SPANDIDOS PUBLICATIONS

Canonical WNT signaling pathway and human AREG

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 February 23, 2006; Accepted March 27, 2006

Abstract. AREG (Amphiregulin), BTC (ß-cellulin), EGF, EPGN (Epigen), EREG (Epiregulin), HBEGF, NRG1, NRG2, NRG3, NRG4 and TGFA (TGFα) constitute EGF family ligands for ERBB family receptors. Cetuximab (Erbitux), Pertuzumab (Omnitarg) and Trastuzumab (Herceptin) are anti-cancer drugs targeted to EGF family ligands, while Gefitinib (Iressa), Erlotinib (Tarceva) and Lapatinib (GW572016) are anti-cancer drugs targeted to ERBB family receptors. AREG and TGFA are biomarkers for Gefitinib non-responders. The TCF/LEF binding sites within the promoter region of human *EGF* family members were searched for by using bioinformatics and human intelligence (Humint). Because three TCF/LEF-binding sites were identified within the 5'-promoter region of human *AREG* gene, comparative genomics analyses on *AREG* orthologs were further performed. The *EPGN-EREG-AREG-BTC* cluster at human chromosome 4q13.3 was linked to the *PPBP-CXCL* segmental duplicons. *AREG* was the paralog of *HBEGF* at human chromosome 5q31.2. Chimpanzee *AREG* gene, consisting of six exons, was located within NW_105918.1 genome sequence. Chimpanzee AREG was a type I transmembrane protein showing 98.0% and 71.4% total amino-acid identity with human AREG and mouse Areg, respectively. Three TCF/LEF-binding sites within human *AREG* promoter were conserved in chimpanzee *AREG* promoter, but not in rodent *Areg* promoters. Primate *AREG* promoters were significantly divergent from rodent *Areg* promoters. *AREG* mRNA was expressed in a variety of human tumors, such as colorectal cancer, liver cancer, gastric cancer, breast cancer, prostate cancer, esophageal cancer and myeloma. Because human *AREG* was characterized as potent target gene of WNT/ß-catenin signaling pathway, WNT signaling activation could lead to Gefitinib resistance through *AREG* upregulation. AREG is a target of systems medicine in the field of 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, AREG, WNT, integrome network, systems medicine

Introduction

AREG (Amphiregulin), BTC (ß-cellulin), EGF, EPGN (Epigen), EREG (Epiregulin), HBEGF, NRG1, NRG2, NRG3, NRG4 and TGFA (TGFα) constitute EGF family ligands for ERBB family receptors (1,2). Ligand binding to ERBB receptors induces receptor dimerization, and the following autophosphorylation of tyrosine residues within the cytoplasmic region of receptors. SH2- or PTB-docking proteins are then recruited to the phosphotyrosine residues of ERBB receptors to activate the intracellular signaling cascades implicated in the regulation of a variety of cellular processes, such as growth, differentiation, apoptosis, adhesion and migration.

Cetuximab (Erbitux), Pertuzumab (Omnitarg) and Trastuzumab (Herceptin) are monoclonal antibodies targeted to EGF family ligands, while Gefitinib (Iressa), Erlotinib (Tarceva) and Lapatinib (GW572016) are small molecule inhibitors for ERBB family receptors (3-5). Cetuximab, Pertuzumab, Trastuzumab, Gefitinib, Erlotinib and Lapatinib are anti-cancer drugs targeted to the ERBB signaling pathways. Upregulation of AREG or TGFA, functioning as ligands for ERBB family members, is associated with Gefitinib non-responsiveness (6).

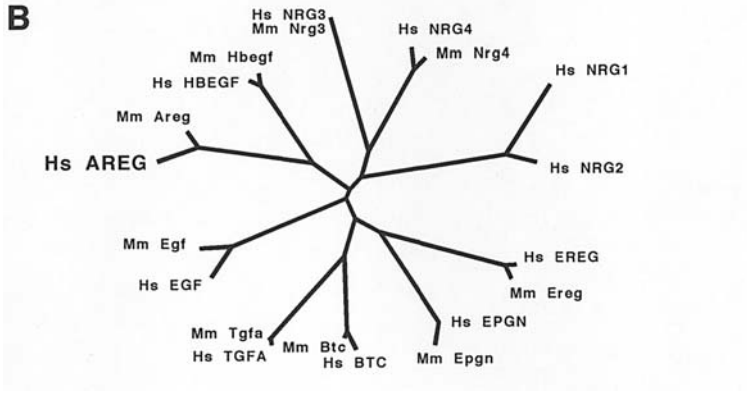
Canonical WNT signaling pathway, transduced to the ß-catenin/TCF signaling cascade, plays a key role during embryogenesis, tissue regeneration and carcinogenesis (7-14); however, WNT-dependent transcriptional regulation of *EGF* family members remains unclear. Here, TCF/LEF binding sites within the promoter region of human *EGF* family members were searched for. Because three TCF/LEF-binding sites were identified within the 5'-promoter region of human *AREG* gene, comparative genomics analyses on *AREG* orthologs were further performed.

Materials and methods

Screening of WNT target gene. Genome sequences corresponding to human *AREG, BTC, EGF, EPGN, EREG, HBEGF, NRG1, NRG2, NRG3, NRG4* and *TGFA* genes were searched for with BLAST programs (http://www.ncbi.nlm.nih.gov) as described previously (15-18). TCF/LEF-binding sites within the 5'-flanking promoter region of the above

A

Human gene	Human complete CDS	Human genome clone including promoter	TCF/LEF binding site within human promoter
AREG	NM_001657.2	AC021180.6	3
BTC	NM_001729.1	AC098825.3	0
EGF	NM_001963.2	AC005509.1	1
EPGN	NM_001013442.1	AC097470.3	1
EREG	NM_001432.1	AC021180.6	1
HBEGF	NM_001945.1	AC008438.6	1
NRG1	NM_013957.1	AC083977.5	0
NRG2	NM_004883.1	AC094080.4	0
NRG3	NM_001010848.2	AL096706.10	1
NRG4	NM_138573.1	AC087456.3	0
TGFA	NM_003236.1	AC069255.18	1



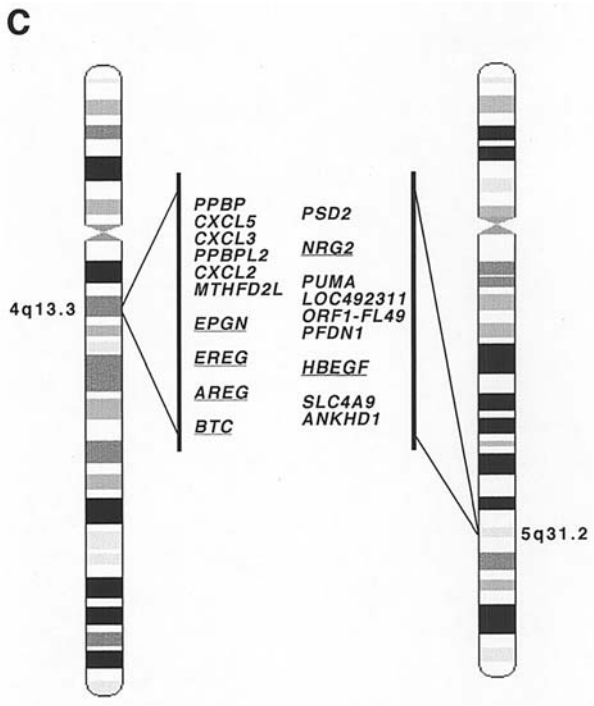


Figure 1. (A), Human *EGF* gene family. Gene symbol, complete coding sequence, genome sequence and the number of TCF/LEF-binding sites within promoter region of 11 *EGF* family genes are listed. Three TCF/LEF-binding sites exist within human *AREG* promoter. (B), Phylogenetic analysis on mammalian EGF family members. Hs, human; Mm, mouse. AREG and HGEGF are paralogs among the EGF family. (C), Intra-species comparative genomics on *AREG* and *HBEGF* loci. *EPGN-EREG-AREG-BTC* gene cluster is located around the local duplication hot spot at human chromosome 4q13.3.

genes were searched for based on bioinformatics and manual inspection as described previously (19-22).

Identification of chimpanzee AREG orthologs. Chimpanzee genome sequence homologous to human *AREG* was searched for with BLAST programs as described previously (23-26). Complete coding sequence (CDS) of chimpanzee AREG was determined by assembling exonic regions.

Comparative integromics analyses. Phylogenetic analyses on EGF family proteins as well as on promoters of mammalian *AREG* and *HBEGF* orthologs were performed by using the CLUSTALW program. Human and chimpanzee *AREG* promoters were then aligned by using Genetyx program and manual curation as described previously (27-30).

In silico expression analyses. Expressed sequence tags (ESTs) derived from human *AREG* gene were searched for by using the BLAST programs as described previously (31-34). The sources of human AREG ESTs were listed up for *in silico* expression analyses.

Results

Screening of TCF/LEF-binding site within promoter region of EGF family genes. Human AREG RefSeq (NM_001657.2), BTC RefSeq (NM_001729.1), EGF RefSeq (NM_001963.2), EPGN RefSeq (NM_001013442.1), EREG RefSeq (NM_001432.1), HBEGF RefSeq (NM_001945.1), NRG1 RefSeq (NM_013957.1), NRG2 RefSeq (NM_004883.1), NRG3 RefSeq (NM_001010848.2), NRG4 RefSeq (NM_138573.1) and TGFA RefSeq (NM_003236.1) were used as query sequences for the BLAST programs to identify genome clones corresponding to *EGF* family genes. The 5'-flanking promoter region of human *AREG, BTC, EGF, EPGN, EREG, HBEGF, NRG1, NRG2, NRG3, NRG4* and *TGFA* genes were identified within AC021180.6, AC098825.3, AC005509.1, AC097470.3, AC021180.6, AC008438.6, AC083977.5, AC094080.4, AL096706.10, AC087456.3, and AC069255.18 genome sequences, respectively (Fig. 1A). TCF/LEF-binding sites within the 5'-promoter region of human *EGF* family genes were then searched for based on manual inspection. Three TCF/LEF-binding sites were identified within human *AREG* promoter (Fig. 1A).

Comparative integromics analyses on AREG. Comparative proteomics analysis was performed at first. Phylogenetic analysis on human and mouse EGF family members revealed that AREG and HBEGF are paralogs (Fig. 1B). Intra-species comparative genomics analysis was next performed. *PUMA, PFDN1, SLC4A9* and *ANKHD1* genes were located around the *HBEGF* gene at human chromosome 5q31.2; however, paralogs of these genes were not located around the *AREG* gene (Fig. 1C).

EPGN, EREG and *BTC* genes were directly linked to the *AREG* gene, which indicated the existence of *EGF* family gene cluster. *PPBP, CXCL5, CXCL3, PPBPL2* and *CXCL2* genes were located centromeric to the *EPGN-EREG-AREG-BTC* gene cluster. The *EGF* family gene cluster and the *PPBP-CXCL* segmental duplicons were closely linked at human chromosome 4q13.3 (Fig. 1C).

Expression profile of human AREG. In silico expression analysis was performed to investigate the expression profile

SPANDIDOS PUBLICATIONS

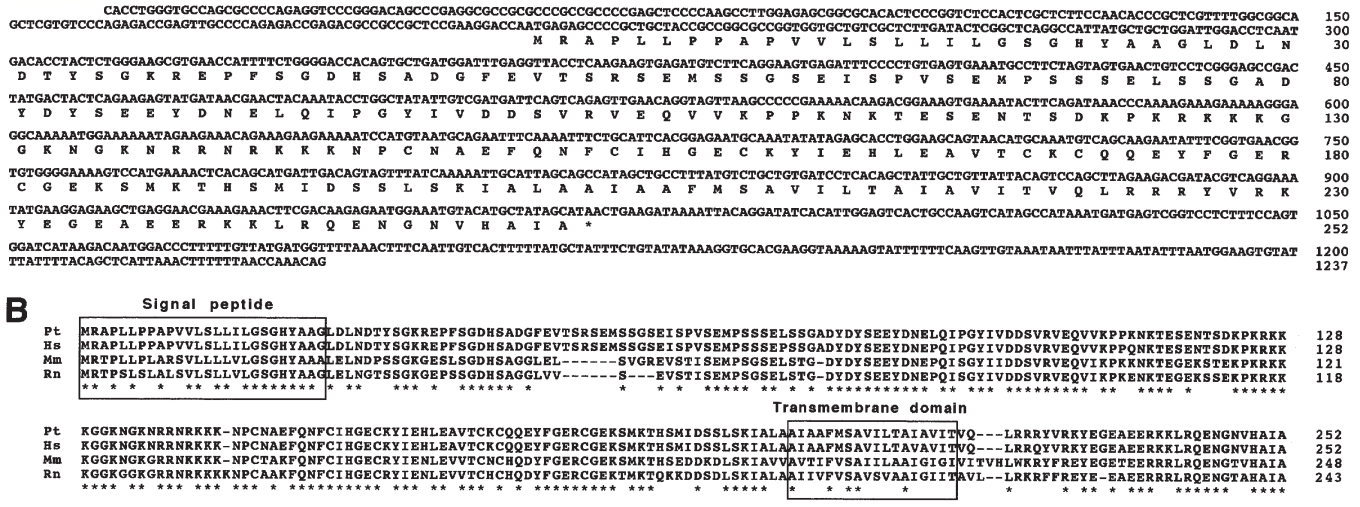


Figure 2. Chimpanzee AREG. (A), Nucleotide and amino-acid sequences of chimpanzee AREG complete CDS. Nucleotides and amino-acid residues are numbered on the right. (B), Alignment of AREG orthologs. Hs, human; Pt, chimpanzee; Mm, mouse; Rn, rat. Conserved amino-acid residues are shown by asterisks.

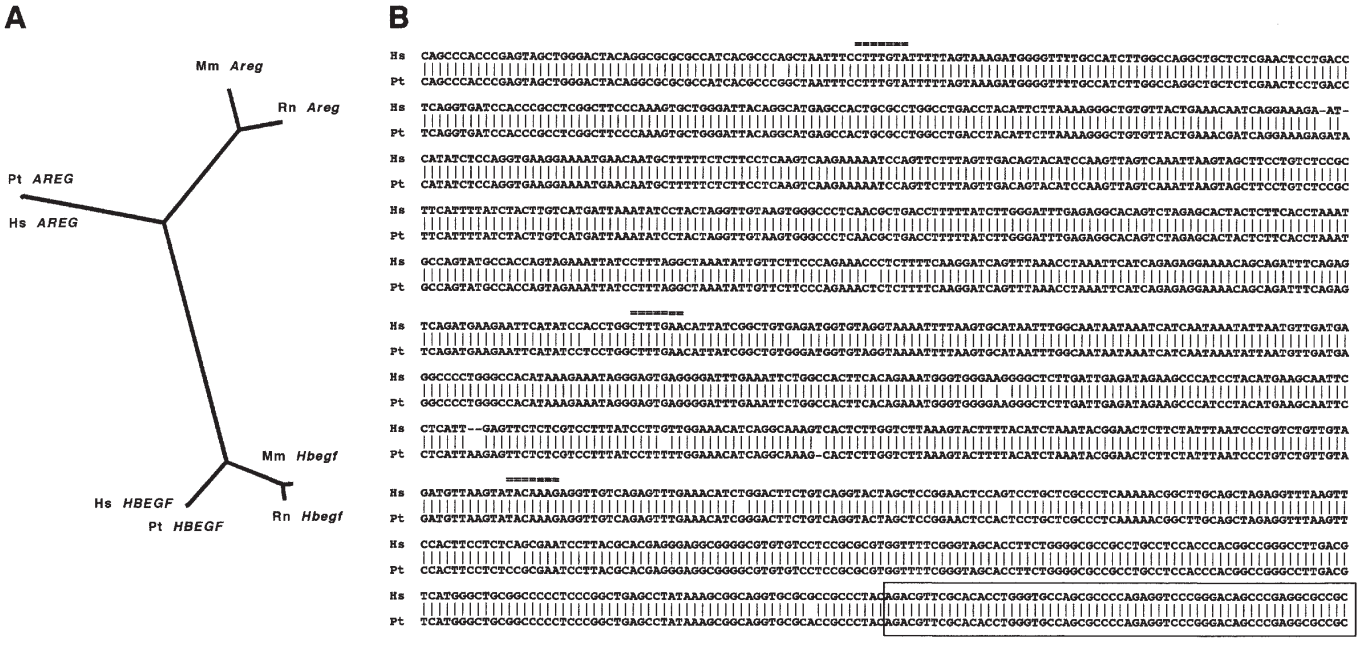


Figure 3. Comparative genomics analyses on *AREG* promoters. Hs, human; Pt, chimpanzee; Mm, mouse; Rn, rat. (A), Phylogenetic analysis on mammalian *AREG* and *HBEGF* promoters. Primate *AREG* promoters are significantly divergent from rodent *Areg* promoters. (B), Alignment of human and chimpanzee *AREG* promoters. Region corresponding to exon 1 of human *AREG* gene is boxed. TCF/LEF-binding sites are shown by double over-lines.

of human *AREG* mRNA. *AREG* mRNA was expressed in a variety of human tumors, such as colorectal cancer, liver cancer, gastric cancer, breast cancer, prostate cancer, esophageal cancer and myeloma.

Identification of chimpanzee AREG ortholog. BLAST programs using human AREG RefSeq revealed that chimpanzee *AREG* gene was located within NW_105918.1 genome sequence. Exon-intron boundaries of chimpanzee *AREG* gene were determined based on the consensus sequence of exon-intron junctions. Chimpanzee *AREG* gene was found consisting of six exons. Chimpanzee AREG complete CDS was determined by assembling exonic regions (Fig. 2A). Genetics program revealed that nucleotide position 211-969 was the coding region. Chimpanzee *AREG* gene was found to encode a 252-amino-acid AREG protein. Chimpanzee AREG was a type I transmembrane protein showing 98.0% and 71.4% total amino-acid identity with human AREG and mouse Areg, respectively (Fig. 2B).

Comparative genomics analyses on AREG promoters. Human *AREG* promoter, chimpanzee *AREG* promoter, mouse *Areg* promoter and rat *Areg* promoter were located within AC021180.6, NW_105918.1, AC115067.7 and AC115469.4

genome sequences, respectively. Human *HBEGF* promoter, chimpanzee *HBEGF* promoter, mouse *Hbegf* promoter and rat *Hbegf* promoter were located within AC008438.6, NW_107080.1, AC147220.3 and AC097756.6 genome sequences, respectively. Phylogenetic analysis on the promoters of mammalian *AREG* and *HBEGF* orthologs revealed that primate *AREG* promoters were significantly divergent from rodent *Areg* promoters (Fig. 3A).

Three TCF/LEF-binding sites within human *AREG* promoter were located about 1200, 600 and 300 bp upstream of the transcription start site (Fig. 3B). Three TCF/LEF-binding sites within human *AREG* promoter were conserved in chimpanzee *AREG* promoter (Fig. 3B), but not in rodent *Areg* promoters.

Discussion

TCF/LEF binding sites within the promoter region of human *AREG, BTC, EGF, EPGN, EREG, HBEGF, NRG1, NRG2, NRG3, NRG4* and *TGFA* genes were searched for in this study. Because three TCF/LEF-binding sites were identified within the 5'-promoter region of the human *AREG* gene (Fig. 1A), comparative genomics analyses on *AREG* orthologs were further performed.

Phylogenetics analyses revealed that *AREG* and *HBEGF* genes were paralogs within the human genome (Fig. 1B). The *EPGN-EREG-AREG-BTC* cluster at human chromosome 4q13.3 was linked to the *PPBP-CXCL* segmental duplicons (Fig. 1B). This fact indicates that human chromosome 4q13.3 was one of the hot spots for local duplications within the human genome.

Chimpanzee *AREG* gene, consisting of six exons, was found to encode a 252-amino-acid protein (Fig. 2A). Chimpanzee AREG was a type I transmembrane protein showing 98.0% and 71.4% total amino-acid identity with human AREG and mouse Areg, respectively. Phylogenetics analysis also indicated that AREG orthologs were relatively divergent compared with other EGF family orthologs. Together these facts indicate the protein evolution of AREG orthologs.

Three TCF/LEF-binding sites within human *AREG* promoter were conserved in chimpanzee *AREG* promoter, but not in rodent *Areg* promoters. Phylogenetic analysis also revealed that primate *AREG* promoters were significantly divergent from rodent *Areg* promoters. Together these facts indicate the promoter evolution of *AREG* orthologs.

In silico expression analysis in this study revealed that *AREG* mRNA was expressed in a variety of human tumors, such as colorectal cancer, liver cancer, gastric cancer, breast cancer, prostate cancer, esophageal cancer and myeloma. Kakiuchi *et al* reported that *AREG* upregulation is one of the biomarkers predicting Gefitinib non-responsiveness (6). Because human *AREG* was characterized as potent target gene of WNT/ß-catenin signaling pathway (Fig. 3B), WNT signaling activation could lead to Gefitinib resistance through *AREG* upregulation. Therefore, AREG is a target of systems medicine, especially in the field of oncology.

References

1. Yarden Y and Sliwkowski MX: Untangling the ErbB signaling network. Nat Rev Mol Cell Biol 2: 127-137, 2001.
2. Schlessinger J: Common elements in cellular signaling via EGF and FGF receptors. Science 306: 1506-1507, 2004.
3. Reichert JM, Rosensweig CJ, Faden LB and Dewitz MC: Monoclonal antibody success in the clinic. Nat Biotech 23: 1073-1078, 2005.
4. Krause DS and van Etten RA: Tyrosine kinases as targets for cancer therapy. N Engl J Med 353: 172-187, 2005.
5. Katoh M: Bioinformatics for cancer management in the post-genome era. Technol Cancer Res Treat 5: 169-176, 2006.
6. Kakiuchi S, Daigo Y, Ishikawa N, *et al*: Prediction of sensitivity of advanced non-small cell lung cancers to gefitinib (Iressa, ZD1839). Hum Mol Genet 13: 3029-3043, 2004.
7. Katoh M: *WNT* and *FGF* gene clusters. Int J Oncol 21: 1269-1273, 2002.
8. 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.
9. 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.
10. 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.
11. Katoh M: Epithelial-mesenchymal transition in gastric cancer. Int J Oncol 27: 1677-1683, 2005.
12. Katoh M: WNT2B: comparative integromics and clinical application. Int J Mol Med 16: 1103-1108, 2005.
13. 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.
14. 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.
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: Identification and characterization of human *HES2, HES3*, and *HES5* genes *in silico*. Int J Oncol 25: 529-534, 2004.
17. 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.
18. Katoh Y and Katoh M: Identification and characterization of rat *Wnt6* and *Wnt10a* genes *in silico*. Int J Mol Med 15: 527-531, 2005.
19. Katoh Y and Katoh M: Comparative genomics on *DKK1* orthologs. Int J Oncol 27: 275-279, 2005.
20. Katoh Y and Katoh M: Comparative genomics on *DKK2* and *DKK4* orthologs. Int J Mol Med 16: 477-481, 2005.
21. Katoh Y and Katoh M: Comparative genomics on *FGF16* orthologs. Int J Mol Med 16: 959-963, 2005.
22. Katoh M and Katoh M: Comparative genomics on *FGF8, FGF17*, and *FGF18* orthologs. Int J Mol Med 16: 493-496, 2005.
23. Katoh Y and Katoh M: Identification and characterization of rat *Wnt1* and *Wnt10b* genes *in silico*. Int J Oncol 26: 841-845, 2005.
24. Katoh M and Katoh M: Comparative genomics on *WNT8A* and *WNT8B* genes. Int J Oncol 26: 1129-1133, 2005.
25. Katoh M: Molecular evolution of *WNT2B* orthologs. Int J Oncol 26: 1135-1139, 2005.
26. Katoh M: Comparative genomics on *WNT3-WNT9B* gene cluster. Int J Mol Med 15: 743-747, 2005.
27. Katoh M and Katoh M: Comparative genomics on *WNT5A* and *WNT5B* genes. Int J Mol Med 15: 749-753, 2005.
28. Katoh Y and Katoh M: Comparative genomics on *WNT11* gene. Int J Mol Med 15: 879-883, 2005.
29. Katoh Y and Katoh M: Comparative genomics on *VANGL1* and *VANGL2* genes. Int J Oncol 26: 1435-1440, 2005.
30. Katoh Y and Katoh M: Comparative genomics on *SFRP1* orthologs. Int J Oncol 27: 861-865, 2005.
31. Katoh Y and Katoh M: WNT antagonist, SFRP1, is Hedgehog signaling target. Int J Mol Med 17: 171-175, 2006.
32. Katoh Y and Katoh M: Comparative genomics on HHIP family orthologs. Int J Mol Med 17: 391-395, 2006.
33. 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.
34. Katoh M and Katoh M: Comparative genomics on Eph family. Int J Oncol 28: 1243-1247, 2006.