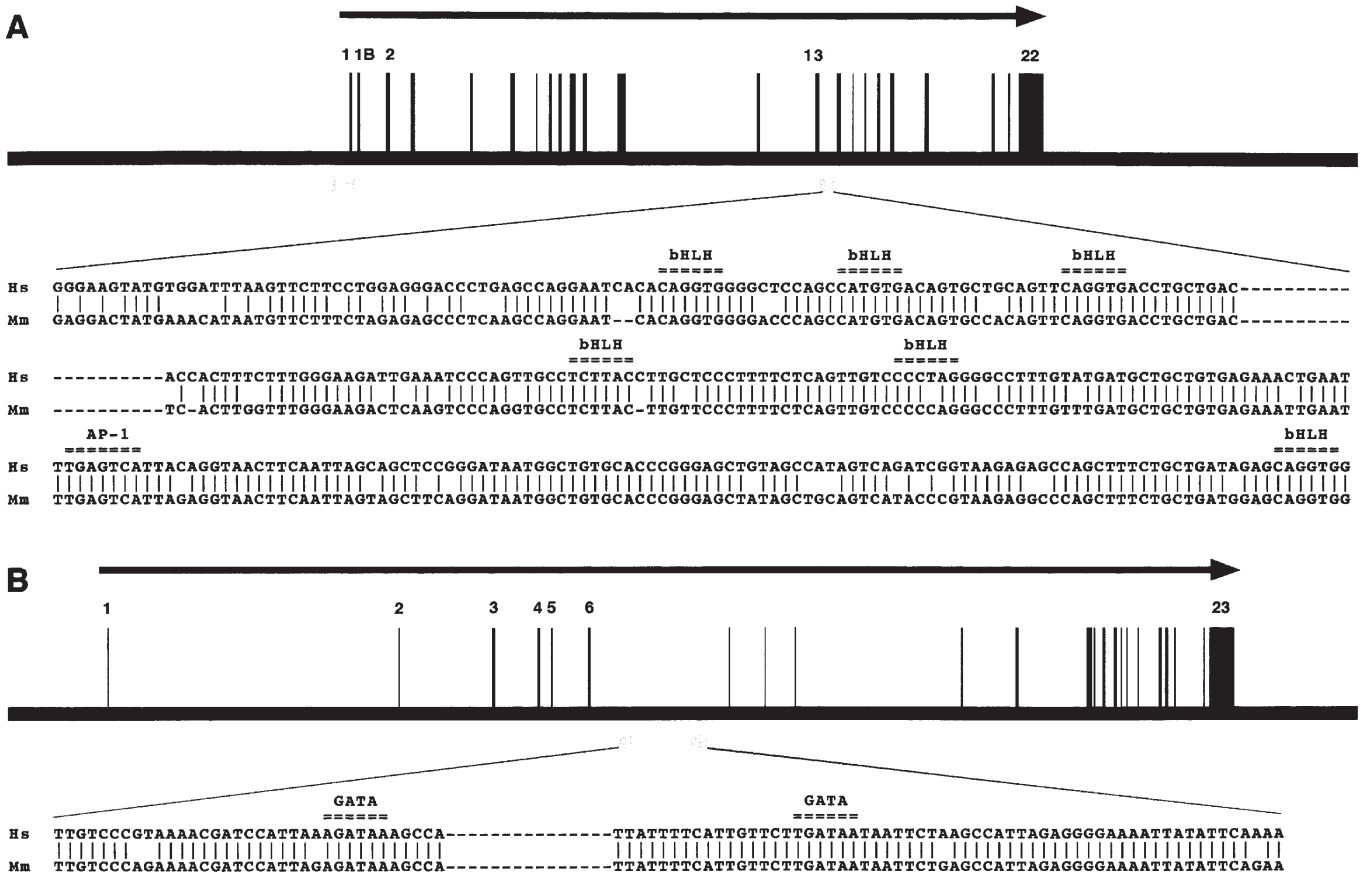

Figure 2. Comparative genomics on *JMJD2A* and *JMJD2B*. (A), Structure of human *JMJD2A* gene and evolutionarily conserved region. *JMJD2A* gene at human chromosome 1p34.2 consists of 23 exons. Conserved regions are shown by gray bars. AP1-binding site and six bHLH-binding sites within intron 13 of human *JMJD2A* gene are conserved in mouse *Jmjd2a* gene. (B), Structure of human *JMJD2B* gene and evolutionarily conserved region. *JMJD2B* gene at human chromosome 19p13.3 consists of 23 exons. Two GATA-binding sites within intron 6 of human *JMJD2B* gene are conserved mouse *Jmjd2b* gene.

Comparative genomics on JMJD2A. BLAST programs revealed that human *JMJD2A* gene was located within AC092815.2 and AL451062.12 genome sequences. Exons 1, 1B, and 2-12 were located within AC092815.2, and exons 12-22 within AL451062.12. Human *JMJD2A* gene, consisting of 23 exons, was about 55-kb in size (Fig. 2A).

Twenty-one JMJD2A ESTs were transcribed from exon 1, and three from exon 1B. *JMJD2A* major transcript consists of exons 1 and 2-22, while *JMJD2A* minor transcript consists of exons 1B and 2-22. Because initiator methionin is located within exon 2, JMJD2A splicing variants encode the same JMJD2A protein.

Human genome sequence around the *JMJD2A* gene was used for the comparative genomics analyses. Although exons 1 and 1B of the *JMJD2A* gene were not well conserved between human and mouse, 5'-promoter region, intron 1, exons 2-13, intron 13, and exons 14-22 were well conserved. AP1-binding site and two clusters of bHLH-binding sites within intron 13 of human *JMJD2A* gene were conserved in mouse *Jmjd2a* gene (Fig. 2A).

Comparative genomics on JMJD2B. BLAST programs revealed that human *JMJD2B* gene was located within AC053467.1, AC005595.1 and AC022517.1 genome sequences. Exon 1 was located within AC053467.1, exons 2-5 within AC005595.1, and exons 6-23 within AC022517.1. Human *JMJD2B* gene, consisting of 23 exons, was about 184-kb in size within NT_011255.14 human genomic contig (Fig. 2B).

Exon 2 was included in seven JMJD2B ESTs, but was spliced out in one JMJD2B EST. *JMJD2B* major transcript consists of exons 1-23, while *JMJD2B* minor transcript consists of exons 1 and 3-23. Because initiator methionin is located within exon 3, JMJD2B splicing variants encode the same JMJD2B protein.

Human genome sequence around the *JMJD2B* gene was used for the comparative genomics analyses. Although 5'-promoter region, exons 1 and 2 of the *JMJD2B* gene were not well conserved between human and mouse, exons 3-6, intron 6, and exons 7-23 were well conserved. Two GATA-binding sites within intron 6 of human *JMJD2B* gene were conserved in mouse *Jmjd2b* gene (Fig. 2B).

Comparative genomics on the JMJD2C. BLAST programs revealed that human *JMJD2C* gene was located within AL354707.17, AL445592.15, AL137020.13 and AL161443.13 genome sequences. Exons 1-4 were located within AL354707.17, exons 5-8 within AL445592.15, exons 9-19 within AL137020.13, and exons 20-22 within AL161443.13. Human *JMJD2C* gene, consisting of 22 exons, was about 407-kb in size within NT_008413.17 human genomic contig (Fig. 3).

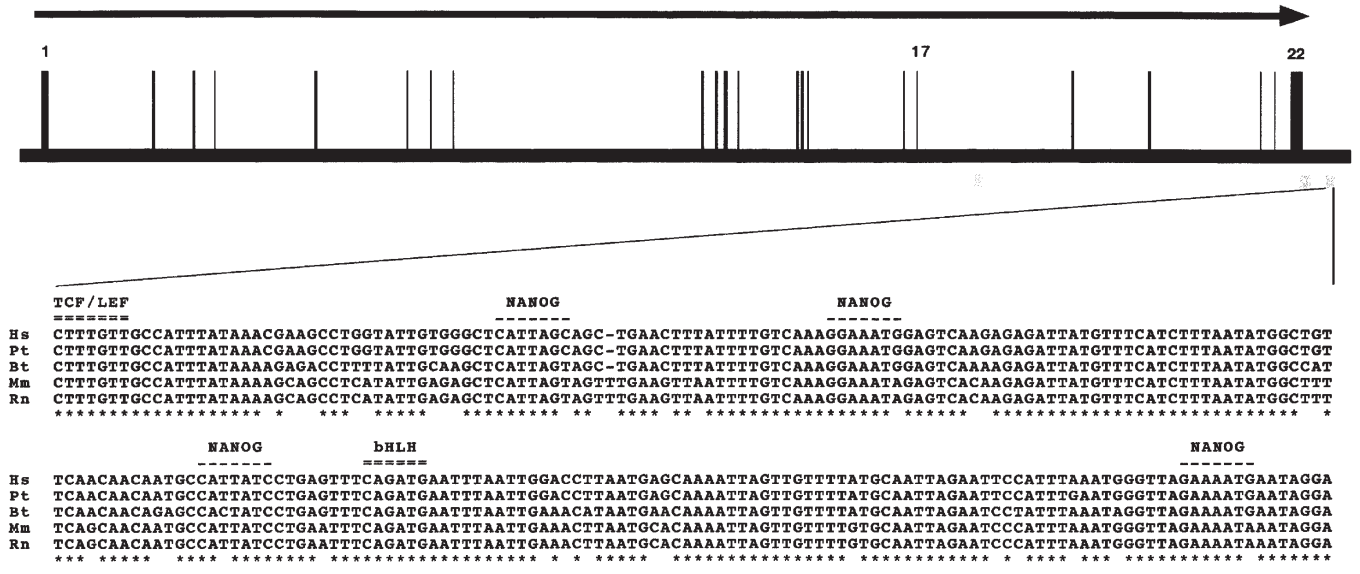

Figure 3. Comparative genomics on *JMJD2C*. *JMJD2C* gene at human chromosome 9p24.1 consists of 22 exons. Conserved regions are shown by gray bars. Four NANOG-binding sites, one TCF/LEF-binding site and one bHLH-binding site are located within evolutionary conserved region at the 3'-flanking region of human *JMJD2C* gene. NANOG- TCF/LEF-, and bHLH-binding sites within the 3'-flanking region of human *JMJD2C* gene are conserved in chimpanzee, cow, mouse and rat *JMJD2C* othologs.

Human genome sequence around *JMJD2C* gene was used for the comparative genomics analyses. Although 5'-promoter region of *JMJD2C* gene was not well conserved between human and mouse, exons 1-17, intron 17, exons 18-22, and 3'-flanking region were well conserved (Fig. 3). Four NANOG-binding sites, one TCF/LEF-binding site, and one bHLH-binding site were located within evolutionary conserved region at the 3'-flanking region of human *JMJD2C* gene. NANOG- TCF/LEF-, and bHLH-binding sites within the 3'-flanking region of human *JMJD2C* gene were conserved in chimpanzee, cow, mouse and rat *JMJD2C* othologs.

Discussion

Comparative integromics on JMJD2A, JMJD2B and JMJD2C were carried out in this study. Mouse *Jmjd2a* was expressed in fertilized egg and 2-cell embryos, while human *JMJD2A* was expressed in undifferentiated and differentiated ES cells. Mouse *Jmjd2b* was expressed in 8-cell embryos and undifferentiated ES cells, while human *JMJD2B* was expressed in undifferentiated and differentiated ES cells. Mouse *Jmjd2c* and human *JMJD2C* were preferentially expressed in undifferentiated ES cells (Fig. 1).

Human *JMJD2A* gene, about 55-kb in size, was found consisting of 23 exons. The 5'-promoter region, intron 1, exons 2-13, intron 13, and exons 14-22 of human *JMJD2A* gene were well conserved in mouse *Jmjd2a* gene. AP1-binding site and six bHLH-binding sites within intron 13 of human *JMJD2A* gene were evolutionarily conserved (Fig. 2A).

Human *JMJD2B* gene, about 184-kb in size, was found consisting of 23 exons. Exons 3-6, intron 6, and exons 7-23 of human *JMJD2B* gene were well conserved in mouse *Jmjd2b* gene. Two GATA-binding sites within intron 6 of human *JMJD2B* gene were evolutionarily conserved (Fig. 2B).

Human *JMJD2C* gene, about 407-kb in size, was found consisting of 22 exons. Exons 1-17, intron 17, exons 18-22, and 3'-flanking region of human *JMJD2C* gene were well conserved in mouse *Jmjd2c* gene. NANOG- TCF/LEF-, and bHLH-binding sites within the 3'-flanking region of human *JMJD2C* gene were conserved in chimpanzee, cow, mouse and rat *JMJD2C* othologs (Fig. 3).

POU5F1/OCT3/OCT4, SOX2, and NANOG are transcription factors expressed in pluripotent stem cells, such as ES cells and embryonic germ (EG) cells (1). POU5F1 and SOX2 are expressed in zygote, blastomere, blastocyst, and epiblast, while NANOG is transiently expressed in blastomere and blastocyst. Notably, epigenetic reprogramming occurs in the early embryogenesis during NANOG positive stages (1). Together these facts indicate that JMJD2C is the evolutionarily conserved target of Homeo-domain transcription factor NANOG, and that JMJD2C is the histone demethylase implicated in the epigenetic reprogramming during early embryogenesis.

References

1. Surani MA, Hayashi K and Hajkova P: Genetic and epigenetic regulators of pluripotency. Cell 128: 747-762, 2007.
2. Kouzarides T: Chromatin modifications and their function. Cell 128: 693-705, 2007.
3. Jenuwein T and Allis CD: Translating the histone code. Science 293: 1074-1080, 2001.
4. Bird A: DNA methylation patterns and epigenetic memory. Genes Dev 16: 6-21, 2002.
5. Turner BM: Defining an epigenetic code. Nat Cell Biol 9: 1-5, 2007.
6. Takeuchi T, Yamazaki Y, Katoh-Fukui Y, *et al*: Gene trap capture of a novel mouse gene, *Jumonji*, required for neural tube formation. Genes Dev 9: 1211-1222, 1995.
7. Takeuchi T, Watanabe Y, Takano-Shimizu T and Kondo S: Roles of Jumonji and *Jumonji* family genes in chromatin regulation and development. Dev Dyn 235: 2449-2459, 2006.
8. Katoh M and Katoh M: Identification and characterization of *TRIP8* gene *in silico*. Int J Mol Med 12: 817-821, 2003.
9. Hoog C, Schalling M, Grunder-Brundell E and Daneholt B: Analysis of murine male germ cell-specific transcript that encodes a putative zinc finger protein. Mol Reprod Dev 30: 173-181, 1991.

10. Hu Z, Gomes I, Horrigan SK, *et al*: A novel nuclear protein, 5qNCA (LOC51780) is a candidate for the myeloid leukemia tumor suppressor gene on chromosome 5 band q31. Oncogene 20: 6946-6954, 2001.
11. Yamane K, Toumazou C, Tsukada Y, *et al*: JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell 125: 483-495, 2006.
12. Katoh M and Katoh M: Identification and characterization of *JMJD2* family genes *in silico*. Int J Oncol 24: 1623-1628, 2004.
13. Chen Z, Zang J, Whetstine J, *et al*: Structural insights into histone demethylation by JMJD2 family members. Cell 125: 691-702, 2006.
14. Cloos PAC, Christensen J, Agger K, *et al*: The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. Nature 442: 307-311, 2006.
15. Katoh M: Paradigm shift in gene-finding method: from bench-top approach to desk-top approach. Int J Mol Med 10: 677-682, 2002.
16. Katoh M and Katoh M: Comparative genomics on mammalian *FGF3-FGF4* locus. Int J Oncol 27: 281-285, 2005.
17. Katoh Y and Katoh M: WNT antagonist, SFRP1, is Hedgehog signaling target. Int J Mol Med 17: 171-175, 2006.
18. Katoh M and Katoh M: CER1 is a common target of WNT and NODAL signaling pathways in human embryonic stem cells. Int J Mol Med 17: 795-799, 2006.
19. Katoh Y and Katoh M: Conserved POU-binding site linked to SP1-binding site within *FZD5* promoter: transcriptional mechanisms of *FZD5* in undifferentiated human ES cells, fetal liver/spleen, adult colon, pancreatic islet, and diffuse-type gastric cancer. Int J Oncol 30: 751-755, 2007.
20. Katoh M and Katoh M: Comparative integromics on FZD7 orthologs: conserved binding sites for PU.1, SP1, CCAAT-box and TCF/LEF/SOX transcription factors within 5'-promoter region of mammalian *FZD7* orthologs. Int J Mol Med 19: 529-533, 2007.