Comparative integromics on *JMJD1C* gene encoding histone demethylase: Conserved POU5F1 binding site elucidating mechanism of *JMJD1C* expression in undifferentiated ES cells and diffuse-type gastric cancer

MASUKO 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 March 22, 2007; Accepted April 30, 2007

Abstract. Epigenetic modifications of genomic DNA and histones alter the chromatin structure to regulate the accessibility of transcription factors to the promoter or enhancer regions. In 2003, we identified and characterized JMJD1C (TRIP8) consisting of TRI8H1 domain with C2HC4-type zinc finger-like motif, TRI8H2 domain with thyroid hormone receptor ß-binding region, and JmjC domain. JMJD1A (TSGA), JMJD1B (5qNCA) and JMJD1C with the common domain architecture are histone H3K9 demethylases implicated in the nuclear hormone receptor-based transcriptional regulation. Here, comparative integromics on *JMJD1C* gene is reported. JMJD1C variant 1, previously reported, consists of exons 1, 2 and 3-26, while JMJD1C variant 2 characterized in this study was transcribed from novel exon 1B located 5' to exon 3. Four human JMJD1C ESTs were transcribed from exon 1, while 14 human JMJD1C ESTs from exon 1B. All of 26 mouse Jmjd1c ESTs were transcribed from exon 1b. These facts indicate that JMJD1C variant 2 transcribed from exon 1B was the major transcript. Human JMJD1C variant 2 with TRI8H1, TRI8H2, and JmjC domains showed 85.7% total-amino-acid identity with mouse Jmjd1c. Human *JMJD1C* mRNA was expressed in undifferentiated embryonic stem (ES) cells, pancreatic islet, diffuse-type gastric cancer, and other tissues or tumors. Mouse *Jmjd1c* mRNA was expressed in fertilized egg, blastocyst, undifferentiated ES cells, embryonic germ cells, c-Kit$^+$/Sca-1$^+$/Lin$^-$ hematopoietic stem cells, pancreatic islet, and other tissues. Comparative genomics analyses revealed that binding sites for POU5F1 (OCT3/OCT4), AP-1, and bHLH transcription factors within the promoter region located 5' to exon 1B of human *JMJD1C* gene were conserved in chimpanzee, cow, mouse and rat *JMJD1C* orthologs. POU5F1-mediated expression of JMJD1C histone demethylase is implicated in the reactivation of silenced genes in undifferentiated ES cells, pancreatic islet, and diffuse-type gastric cancer.

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 code, histone demethylase, JmjC domain, blastocyst, embryonic stem cells, hematopoietic stem cells, gastric cancer

Introduction

Embryogenesis generating whole body from fertilized egg in the cell autonomous manner is regulated by the network of transcription factors as well as by the epigenetic modifications of chromatin structure around the key genes (1-3). Cancer cells acquire malignant phenotypes during multi-stage carcinogenesis due to the accumulation of epigenetic changes and genetic alterations of cancer-associated genes (4-6). Epigenetics to investigate the epigenetic regulation during embryogenesis and carcinogenesis is the scientific frontier in the post-genome era.

Chromatin structure is composed of a nucleosome unit with 147 bp genomic DNA wrapped around the core histone octamer (7-9). Genomic DNA is modified through methylation of cytosine bases located 5' to a guanosine in a CpG dinucleotide. Histones are modified through methylation, acetylation, phosphorylation, ubiquitylation, sumoylation, and ADP-ribosylation. Hypermethylation of genomic DNA as well as methylation at Lysine (K) 9 residue of histone H3 (H3K9), H3K27 and H4K20 are associated with regions of transcriptionally silenced chromatin. On the other hand, hypomethylation of genomic DNA as well as methylation at histone H3K4, H3K36, H3K79 and acetylation at histone H3K9 are associated with regions of transcriptionally active chromatin. Epigenetic modifications of genomic DNA and histones alter the chromatin structure to regulate the accessibility of transcription factors to the promoter or enhancer regions.

In 2003, we identified and characterized JMJD1C (TRIP8) consisting of TRI8H1 domain with C2HC4-type zinc finger-like motif, TRI8H2 domain with thyroid hormone receptor ß-binding region, and JmjC domain (10). JMJD1A (TSGA),

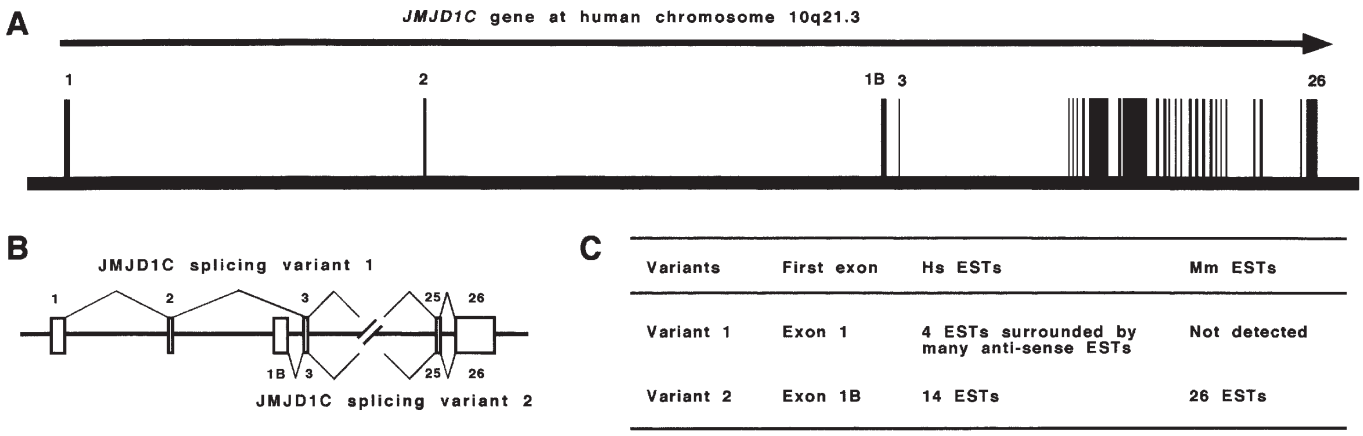

Figure 1. *JMJD1C* splicing variants. (A), Structure of *JMJD1C* gene. Exon 1B located within intron 2 was identified in this study. (B), Alternative promoter of *JMJD1C* gene. JMJD1C variant 1 previously reported consists of exons 1, 2 and 3-26, while JMJD1C variant 2 is transcribed from exon 1B. (C), Comparison between JMJD1C variants 1 and 2. JMJD1C variant 2 was the major transcript, which is evolutionarily conserved.

JMJD1B (5qNCA) and JMJD1C with the common domain architecture are JMJD1 family histone demethylases (10-12).

Here, comparative integromics on *JMJD1C* gene is reported. JMJD1C variants 1 and 2 are splicing variants generated due to alternative promoter. JMJD1C variant 2 characterized in this study was the evolutionarily conserved major transcript. *JMJD1C* mRNA was expressed in undifferentiated embryonic stem (ES) cells, pancreatic islet, diffuse-type gastric cancer, and other tissues or tumors. Transcriptional mechanism of JMJD1C variant 2 in undifferentiated ES cells and diffuse-type gastric cancer will also be described.

Materials and methods

Identification and characterization of novel exon of human JMJD1C gene. Human genome sequences homologous to human JMJD1C RefSeq (NM_032776.1) were searched for with BLAST programs as described previously (13,14). Human ESTs homologous to genome sequence around the *JMJD1C* gene were next searched for to identify novel exon derived from the *JMJD1C* gene. JMJD1C isoform incorporating the novel exon was then characterized.

In silico expression analyses. Expressed sequence tags (ESTs) derived from human *JMJD1C* and mouse *Jmjd1c* genes were searched for using the BLAST programs as described previously (15,16). The sources of human ESTs were listed up for *in silico* expression analyses.

Comparative proteomics analyses. Amino-acid sequences of JMJD1C orthologs were aligned using Genetyx program. Domain architecture of JMJD1C orthologs was analyzed using RPS-BLAST program.

Comparative genomics analyses. Human genome sequence around the *JMJD1C* gene was compared with chimpanzee, cow, mouse, and rat genome sequences to identify evolutionarily conserved regions as described previously (17,18). Binding sites for transcription factors, such as TCF/LEF, POU5F1, SOX2 and NANOG were then searched for as described previously (19,20).

Results

Alternative promoters of human JMJD1C gene. BLAST programs using NM_032776.1 RefSeq revealed that human *JMJD1C* gene is located within the human genome sequences AC022022.10, AL607128.8, AL713895.8 and AL590502.12. We assembled these genome sequences around the human *JMJD1C* gene to construct the 'JMJD1C genome contig' of 310 kb in size. BLAST programs using the JMJD1C genome contig revealed that a novel exon was located between exons 2 and 3 of the *JMJD1C* gene (Fig. 1A). Human ESTs incorporating the novel exon were next searched for to characterize the novel exon. Ten ESTs were spliced from the novel exon to exon 3; however, no EST was spliced from exon 2 to the novel exon. Based on these facts, it was concluded that the novel exon was an alternative first exon. Therefore, the novel exon was designated exon 1B (Fig. 1A).

JMJD1C transcript previously reported consists of exons 1, 2 and 3-26 (10), while JMJD1C transcript characterized in this study was transcribed from novel exon 1B located 5' to exon 3 (Fig. 1B). Because some exons between exons 3 and 26 were spliced out due to alternative splicing of the cassette splicing type, JMJD1C variants 1 and 2 were defined as JMJD1C transcripts starting from exons 1 and 1B, respectively.

BLAST programs revealed that 4 human JMJD1C ESTs were transcribed from exon 1, and that 14 human JMJD1C ESTs were transcribed from exon 1B (Fig. 1C). In addition, all of 26 mouse Jmjd1c ESTs were transcribed from exon 1b (Fig. 1C). Based on these facts, it was concluded that JMJD1C variant 2 transcribed from exon 1B was the major transcript.

Comparative proteomics on JMJD1C orthologs. Human NM_004241.2 RefSeq and mouse AK173162.1 cDNA were the representative full-length cDNAs corresponding to JMJD1C variant 2. NP_004232.2 amino-acid sequence translated from NM_004241.2 RefSeq was used as the representative human JMJD1C amino-acid sequence. Nucleotide position 312-7598 was translated into amino-acid sequence in AK173162.1 cDNA; however, nucleotide position 657-7598

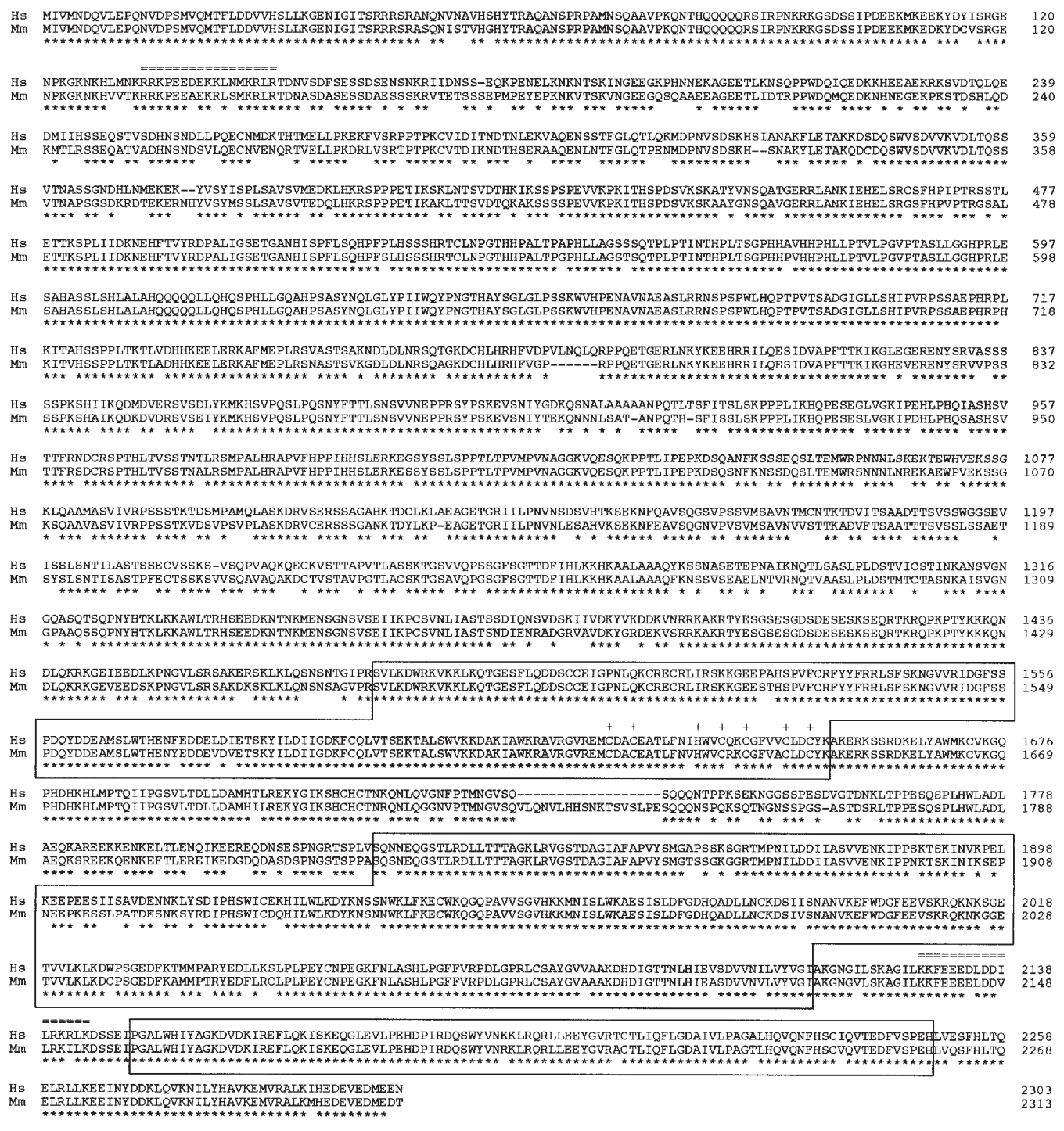

Figure 2. JMJD1C orthologs. Human JMJD1C variant 2 and mouse Jmjd1c are aligned. Amino-acid residues are numbered on the right. Conserved amino-acid residues are shown by asterisks below the alignment. TRI8H1 domain with C2HC4-type zinc finger-like motif, TRI8H2 domain with thyroid hormone receptor ß (TRß)-binding region, and JmjC domain are shown by open boxes. Two bipartite nuclear localization signals are shown by double overlines. Human JMJD1C variant 2 shows 85.7% total-amino-acid identity with mouse Jmjd1c.

was translated to determine the representative mouse Jmjd1c amino-acid sequence in this study (Fig. 2). Human JMJD1C and mouse Jmjd1c showed 85.7% total-amino-acid identity.

TRI8H1 domain with C2HC4-type zinc finger-like motif, TRI8H2 domain with thyroid hormone receptor ß (TRß)-binding region, JmjC histone demethylase domain, and two bipartite nuclear localization signals previously identified in human JMJD1C variant 1 (10) were well conserved between human JMJD1C variant 2 and mouse Jmjd1c (Fig. 2).

Expression profile of JMJD1C orthologs. Thirteen JMJD1C ESTs were derived from ES cells. Among these human JMJD1C ESTs, 10 were expressed in undifferentiated human ES cells. Eight Jmjd1c ESTs were derived from undifferentiated mouse ES cells, and 18 from mouse embryonic germ (EG) cells. In addition, one Jmjd1c EST was derived from mouse fertilized egg, and 14 from mouse blastocyst. These facts indicate that mammalian *JMJD1C* orthologs were expressed in pluripotent stem cells during early embryogenesis.

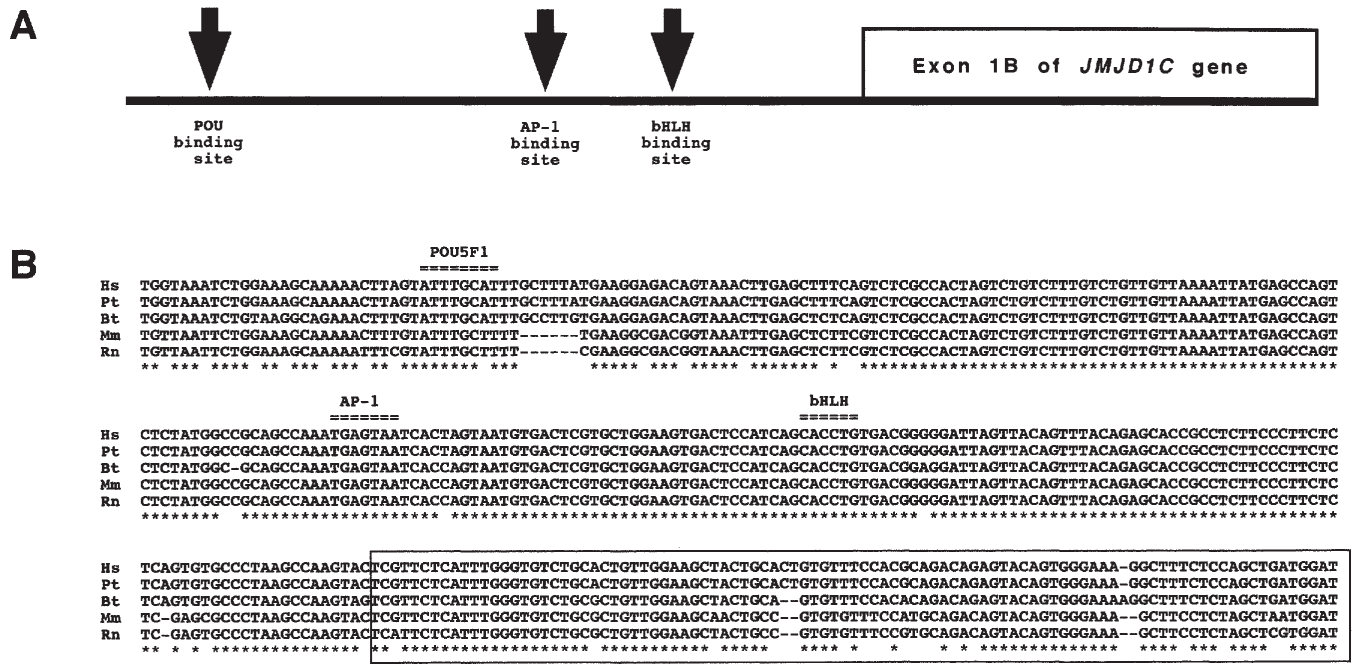

Figure 3. Promoter of the *JMJD1C* major transcript. (A), Schematic representation of promoter of *JMJD1C* major transcript transcribed from exon 1B. Binding sites for POU5F1, AP-1, and bHLH transcription factors are shown by arrows. (B), Alignment of mammalian *JMJD1C* promoter region. Hs, human; Pt, chimpanzee; Bt, cow; Mm, mouse; Rn, rat. Binding sites for POU5F1, AP-1, and bHLH transcription factors are conserved among promoter regions of mammalian *JMJD1C* orthologs.

Thity-five Jmjd1c ESTs were derived from mouse hematopoietic stem cells. Among these mouse Jmjd1c ESTs, 24 were derived from c-Kit$^+$, Sca-1$^+$ and Lin$^-$ population of mouse hematopoietic stem cells.

Mammalian *JMJD1C* orthologs were also expressed in adult tissues and tumors. Eight JMJD1C ESTs were derived from human islet, and three from human insulinoma. Six Jmjd1c ESTs were derived from mouse pancreatic islet. Four JMJD1C ESTs were derived from diffuse-type human gastric cancer.

Comparative genomics on JMJD1C promoters. Comparative genomics analyses revealed that the 5'-promoter for human JMJD1C variant 2 rather than variant 1 was well conserved in mouse (Fig. 3A). The 5'-promoter for human JMJD1C variant 2 was conserved not only in mouse, but also in chimpanzee, cow and rat (Fig. 3B).

Transcription factor-binding sites within the 5'-promoter for human JMJD1C variant 2 were next searched for based on Genetyx program, Match program, and manual annotation as previously described (19,20). POU-, AP-1-, and bHLH-binding sites were identified within the 5'-promoter for human JMJD1C variant 2. In addition, these POU-, AP-1-, and bHLH-binding sites were conserved in chimpanzee, cow, mouse, and rat *JMJD1C* orthologs (Fig. 3B).

Discussion

Comparative integromics on *JMJD1C* gene were investigated in this study. JMJD1C variant 1, previously reported, consists of exons 1, 2 and 3-26, while JMJD1C variant 2 characterized in this study was transcribed from novel exon 1B located 5' to exon 3. Four human JMJD1C ESTs were transcribed from exon 1, while 14 human JMJD1C ESTs from exon 1B. All of 26 mouse Jmjd1c ESTs were transcribed from exon 1b (Fig. 1). These facts indicate that JMJD1C variant 2 transcribed from exon 1B was the major transcript.

JMJD1C belongs to the Jumonji family characterized by the JmjC domain. In 1995, Takeuchi *et al* cloned and characterized mouse *Jumonji* (*Jmj*) gene encoding a transcriptional repressor with JmjN and JmjC domains, which is the founding member of the Jumonji family (21). We identified and characterized *JMJD1C, JMJD2A, JMJD2B, JMJD2C* and *JMJD2D* genes encoding Jumonji family proteins in 2003 and 2004 (10,22). 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 (12,23).

Human JMJD1C variant 2 with TRI8H1, TRI8H2, and JmjC domains showed 85.7% total-amino-acid identity with mouse Jmjd1c (Fig. 2). JMJD1C orthologs share a common domain architecture with JMJD1A and JMJD1B (10). Hoog *et al* cloned and characterized mouse Jmjd1a as a male germ cell-specific transcript in 1991 (24). Yamane *et al* reported JMJD1A, using their in-house name JHDM2A, as a histone H3K9 demethylase implicated in the androgen receptor-mediated gene activation in 2006 (25). Hu *et al* cloned and characterized JMJD1B at human chromosome 5q35, which is deleted in myeloid leukemia in 2001 (26). JMJD1C, originally designated TRIP8, is a thyroid hormone receptor ß-binding protein (10,11). JMJD1 family members are histone demethylases, implicated in the transcriptional regulation of target genes for nuclear hormone receptors.

Human *JMJD1C* mRNA was expressed in undifferentiated ES cells, pancreatic islet, diffuse-type gastric cancer, and other tissues or tumors. Mouse *Jmjd1c* mRNA was expressed in

fertilized egg, blastocyst, undifferentiated ES cells, EG cells, c-Kit$^+$/Sca-1$^+$/Lin$^-$ hematopoietic stem cells, and other tissues. Comparative genomics analyses revealed that binding sites for POU5F1, AP-1, and bHLH transcription factors within the promoter region locate 5' to exon 1B of human *JMJD1C* gene and were conserved in chimpanzee, cow, mouse and rat *JMJD1C* orthologs (Fig. 3). We recently reported preferential expression of *POU5F1* in undifferentiated ES cells, pancreatic islet, and diffuse-type gastric cancer (19). Together these facts indicate that POU5F1-mediated expression of JMJD1C histone demethylase is implicated in the reactivation of silenced genes in undifferentiated ES cells, pancreatic islet, and diffuse-type gastric cancer.

References

1. Boyer LA, Lee TI, Cole MF, *et al*: Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122: 947-956, 2005.
2. Lee TI, Jenner RG, Boyer LA, *et al*: Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125: 301-313, 2006.
3. Yamanaka Y, Ralston A, Stephenson RO and Rossant J: Cell and molecular regulation of the mouse blastocyst. Dev Dyn 235: 2301-2314, 2006.
4. Jones PA and Baylin SB: The functional role of epigenetic events in cancer. Nat Rev Genet 3: 415-428, 2002.
5. Baylin SB and Ohm JE: Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat Rev Cancer 6: 107-116, 2006.
6. Katoh M and Katoh M: Bioinformatics for cancer management in the post-genome era. Technol Cancer Res Treat 5: 169-176, 2006.
7. Jenuwein T and Allis CD: Translating the histone code. Science 293: 1074-1080, 2001.
8. Bird A: DNA methylation patterns and epigenetic memory. Genes Dev 16: 6-21, 2002.
9. Turner BM: Defining an epigenetic code. Nat Cell Biol 9: 1-5, 2007.
10. Katoh M and Katoh M: Identification and characterization of *TRIP8* gene *in silico*. Int J Mol Med 12: 817-821, 2003.
11. Lee JW, Choi HS, Gyuris J, Brent R and Moore DD: Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. Mol Endocrinol 9: 243-254, 1995.
12. Klose RJ, Kallin EM and Zhang Y: JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet 7: 715-727, 2006.
13. Katoh M: Paradignshift in gene-finding method: from bench-top approach to desk-top approach. Int J Mol Med 10: 677-682, 2002.
14. Katoh M: Molecular evolution of *WNT2B* orthologs. Int J Oncol 26: 1135-1139, 2005.
15. Katoh M and Katoh M: Identification and characterization of human *HESL*, rat *Hesl* and rainbow trout *hesl* genes *in silico*. Int J Mol Med 14: 747-751, 2005.
16. Katoh M and Katoh M: Comparative genomics on mammalian *FGF3-FGF4* olocus. 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.
21. 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.
22. Katoh M and Katoh M: Identification and characterization of *JMJD2* family genes *in silico*. Int J Oncol 24: 1623-1628, 2004.
23. 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.
24. 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.
25. 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.
26. 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.