SPANDIDOS PUBLICATIONS

Comparative genomics on *PROM1* gene encoding stem cell marker CD133

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 March 12, 2007; Accepted April 11, 2007

Abstract. Stem cells are characterized by self-renewal and multipotency to produce multiple lineages of progenitor and differentiated cells. *PROM1* gene encodes CD133 protein, which is a cell surface marker of hematopoietic stem cells, prostatic epithelial stem cells, pancreatic stem cells, leukemic stem cells, liver cancer stem cells, and colorectal cancer stem cells. Here, comparative integromics analyses on PROM1 orthologs were performed. Human PROM1 RefSeq NM_006017.1 was a truncated transcript, while AK027422.1 was the representative human PROM1 cDNA. Chimpanzee *PROM1* gene, consisting of 27 exons, was identified within NW_001234057.1 genome sequence. Chimpanzee 5-transmembrane protein CD133 showed 99.2% and 60.9% total-amino-acid identity with human and mouse CD133 orthologs, respectively. Only 2 of 8 Asn-linked glycosylation sites in primate CD133 orthologs were conserved in rodent CD133 orthologs. Comparative proteomics revealed that CD133 orthologs were relatively divergent between primates and rodents. *PROM1* mRNA was expressed in human embryonic stem (ES) cells, trachea, small intestine, NT2 cells, diffuse-type gastric cancer, and colorectal cancer. Human *PROM1* mRNA transcribed from exon 1A was the major transcript. Comparative genomics revealed that the region around exon 1A corresponding to 5'-UTR of human *PROM1* mRNA was not conserved in mouse and rat. Intron 2 of *PROM1* orthologs was relatively well conserved among mammals. Tandem TCF/LEF-binding sites with 7-bp spacing within intron 2 were conserved among human, chimpanzee, mouse, and rat *PROM1* orthologs. Together these facts indicate that canonical WNT signaling activation is implicated in CD133 expression in ES cells, adult stem cells, and cancer stem cells.

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: embryonic stem cells, cancer stem cells, WNT, endoderm, gastrointestinal tract, pancreas, regenerative medicine, systems medicine

Introduction

Stem cells are characterized by self-renewal and multipotency. Embryonic stem (ES) cells derived from inner cell mass of blastocyst are implicated in tissue morphogenesis, while normal stem cells derived from adult tissues are implicated in tissue homeostasis (1-4). Epigenetic changes and genetic alterations in normal stem cells or progenitor cells give rise to cancer stem cells (4-6).

WNT signals are implicated in embryogenesis, tissue repair during chronic inflammation, and carcinogenesis (5-8). WNT signals are transduced through Frizzled family receptors to the canonical and non-canonical pathways in context-dependent manner (9-11). Canonical WNT signals induce assembly of Frizzled-Dishevelled and LRP5/LRP6-AXIN-FRAT complexes to release ß-catenin from ubiquition-dependent degradation. Stabilized ß-catenin is translocated into the nucleus to generate PYGO-BCL9/BCL9L-ß-catenin-TCF/LEF complex for the transcriptional upregulation of target genes (12,13). *MYC, WISP1, CCND1, DKK1, FGF20, JAG1,* and *GCG* are target genes of the canonical WNT signaling pathway (14-20). *SFRP1* and *SFRP2* genes encoding secreted WNT antagonists are silenced due to promoter hypermethylation in colorectal cancer (21,22). *APC* gene encoding a negative regulator of the canonical WNT signaling cascade and *CTNNB1* gene encoding ß-catenin are mutated in colorectal cancer (23). Reya and Clevers proposed that dysregulation of the canonical WNT signaling pathway plays a key role in colorectal carcinogenesis (5).

CD133 protein encoded by *PROM1* gene was initially identified as a cell surface marker for human hematopoietic stem or progenitor cells (24). CD133 is expressed on other adult tissue stem cells, such as prostatic epithelial stem cells and pancreatic stem cells (25,26). CD133 is also expressed on cancer stem cells, such as leukemic stem cells and liver cancer stem cells (27,28). Recently, Ricci-Vitiani *et al* reported that the high-density CD133$^+$ cells are colon cancer stem cells responsible for tumor formation and maintenance (29). Here, comparative integromics analyses on PROM1 orthologs were performed. Chimpanzee *PROM1* gene was identified and characterized. Expression profile of human *PROM1*

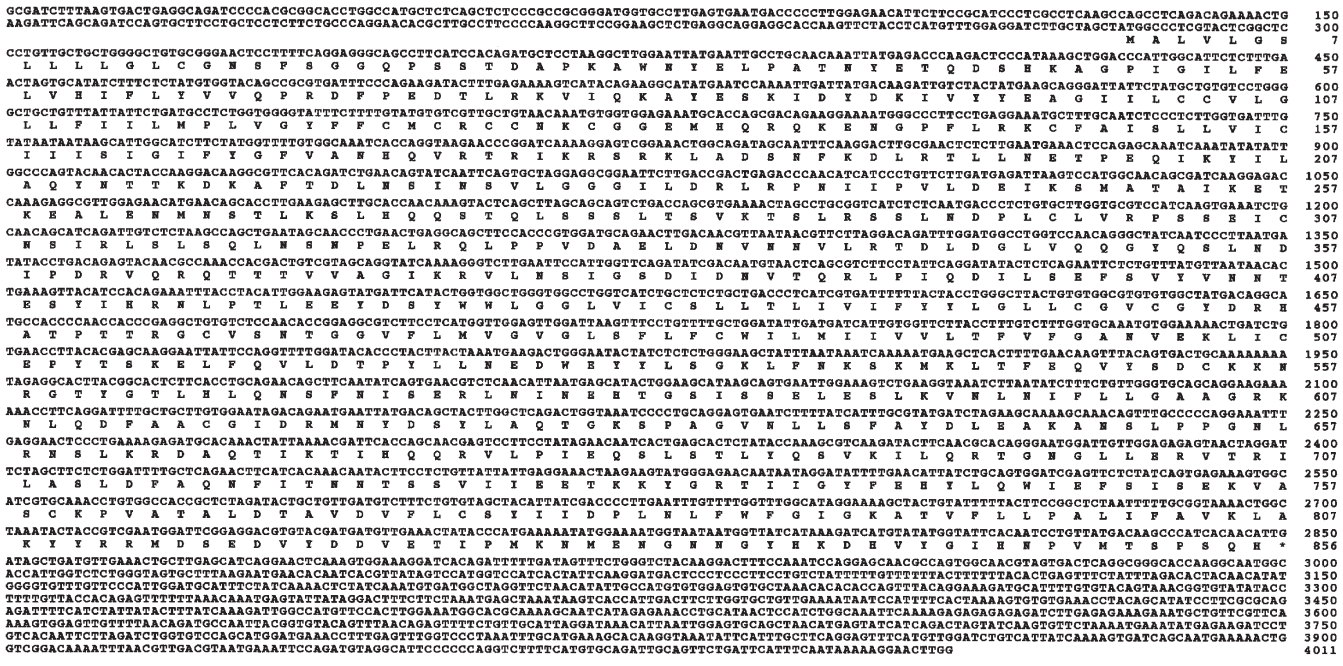

Figure 1. Nucleotide and amino-acid sequences of chimpanzee PROM1. Nucleotides and amino-acid residues are numbered on the right.

mRNA, and transcription factor-binding sites conserved among mammalian *PROM1* orthologs will also be reported.

Materials and methods

Identification and characterization of chimpanzee PROM1 ortholog. Chimpanzee genome sequence homologous to human *PROM1* was searched for with BLAST programs as described previously (30,31). 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 as described previously (32,33). Complete coding sequence (CDS) of chimpanzee PROM1 was determined by assembling exonic regions.

Comparative genomics analyses. Human genome sequence around the *PROM1* gene was compared with chimpanzee, 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 (34,35).

Comparative proteomics analyses. Domain architecture and membrane topology of CD133 orthologs were analyzed using RPS-BLAST and PSORT II programs.

In silico expression analyses. Expressed sequence tags (ESTs) derived from human *PROM1* gene were searched for using the BLAST programs as described previously (36,37). The sources of human ESTs were listed up for *in silico* expression analyses.

Results

Chimpanzee PROM1 gene. BLAST programs revealed that human PROM1 RefSeq NM_006017.1 was a truncated transcript, and AK027422.1 was the representative PROM1 cDNA. Chimpanzee *PROM1* gene located within NW_001234057.1 genome sequence was identified using the human PROM1 representative cDNA as a query sequence for the BLAST programs. Exon-intron boundaries of chimpanzee *PROM1* gene were determined based on the consensus sequence of exon-intron junctions. Chimpanzee *PROM1* gene was found consisting of 27 exons. Complete CDS of chimpanzee PROM1 was determined by assembling exonic regions. Nucleotide position 281-2851 was the coding region. Chimpanzee *PROM1* gene was found to encode a 856-amino-acid CD133 protein (Fig. 1).

Comparative proteomics on PROM1 orthologs. PSORT II analyses predicated that chimpanzee and human CD133 were 6-transmembrane proteins with N-terminal signal peptide, and that mouse and rat CD133 were 5-transmembrane proteins with N-terminal signal peptide. However, Kyte and Doolittle hydrophobicity analyses revealed that the hydrophobicity around the first transmembrane region predicted by the PSORT II analyses in primate CD133 orthologs were not significant. Because PSORT II prediction of transmembrane domains for multi-span-transmembrane proteins are not always correct, we annotated that chimpanzee and human CD133 are 5-transmembrane proteins like other mammalian CD133 orthologs (Fig. 2).

Chimpanzee CD133 showed 99.2% and 60.9% total-amino-acid identity with human and mouse CD133 orthologs, respectively. Only 2 of 8 Asn-linked glycosylation sites in primate CD133 orthologs were conserved in rodent CD133 orthologs. Based on comparative proteomics, it was concluded that CD133 orthologs were relatively divergent between primates and rodents.

In silico expression analysis on human PROM1. Expression of human *PROM1* mRNA was detected in undifferentiated

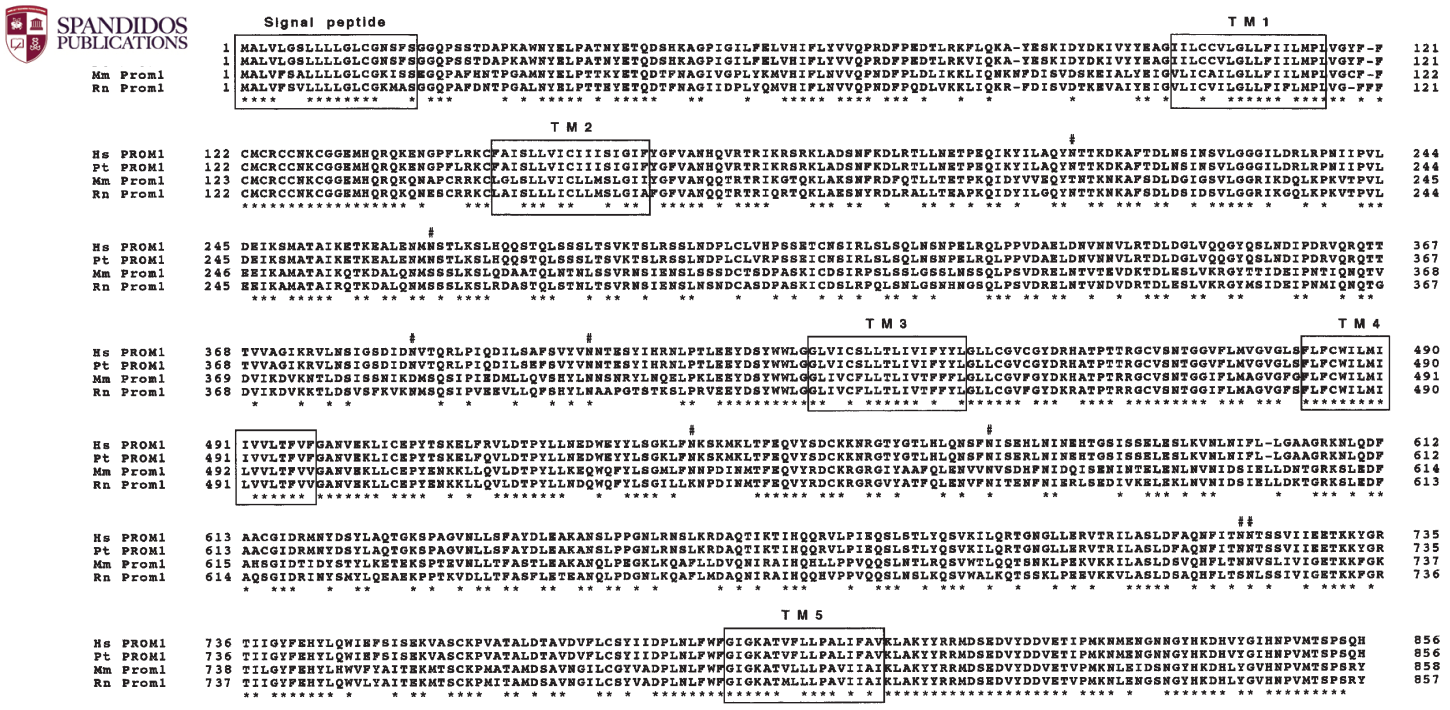

Figure 2. Mammalian PROM1 (CD133) orthologs. Hs, human; Pt, chimpanzee; Mm, mouse; Rn, rat. Signal peptide and five transmembrane domains (TM1-TM5) are boxed. Asn-linked glycosylation sites conserved in primate CD133 orthologs are indicated by a sharp.

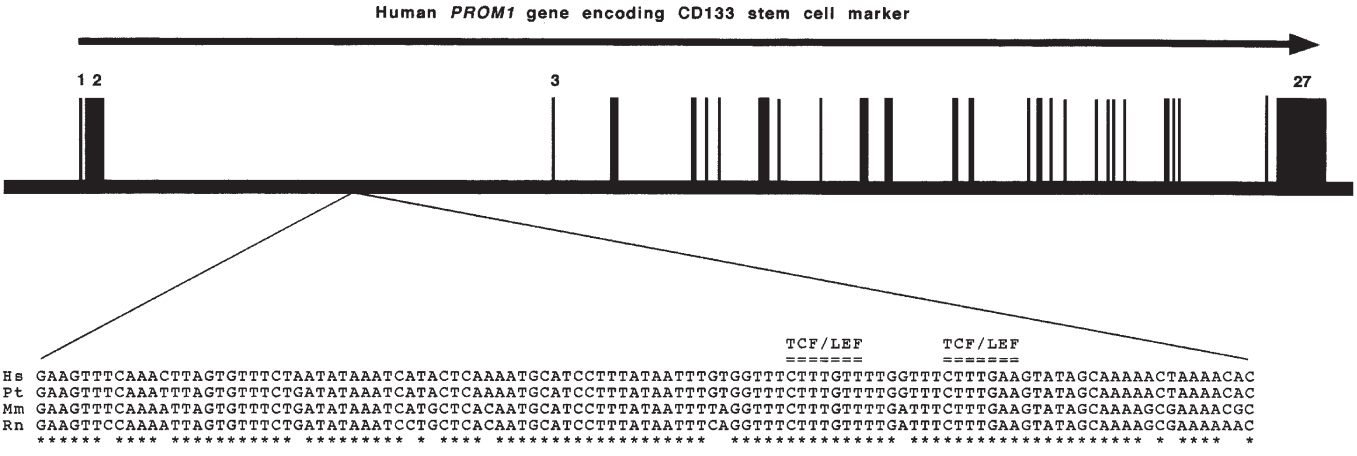

Figure 3. Structure of human *PROM1* gene and conserved tandem TCF/LEF sites. Hs, human; Pt, chimpanzee; Mm, mouse; Rn, rat. Although multiple first exons exist in the 5'-position of exon 2, only exon 1A corresponding to the major transcript is shown in this figure. Tandem TCF/LEF sites within intron 2 are conserved among mammalian *PROM1* orthologs.

embryonic stem (ES) cells, ES cells-derived embryoid body, ES cell-derived hepatocyte progenitors, trachea, small intestine, NT2 cells, diffuse-type gastric cancer, and colorectal cancer using *in silico* expression analysis.

In addition, analyses on the first exon of *PROM1* transcript revealed that human *PROM1* mRNA transcribed from exon 1A, like AK027422.1 representative transcript, was the major transcript.

Comparative genomics on PROM1 orthologs. Chimpanzee *PROM1* gene was located within NW_001234057.1 genome sequence as mentioned above. BLAST programs using human representative PROM1 cDNA AK027422.1 revealed that human *PROM1* gene was located within AC108063.3 and AC005598.6 genome sequences. BLAST programs next revealed that mouse *Prom1* gene was located within AC103621.6 and AC102476.14 genome sequences, and rat *Prom1* gene within AC114149.4 and AC134757.2 genome sequences.

Comparison between human and mouse genome sequences around the *PROM1* orthologs revealed that region around exon 1A including intron 1 of human *PROM1* gene was not conserved in mouse, and that exon 2 with initiator methionine and three regions within intron 2 of human *PROM1* gene were conserved in mouse.

Conserved transcription factor-binding sites in conserved regions within intron 2 were next searched for. Double TCF/LEF-binding sites with 7-bp spacing within intron 2 were conserved among human, chimpanzee, mouse, and rat *PROG1* orthologs (Fig. 3).

Discussion

Comparative integromics analyses on PROM1 orthologs were performed in this study. Human PROM1 RefSeq NM_006017.1 was a truncated transcript, while AK027422.1 was the representative human PROM1 cDNA. Chimpanzee *PROM1* gene, consisting of 27 exons, was identified within NW_001234057.1 genome sequence. Chimpanzee *PROM1* gene was found to encode a 856-amino-acid CD133 protein (Fig. 1). Five-transmembrane protein CD133 showed 99.2% and 60.9% total-amino-acid identity with human and mouse CD133 orthologs, respectively (Fig. 2). Only 2 of 8 Asn-linked glycosylation sites in primate CD133 orthologs were conserved in rodent CD133 orthologs. Comparative proteomics revealed that CD133 orthologs were relatively divergent between primates and rodents.

In silico expression analyses revealed that *PROM1* mRNA was expressed in undifferentiated human ES cells, trachea, small intestine, NT2 cells, diffuse-type gastric cancer, and colorectal cancer. This is the first report on CD133 expression on diffuse-type gastric cancer.

Shmelkov *et al* determined the structure of human *PROM1* gene to reveal the existence of multiple alternative first exons in the 5' position of exon 2. They also reported putative transcription factor-binding sites within the 5'-promoter region of alternative first exons (38).

We clarified that exon 1A corresponding to 5'-UTR was the first exon for the major transcript derived from human *PROM1* gene. However, the region around exon 1A of human *PROM1* gene was not conserved in mouse and rat. Instead, intron 2 of *PROM1* orthologs was relatively well conserved among mammals. Tandem TCF/LEF-binding sites with 7-bp spacing within intron 2 were conserved among human, chimpanzee, mouse, and rat *PROM1* orthologs (Fig. 3). Together these facts indicate that canonical WNT signaling activation plays a key role for the CD133 expression in ES cells, adult stem cells, and cancer stem cells.

References

1. Katoh M: Regulation of WNT signaling molecules by retinoic acid during neuronal differentiation in NT2 cells: threshold model of WNT action. Int J Mol Med 10: 683-687, 2002.
2. Clevers H: Stem cells, asymmetric division and cancer. Nat Genet 37: 1027-1028, 2005.
3. Loh YH, Wu Q, Chew JL, *et al*: The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38: 431-440, 2006.
4. Passegué E: A game of subversion. Nature 442: 754-755, 2006.
5. Reya T and Clevers H: Wnt signalling in stem cells and cancer. Nature 434: 843-850, 2005.
6. Katoh M and Katoh M: Bioinformatics for cancer management in the post-genome era. Technol Cancer Res Treat 5: 169-176, 2006.
7. Katoh M: *WNT* and *FGF* gene clusters. Int J Oncol 21: 1269-1273, 2002.
8. Katoh M and Katoh M: Cross-talk of WNT and FGF signaling pathways at GSK3ß to regulate ß-catenin and SNAIL signaling cascades. Cancer Biol Ther 5: 1059-1064, 2006.
9. Swain RK, Katoh M, Medina A and Steinbeisser H: *Xenopus* frizzled-4S, a splicing variant of Xfz4, is a context-dependent activator and inhibitor of Wnt/ß-catenin signaling. Cell Commun Signal 3: 12, 2005.
10. Katoh M: WNT/PCP signaling pathway and human cancer. Oncol Rep 14: 1583-1588, 2005.
11. Katoh M and Katoh M: STAT3-induced WNT5A signaling loop in embryonic stem cells, adult normal tissues, chronic persistent inflammation, rheumatoid arthritis, and cancer. Int J Mol Med 19: 273-278, 2007.
12. Kramps T, Peter O, Brunner E, *et al*: Wnt/wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear ß-catenin-TCF complex. Cell 109: 47-60, 2002.
13. Katoh M and Katoh M: Identification and characterization of human *BCL9L* gene and mouse *Bcl9l* gene *in silico*. Int J Mol Med 12: 643-649, 2003.
14. He TC, Sparks AB, Rago C, *et al*: Identification of *c-MYC* as a target of the APC pathway. Science 281: 1509-1512, 1998.
15. Pennica D, Swanson TA, Welsh JW, *et al*: *WISP* genes are members of the connective tissue growth factor family that are up-regulated in *Wnt1*-transformed cells and aberrantly expressed in human colon tumors. Proc Natl Acad Sci USA 95: 14717-14722, 1998.
16. Tetsu O and McCormick F: ß-catenin regulates expression of *cyclin D1* in colon carcinoma cells. Nature 398: 422-426, 1999.
17. Chamorro MN, Schwartz DR, Vonica A, *et al*: *FGF20* and *DKK1* are transcriptional target of ß-catenin and FGF20 is implicated in cancer and development. EMBO J 24: 73-84, 2005.
18. Katoh M and Katoh M: Comparative genomics on *FGF20* orthologs. Oncol Rep 14: 287-290, 2005.
19. Katoh M and Katoh M: Notch ligand, JAG1, is evolutionarily conserved target of canonical WNT signaling pathway in progenitor cells. Int J Mol Med 17: 681-685, 2006.
20. Yi F, Brubaker PL and Jin T: TCF-4 mediates cell type-specific regulation of proglucagon gene expression by ß-catenin and glycogen synthase kinase-3ß. J Biol Chem 280: 1457-1464, 2005.
21. Suzuki H, Watkins DN, Jair KW, *et al*: Epigenetic inactivation of *SFRP* genes allows constitutive WNT signaling in colorectal cancer. Nat Genet 36: 417-422, 2004.
22. Katoh Y and Katoh M: WNT antagonist, SFRP1, is Hedgehog signaling target. Int J Mol Med 17: 171-175, 2006.
23. Nishisho I, Nakamura Y, Miyoshi Y, *et al*: Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science 253: 665-669, 1991.
24. Yin AH, Miraglia S, Zanjani ED, *et al*: AC133, a novel marker for human hematopoietic stem and progenitor cells. Blood 90: 5002-5012, 1997.
25. Richardson GD, Robson CN, Lang SH, *et al*: CD133, a novel marker for human prostatic epithelial stem cells. J Cell Sci 117: 3539-3545, 2004.
26. Oshima Y, Suzuki A, Kawashimo K, *et al*: Isolation of mouse pancreatic ductal progenitor cells expressing CD133 and c-Met by flow cytometric cell sorting. Gastroenterology (In press).
27. Horn PA, Tesch H, Staib P, *et al*: Expression of AC133, a novel hematopoietic precursor antigen, on acute myeloid leukemia cells. Blood 93: 1435-1437, 1999.
28. Suetsugu A, Nagaki M, Aoki H, *et al*: Characterization of $CD133^+$ hepatocellular carcinoma cells as cancer stem/progenitor cells. Biochem Biophys Res Commun 351: 820-824, 2006.
29. Ricci-Vitiani L, Lombardi DG, Pilozzi E, *et al*: Identification and expansion of human colon-cancer-initiating cells. Nature 445: 111-115, 2007.
30. Katoh M: Paradigm shift in gene-finding method: from bench-top approach to desk-top approach. Int J Mol Med 10: 677-682, 2002.
31. 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.
32. Katoh M: Molecular evolution of *WNT2B* orthologs. Int J Oncol 26: 1135-1139, 2005.
33. Katoh Y and Katoh M: Comparative genomics on HHIP family orthologs. Int J Mol Med 17: 391-395, 2006.
34. Katoh Y and Katoh M: FGF signaling inhibitor, SPRY4, is evolutionarily conserved target of WNT signaling pathway in progenitor cells. Int J Mol Med 17: 529-532, 2006.
35. 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.
36. Katoh Y and Katoh M: Comparative integromics on BMP/GDF family. Int J Mol Med 17: 951-955, 2006.
37. 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.
38. Shmelkov SV, Jun L, St. Clair R, *et al*: Alternative promoters regulate transcription of the gene that encodes stem cell surface protein AC133. Blood 103: 2055-2061, 2004.