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.

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 β-catenin from ubiquition-dependent degradation. Stabilized β-catenin is translocated into the nucleus to generate PYGO-BCL9/BCL9L-β-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 β-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*
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 to be composed 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 predicted 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
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 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