Comparative integromics on Eph family

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 January 18, 2006; Accepted February 15, 2006

Abstract. EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA7, EPHA8, EPHA10, EPHB1, EPHB2, EPHB3, EPHB4 and EPHB6 are EPH family receptors for Ephrin family ligands. Ephrin/EPH signaling pathway networks with the WNT signaling pathway during embryogenesis, tissue regeneration, and carcinogenesis. TCF/LEF-binding sites within the promoter region of human *EPH* family members were searched for by using bioinformatics and human intelligence. Because five TCF/LEF-binding sites were identified within the 5'-promoter region of the *EPHA7* gene, comparative genomics analyses on *EPHA7* orthologs were further performed. *EPHA7-MANEA-FHL5* locus at human chromosome 6q16.1 and *EPHA10-MANEAL-FHL3* locus at human chromosome 1p34.3 were paralogous regions within the human genome. Human *EPHA7* mRNA was expressed in embryonic stem (ES) cells, neural tissues, duodenal cancer and parathyroid tumors, while mouse *Epha7* mRNA was expressed in fertilized egg, Rathke's pouche, visual cortex, pituitary gland, other neural tissues, pancreas, lung tumors and mammary tumors. The chimpanzee *EPHA7* gene and cow *Epha7* gene were identified within NW_107969.1 and AC155055.2 genome sequences, respectively. Five TCF/LEF-binding sites within human *EPHA7* promoter were conserved in the chimpanzee *EPHA7* promoter, and three TCF/LEF-binding sites in the cow *Epha7* promoter, but none in the mouse *Epha7* promoter. Primates and cow *EPHA7* orthologs were identified as evolutionarily conserved targets of the WNT/ß-catenin signaling pathway. D6S1056 microsatellite marker within *EPHA7* gene is deleted in prostate cancer. Deletion and/or promoter CpG hypermethylation could explain the *EPHA7* down-regulation in human tumors. *EPHA7* is a target of systems medicine, especially in the fields of regenerative medicine and oncology.

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@ncc.go.jp

Key words: bioinformatics, comparative genomics, comparative proteomics, EFN, EPH, WNT, integrome network, systems medicine

Introduction

EPH family members are receptors for Ephrin family ligands (1-4). EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA8 and EPHA10 are classified as EPHA subfamily members, while EPHB1, EPHB2, EPHB3, EPHB4 and EPHB6 are classified as the EPHB subfamily members. EPH family members consist of extracellular Ephrin-binding domain, cysteine-rich domain, two fibronectin type III repeats as well as cytoplasmic tyrosine kinase domain and C-terminal SAM motif.

EFNA1, EFNA2, EFNA3, EFNA4, EFNA5, EFNB1, EFNB2 and EFNB3 are EFRIN (EFN) family ligands for EPH family receptors (1-4). EFNA1, EFNA2, EFNA3, EFNA4 and EFNA5 are EFNA subfamily members characterized as GPI-anchored cell-surface proteins with EPH-binding domain, while EFNB1, EFNB2 and EFNB3 are EFNB subfamily members characterized as transmembrane proteins with an extracellular EPH-binding domain and a cytoplasmic PDZ-binding motif.

Ephrin/EPH signaling pathway networks with WNT signaling pathway in a variety of processes, such as axon guidance and gastrointestinal morphogenesis, during embryogenesis, tissue regeneration and carcinogenesis (5). Canonical WNT signaling pathway activates the transcription of target genes, such as *DKK1, DKK4, FGF18* and *FGF20*, depending on the transcriptional complex consisting of TCF/LEF, ß-catenin, BCL9/BCL9L and PYGO1/PYGO2 (6-22).

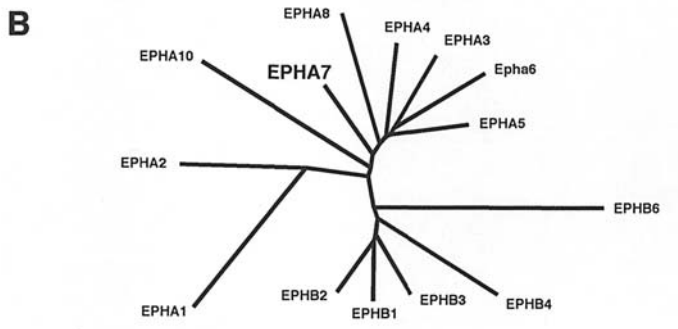
TCF/LEF-binding sites within the promoter region of human *EPH* family members were searched for by using bioinformatics and human intelligence. Because five TCF/LEF-binding sites were identified within the 5'-promoter region of the *EPHA7* gene, comparative genomics analyses on *EPHA7* orthologs were further performed.

Materials and methods

WNT target gene screening. Genome sequences corresponding to human *EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA8, EPHA10, EPHB1, EPHB2, EPHB3, EPHB4* and *EPHB6* genes were searched for with BLAST programs (http://www.ncbi.nlm.nih.gov) as described previously (23-28). TCF/LEF-binding sites within the 5'-flanking promoter region of the above genes were searched for based on bioinformatics and manual inspection as described previously (29,30).

A

Human gene	Human complete CDS	Human genome clone including promoter	TCF/LEF binding site within human promoter
EPHA1	NM_005232.3	AC092214.3	0
EPHA2	NM_004431.2	AL451042.10	1
EPHA3	NM_005233.3	AC138973.2	2
EPHA4	NM_004438.3	AC079834.8	0
EPHA5	NM_004439.4	AC104137.5	2
EPHA6	AK092565.1	AC109782.7	0
EPHA7	NM_004440.2	AL121966.10	5
EPHA8	NM_020526.3	AL035703.21	0
EPHA10	AJ872185.1	AC104336.2	2
EPHB1	NM_004441.3	AC016931.21	0
EPHB2	NM_017449.2	AL158086.32	0
EPHB3	NM_004443.3	AC112643.12	1
EPHB4	NM_004444.4	AF312032.1	0
EPHB6	NM_004445.2	AC104597.3	2



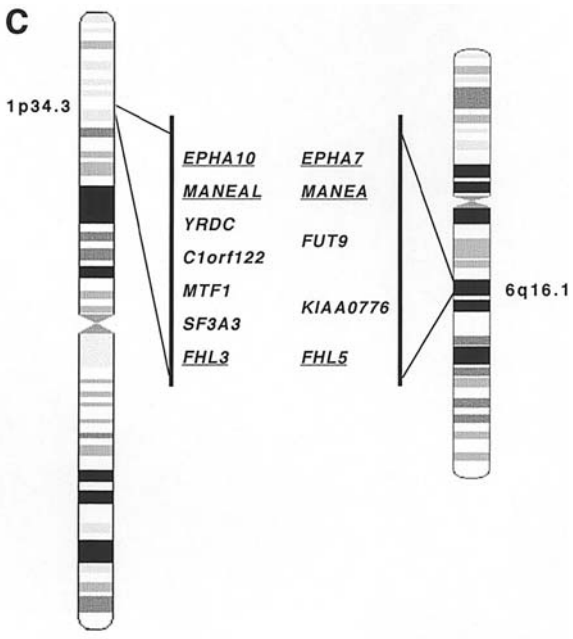


Figure 1. (A), The human *EPH* gene family. Gene symbol, complete coding sequence, genome sequence and the number of TCF/LEF-binding sites within the promoter region of 14 *EPH* family genes are listed. Five TCF/LEF-binding sites exist within human *EPHA7* promoter. (B), Phylogenetic analysis on the EPH family. EPHA7 is more related to EPHA8 and EPHA10 than the other members of the EPHA family. (C), Intra-species comparative genomics on *EPHA7* and *EPHA10* loci. *EPHA7-MANEA-FHL5* locus at human chromosome 6q16.1 and *EPHA10-MANEAL-FHL3* locus at human chromosome 1p34.3 are paralogous regions within the human genome.

Identification of the chimpanzee and cow EPHA7 orthologs. Chimpanzee and cow genome sequences homologous to human *EPHA7* were searched for with BLAST programs as described previously (31-36). TCF/LEF-binding sites within the 5'-flanking promoter region of *EPHA7* orthologs were also searched for.

Comparative proteomics analysis. Phylogenetic analysis on EPH family proteins was performed by using the CLUSTALW program.

Comparative genomics analyses. Phylogenetic analysis on the promoter of *EPHA7* orthologs was performed by using the CLUSTALW program. Promoter region of human, chimpanzee and cow *EPHA7* orthologs were aligned by using the Genetyx program and manual curation.

In silico expression analyses. Expressed sequence tags (ESTs) derived from human *EPHA7* gene and mouse *Epha7* gene were searched for by using the BLAST programs. The sources of human EPHA7 ESTs and those of mouse Epha7 ESTs were listed up for *in silico* expression analyses.

Results

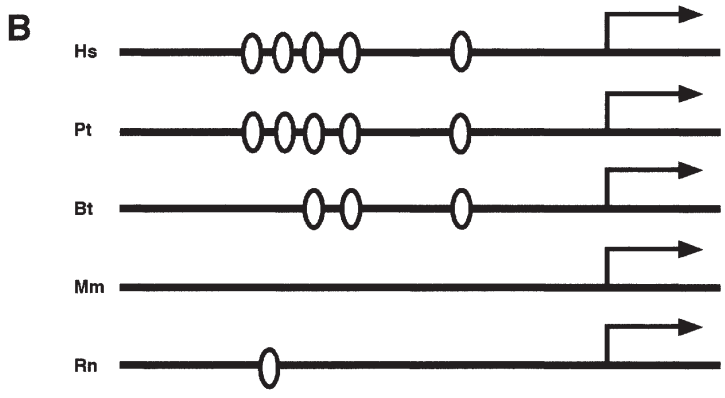
Screening of TCF/LEF-binding site within promoter region of EPH family genes. Human EPHA1 RefSeq (NM_005232.3), EPHA2 RefSeq (NM_004431.2), EPHA3 RefSeq (NM_005233.3), EPHA4 RefSeq (NM_004438.3), EPHA5 RefSeq (NM_004439.4), EPHA6 RefSeq (AK092565.1), EPHA7 RefSeq (NM_004440.2), EPHA8 RefSeq (NM_020526.3), EPHA10 coding sequence (AJ872185.1), EPHB1 RefSeq (NM_004441.3), EPHB2 RefSeq (NM_017449.2), EPHB3 RefSeq (NM_004443.3), EPHB4 RefSeq (NM_004444.4) and EPHB6 RefSeq (NM_004445.2) were used as query sequences for the BLAST programs to identify genome clones corresponding to *EPH* family genes. The 5'-flanking promoter region of human *EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA8, EPHA10, EPHB1, EPHB2, EPHB3, EPHB4* and *EPHB6* genes were identified within AC092214.3, AL451042.10, AC138973.2, AC079834.8, AC104137.5, AC109782.7, AL121966.10, AL035703.21, AC104336.2, AC016931.21, AL158086.32, AC112643.12, AF312032.1 and AC104597.3 genome sequences, respectively (Fig. 1A). TCF/LEF-binding sites within the 5'-promoter region of human *EPH* family genes were then searched for based on manual inspection. Five TCF/LEF-binding sites were identified within the human *EPHA7* promoter (Fig. 1A).

Comparative integromics analysis on EPHA7. Comparative proteomics analysis on EPHA7 was at first performed. Because human EPHA6 RefSeq encoded a C-terminally truncated isoform, mouse Epha6 was used for the phylogenetic analysis in this study. Phylogenetic analysis revealed that EPHA7 was more related to EPHA8 and EPHA10 than the other EPHA subfamily members (Fig. 1B).

Intra-species comparative genomics analysis was next performed. *MDN1, CASP8AP2, CX62, BACH2* and *MAP3K7* genes were located centromeric to *EPHA7* gene, while *MANEA, FUT9, KIAA0776* and *FHL5* were located telomeric to *EPHA7* gene. Paralogs corresponding to these genes around the *EPHA7* locus were searched for by using the BLAST programs. *MANEA-like* (*MANEAL*) and *FHL5* genes, located around the *EPHA10* locus at human chromosome 1p34.3, were paralogs of *MANEA* and *FHL3* genes, respectively. These facts indicate that *EPHA7-MANEA-FHL5* locus at human chromosome

A

	EPHA7 ortholog	RefSeq	Genome clone including promoter	TCF/LEF binding site within promoter
Hs	*EPHA7*	NM_004440.2	AL121966.10	5
Pt	*EPHA7*	Not determined	NW_107969.1	5
Bt	*Epha7*	Not determined	AC155055.2	3
Mm	*Epha7*	NM_010141.2	BX000989.10	0
Rn	*Epha7*	NM_134331.1	AC106555.4	1



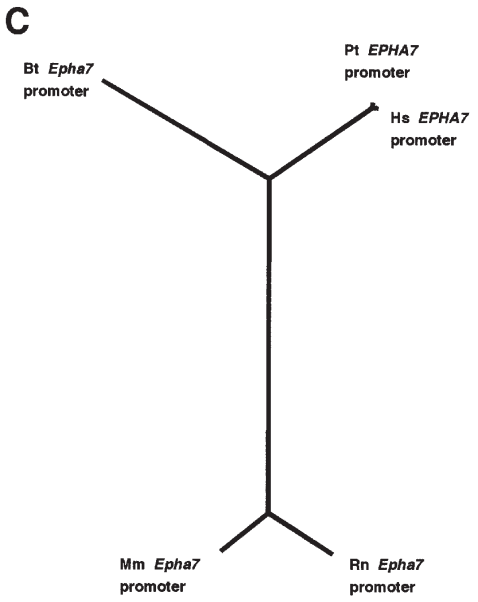


Figure 2. Inter-species comparative genomics analyses on *EPHA7* promoters. Hs, human; Pt, chimpanzee; Bt, cow; Mm, mouse; Rn, rat. (A), Promoter of mammalian *EPHA7* orthologs. (B), Schematic representation of mammalian *EPHA7* promoters. TCF/LEF-binding sites are shown by oval symbols. (C), Phylogenetic analysis on mammalian *EPHA7* promoters. Human, chimpanzee and cow *EPHA7* promoters are significantly divergent from the mouse and rat *Epha7* promoters.

6q16.1 and *EPHA10-MANEAL-FHL3* locus at human chromosome 1p34.3 were paralogous regions within the human genome (Fig. 1C).

Expression profile of human EPHA7 and mouse Epha7 mRNAs. *In silico* expression analyses were performed to compare the expression profile of human *EPHA7* and mouse *Epha7* mRNAs. Human *EPHA7* mRNA was expressed in embryonic stem (ES) cells, neural tissues, duodenal cancer and parathyroid tumors, while mouse *Epha7* mRNA was expressed in fertilized egg, Rathke's pouche, visual cortex, pituitary gland, other neural tissues, pancreas, lung tumors and mammary tumors.

Identification of the chimpamzee and cow EPHA7 orthologs. BLAST programs using human EPHA7 RefSeq revealed that chimpanzee *EPHA7* gene was located within NW_107969.1 genome sequence. Exon-intron boundaries of chimpanzee *EPHA7* gene were determined based on the consensus sequence of exon-intron junctions. Although 3'-part of exon 8 was located within the sequence gap, chimpanzee *EPHA7* gene was found consisting of 17 exons.

BLAST programs using human EPHA7 RefSeq revealed that exons 1-3 of cow *Epha7* gene were located within the AC155055.2 genome sequence.

Comparative genomics analyses on EPHA7 promoters. Human *EPHA7* promoter, chimpanzee *EPHA7* promoter and cow *Epha7* promoter were located within AL121966.10, NW_107969.1 and AC155055.2 genome sequences, respectively, as mentioned above. BLAST programs revealed that mouse and rat *Epha7* promoters were located within BX000989.10 and AC106555.4 genome sequences, respectively (Fig. 2A). GC content of human, chimpanzee and cow *EPHA7* promoters were 46.3%, that of mouse *Epha7* promoter was 50.4%, and that of rat *Epha7* promoter was 48.0%.

Five TCF/LEF-binding sites within human *EPHA7* promoter were located about 1200 bp, 1150 bp, 1000 bp, 900 bp, and 550 bp upstream of the transcription start site (Fig. 2B). Five TCF/LEF-binding sites within human *EPHA7* promoter were completely conserved in the chimpanzee *EPHA7* promoter. Although the first two TCF/LEF-binding sites within human *EPHA7* promoter have undergone nucleotide substitutions, the third to fifth TCF/LEF-binding sites within human *EPHA7* promoter were completely conserved in the cow *Epha7* promoter. On the other hand, the TCF/LEF-binding site was not identified within the mouse *Epha7* promoter (Fig. 2B).

Phylogenetic analysis revealed that human, chimpanzee and cow *EPHA7* promoters were completely divergent from mouse and rat *Epha7* promoters (Fig. 2C). Alignment of human, chimpanzee and cow *EPHA7* promoters revealed that the 5'-promoter region of human, chimpanzee and cow *EPHA7* orthologs with multiple TCF/LEF sites were well conserved (Fig. 3). Based on these facts, it was concluded that primates and cow *EPHA7* orthologs were evolutionarily conserved targets of the WNT/ß-catenin signaling pathway.

Discussion

TCF/LEF-binding sites within the promoter region of human *EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA8, EPHA10, EPHB1, EPHB2, EPHB3, EPHB4* and *EPHB6* genes were searched for in this study. Because five TCF/LEF-binding sites were identified within the 5'-promoter region of *EPHA7* gene at human chromosome 6q16.1 (Fig. 1A), comparative genomics analyses on *EPHA7* orthologs were

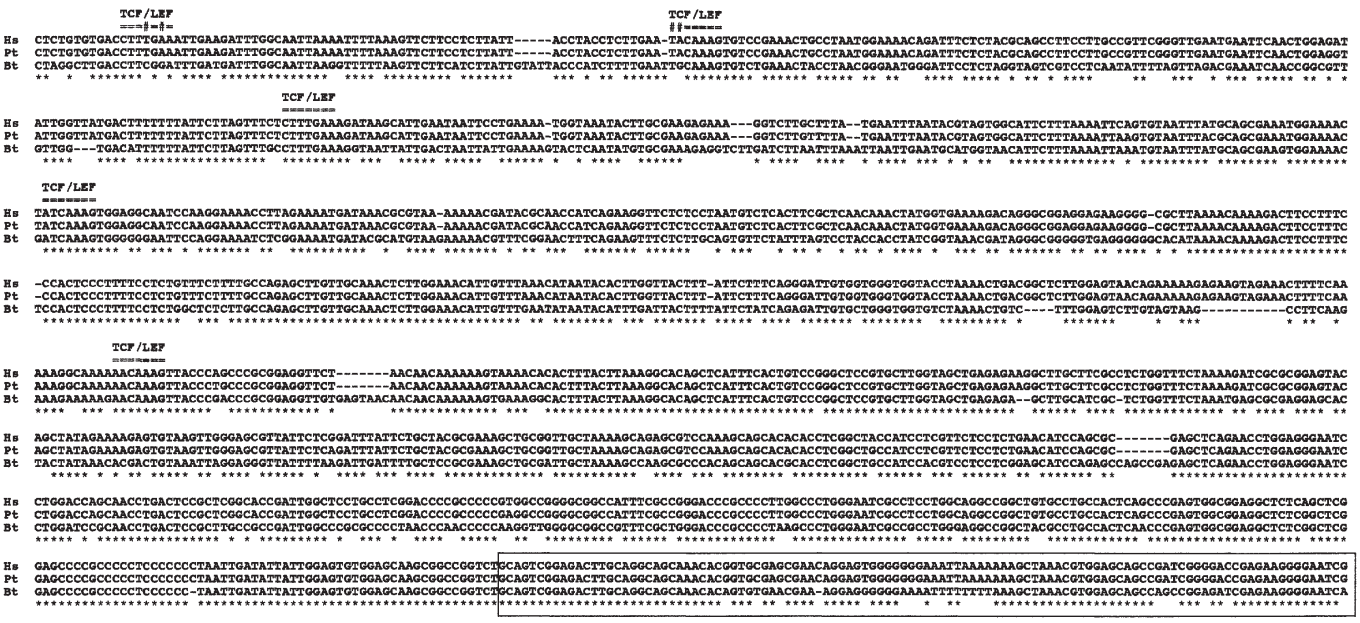


Figure 3. Alignment of human, chimpanzee and cow *EPHA7* promoters. Hs, human; Pt, chimpanzee; Bt, cow. Region corresponding to exon 1 of human *EPHA7* gene is boxed. Five TCF/LEF-binding sites, conserved in primate *EPHA7* promoters, are shown by double over-lines. Nucleotide changes to disrupt the conserved TCF/LEF-binding sites in cow *Epha7* promoter are shown by a sharp. Primates and cow *EPHA7* promoters with multiple TCF/LEF-binding sites are well conserved.

further performed. *EPHA7-MANEA-FHL5* locus at human chromosome 6q16.1 and *EPHA10-MANEAL-FHL3* locus at human chromosome 1p34.3 were paralogous regions within the human genome (Fig. 1C).

The chimpanzee *EPHA7* and cow *Epha7* genes were identified within NW_107969.1 and AC155055.2 genome sequences, respectively (Fig. 2A). Five TCF/LEF-binding sites within human *EPHA7* promoter were conserved in chimpanzee *EPHA7* promoter, and three TCF/LEF-binding sites in the cow *Epha7* promoter, but none in the mouse *Epha7* promoter (Fig. 2B). Human *EPHA7* mRNA and rodent *Epha7* mRNA were expressed in early embryonic cells, neural tissues and several tumor types; however, primate *EPHA7* promoters were significantly divergent from the rodent *Epha7* promoters (Fig. 2C). More detailed comparison on the expression profile of human *EPHA7* mRNA and rodent *Epha7* mRNA should be performed in the future.

Primates and cow *EPHA7* orthologs were identified as evolutionarily conserved targets of the WNT/ß-catenin signaling pathway. Although WNT/ß-catenin signaling pathway is activated in most cases of human colorectal cancer (37), expression of *EPHA7* mRNA in human colorectal cancer was not detected by *in silico* expression analyses. Ephrin/EPH signaling, implicated in the maintenance of colorectal mucosal homeostasis, is down-regulated in mouse advanced colorectal cancer (5,38). Microsatellite markers D6S1942, D6S1293 and D6S1056 were identified within the human *EPHA7* gene in this study. Among these microsatellite markers, D6S1056 is deleted in prostate cancer (39). *EPHA7* might be down-regulated in colorectal cancer due to a deletion. Alternatively, *EPHA7* mRNA might be down-regulated in extra-neural tissues due to promoter CpG hypermethylation.

Because *EPHA7* is an evolutionarily conserved target of the WNT/ß-catenin signaling pathway at least in primates and cow, *EPHA7* is a target of systems medicine, especially in the fields of regenerative medicine and oncology.

References

1. Flanagan JG and Vanderhaeghen P: The ephrins and Eph receptors in neural development. Annu Rev Neurosci 21: 309-345, 1998.
2. Holder N and Klein R: Eph receptors and ephrins: effectors of morphogenesis. Development 126: 2033-2044, 1999.
3. Wilkinson DG: Multiple roles of Eph receptors and ephrinin neural development. Nat Rev Neurosci 2: 155-164, 2001.
4. Mancia F and Shapiro L: ADAM and Eph: how Ephrin-signaling cells become detached. Cell 123: 185-187, 2005.
5. Battle E, Henderson JT, Beghtel H, *et al*: ß-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB. Cell 111: 251-263, 2002.
6. Katoh M: *WNT* and *FGF* gene clusters. Int J Oncol 21: 1269-1273, 2002.
7. 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.
8. Heller RS, Klein T, Ling Z, Heimberg H, Katoh M, Madsen OD and Serup P: Expression of *WNT, Frizzled, sFRP*, and *DKK* genes in adult human pancreas. Gene Expr 11: 141-147, 2003.
9. Katoh M: WNT2 and human gastrointestinal cancer. Int J Mol Med 12: 811-816, 2003.
10. 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.
11. Garciadiego-Cazares D, Rosales C, Katoh M and Chimal-Monroy J: Coordination of chondrocyte differentiation and joint formation by α5ß1 integrin in the developing appendicular skeleton. Development 131: 4735-4742, 2004.
12. Katoh M and Katoh M: Comparative genomics on *WNT8A* and *WNT8B* genes. Int J Oncol 26: 1129-1133, 2005.
13. Katoh M: Molecular evolution of *WNT2B* orthologs. Int J Oncol 26: 1135-1139, 2005.
14. Katoh M: Comparative genomics on *WNT3-WNT9B* gene cluster. Int J Mol Med 15: 743-747, 2005.

15. 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.
16. Katoh M: Epithelial-mesenchymal transition in gastric cancer. Int J Oncol 27: 1677-1683, 2005.
17. Katoh M: WNT2B: comparative integromics and clinical application. Int J Mol Med 16: 1103-1108, 2005.
18. Katoh Y and Katoh M: Comparative genomics on *DKK1* orthologs. Int J Oncol 27: 275-279, 2005.
19. Katoh Y and Katoh M: Comparative genomics on *DKK2* and *DKK4* orthologs. Int J Mol Med 16: 477-481, 2005.
20. Katoh M and Katoh M: Comparative genomics on *FGF20* orthologs. Oncol Rep 14: 287-290, 2005.
21. Katoh M and Katoh M: Comparative genomics on *FGF8, FGF17*, and *FGF18* orthologs. Int J Mol Med 16: 493-496, 2005.
22. Katoh Y and Katoh M: Comparative genomics on *FGF16* orthologs. Int J Mol Med 16: 959-963, 2005.
23. Katoh M: Paradigm shift in gene-finding method: from bench-top approach to desk-top approach. Int J Mol Med 10: 677-682, 2002.
24. Katoh M and Katoh M: Evolutionary conservation of *CCND1-ORAOV1-FGF19-FGF4* locus from zebrafish to human. Int J Mol Med 12: 45-50, 2003.
25. Katoh M and Katoh M: *CLDN23* gene, frequently down-regulated in intestinal-type gastric cancer, is a novel member of *CLAUDIN* gene family. Int J Mol Med 11: 683-689, 2003.
26. Katoh M and Katoh M: Identification and characterization of human *MPP7* gene and mouse *Mpp7* gene *in silico*. Int J Mol Med 13: 333-338, 2004.
27. Katoh M and Katoh M: Identification and characterization of *Crumbs* homolog 2 gene at human chromosome 9q33.3. Int J Oncol 24: 743-749, 2004.
28. Katoh M and Katoh M: Identification and characterization of human *HES2, HES3*, and *HES5* genes *in silico*. Int J Oncol 25: 529-534, 2004.
29. Katoh Y and Katoh M: Comparative genomics on HHIP family orthologs. Int J Mol Med 17: 391-395, 2006.
30. 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.
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, 2004.
32. Katoh M and Katoh M: Comparative genomics on *WNT5A* and *WNT5B* genes. Int J Mol Med 15: 749-753, 2005.
33. Katoh Y and Katoh M: Comparative genomics on *WNT11* gene. Int J Mol Med 15: 879-883, 2005.
34. Katoh Y and Katoh M: Comparative genomics on *VANGL1* and *VANGL2* genes. Int J Oncol 26: 1435-1440, 2005.
35. Katoh Y and Katoh M: Comparative genomics on *SFRP1* orthologs. Int J Oncol 27: 861-865, 2005.
36. Katoh Y and Katoh M: WNT antagonist, SFRP1, is Hedgehog signaling target. Int J Mol Med 17: 171-175, 2006.
37. Morin PJ, Sparks AB, Korinek V, *et al*: Activation of ß-catenin-Tcf signaling in colon cancer by mutations in *ß-catenin* or *APC*. Science 275: 1787-1890, 1997.
38. Battle E, Bacani J, Begthel H, *et al*: EphB receptor activity suppresses colorectal cancer progression. Nature 435: 1126-1130, 2005.
39. Srikantan V, Sesterhenn IA, Davis L, *et al*: Allelic loss on chromosome 6q in primary prostate cancer. Int J Cancer 84: 331-335, 1999.