

# Comparative integromics on FAT1, FAT2, FAT3 and FAT4

YURIKO KATOH<sup>1</sup> and MASARU KATOH<sup>2</sup>

<sup>1</sup>M&M Medical BioInformatics, Hongo 113-0033; <sup>2</sup>Genetics and Cell Biology Section,  
National Cancer Center Research Institute, Tokyo 104-0045, Japan

Received May 2, 2006; Accepted June 5, 2006

**Abstract.** WNT5A, WNT5B, WNT11, FZD3, FZD6, VANGL1, VANGL2, DVL1, DVL2, DVL3, PRICKLE1, PRICKLE2, ANKRD6, NKD1, NKD2, DAAM1, DAAM2, CELSR1, CELSR2, CELSR3, ROR1 and ROR2 are planar cell polarity (PCP) signaling molecules implicated in the regulation of cellular polarity, convergent extension, and invasion. FAT1, FAT2, FAT3 and FAT4 are Cadherin superfamily members homologous to *Drosophila* Fat, functioning as a positive regulator of PCP in the *Drosophila* wing. Complete coding sequence (CDS) for human FAT1 (NM\_005245.3) and FAT2 (NM\_001447.1) are available, while artificial CDS for human FAT3 (XM\_926199 and XM\_936538) and partial CDS for FAT4 (NM\_024582.2). Here, complete CDS of human FAT3 and FAT4 were determined by using bioinformatics and human intelligence (Humint). *FAT3* gene, consisting of 26 exons, encoded a 4557-aa protein with extracellular 33 Cadherin repeats, one Laminin G (LamG) domain and two EGF domains. *FAT4* gene encoded a 4924-aa protein with extracellular 34 Cadherin repeats, two LamG domains and three EGF domains. Cytoplasmic VCSVxPxLP and SDYxS motifs were identified as novel motifs conserved among FAT1, FAT2 and FAT3 orthologs. Domain architecture comparison and phylogenetic analysis revealed that FAT1, FAT2 and FAR3 were divergent from FAT4. *FAT1-MTNR1A* locus at 4q35.2 and *FAT3-MTNR1B* locus at 11q14.3-q21 were paralogous regions within the human genome. *FAT1* mRNA was expressed in embryonic stem (ES) cells, neural tissues, gastric cancer, pancreatic cancer, colorectal cancer, breast cancer, lung cancer and brain tumors. *FAT2* mRNA was expressed in infant brain, cerebellum, gastric cancer, pancreatic cancer, ovarian cancer, esophageal cancer, skin squamous cell carcinoma, head and neck cancer. *FAT3* mRNA was expressed in ES cells, primitive neuroectoderm, fetal brain, infant brain, adult neural tissues and prostate. *FAT4* mRNA was expressed in fetal brain, infant brain, brain

tumor and colorectal cancer. FAT family members were revealed to be targets of systems medicine in the fields of oncology and neurology.

## Introduction

*Drosophila* Frizzled, Dishevelled, Diego, Starry night (Flamingo), Van Gogh (Strabismus) and Prickle are core planar cell polarity (PCP) signaling molecules (1-7). Asymmetrical localization of Frizzled - Dishevelled - Diego - Starry night complex and Van Gogh - Prickle complex induces PCP in the *Drosophila* wing. Human WNT5A, WNT5B, WNT11, FZD3, FZD6, VANGL1, VANGL2, DVL1, DVL2, DVL3, PRICKLE1, PRICKLE2, ANKRD6, NKD1, NKD2, DAAM1, DAAM2, CELSR1, CELSR2, CELSR3, ROR1 and ROR2 are PCP signaling molecules implicated in the regulation of cellular polarity, convergent extension, and invasion (7-26). Activation of PCP signaling pathway controls tissue polarity and cell movement through the activation of RHOA, c-Jun N-terminal kinase (JNK), and nemo-like kinase (NLK) signaling cascades.

Fat, Four-jointed and Dachshous are additional PCP signaling molecules in the *Drosophila* wing (27-29). *Drosophila* Fat, functioning as a positive regulator of PCP in the *Drosophila* wing, is a member of the Cadherin superfamily. *Drosophila* Fat-like, implicated in the morphogenesis of tubular structures of ectodermal origin, is the homolog of Fat (30). Fat1, Fat2, Fat3 and Fat4 are rodent homologs of *Drosophila* Fat and Fat-like (31-33). Complete coding sequence (CDS) for human FAT1 (NM\_005245.3) and FAT2 (NM\_001447.1) are available in the public database, while artificial CDS for human FAT3 (XM\_926199 and XM\_936538) and partial CDS for FAT4 (NM\_024582.2). Here, complete CDS of human FAT3 and FAT4 were determined by using bioinformatics and human intelligence (Humint). Comparative genomics analyses, proteomics analyses and expression profile analyses on the FAT family members were then performed.

## Materials and methods

**Determination of complete CDS for FAT3 and FAT4.** Human cDNAs, expressed sequence tags (ESTs) and genome sequences, derived from *FAT3* and *FAT4* transcripts, were searched for with BLAST programs as described previously (34-36). Complete CDS of human FAT3 and FAT4 were determined by assembling exonic regions.

**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; mkatoh@seaple.ne.jp

**Key words:** bioinformatics, comparative genomics, comparative proteomics, Cadherin, integrome network, systems medicine

**Comparative genomics analyses.** Intra-species comparative genomics analyses were performed as described previously (37-39). Genome sequences corresponding to human *FAT1*, *FAT2*, *FAT3* and *FAT4* genes were searched for with BLAST programs (<http://www.ncbi.nlm.nih.gov>). TCF/LEF-binding sites within the 5'-flanking promoter region of above genes were searched for based on bioinformatics and manual inspection as described previously (40-42).

**Comparative proteomics analyses.** Domain architecture analyses of FAT family members were performed by using RPS-BLAST and PSORT II programs. Phylogenetic analysis on FAT family proteins was performed by using the CLUSTALW program. Human *FAT1*, *FAT2*, *FAT3*, rat *Fat1*, *Fat2* and *Fat3* were then aligned by using Genetyx program and manual curation as described previously (43-45).

**In silico expression analyses.** Expressed sequence tags (ESTs) derived from human *FAT1*, *FAT2*, *FAT3* and *FAT4* genes were searched for by using the BLAST programs as described previously (46-48). The sources of human ESTs derived from *FAT* family genes were listed up for *in silico* expression analyses.

## Results

**Complete CDS of human *FAT3*.** BLAST programs using rat *Fat3* RefSeq (NM\_138544.1) revealed that human *FAT3* gene was located within AP000722.5, AP000805.4, AP002514.5, AP003718.3, AC067807.10 and AP003171.2 genome sequences. Because the first exon corresponding to the 5'-UTR of human *FAT3* mRNA was not detected based on the BLAST programs, we searched for human EST spanning the 5'-UTR to identify BF953408.1 EST. By using the nucleotide sequence of BF953408.1 EST as a query sequence for the BLAST programs, the first exon of human *FAT3* gene was identified within AP003124.3 genome sequence.

Exon-intron boundaries of human *FAT3* gene were determined based on the consensus sequence of exon-intron junctions. Exon 1 was located within AP003124.3 genome sequence as mentioned above, exons 2 within AP000722.5, exon 3 within AP000805.4, exon 4 within AP002514.5, exons 5-17 within AP003718.3, exons 5-22 within AC067807.10, and exons 22-26 within AP003171.2. Exons 2, 10 and 26 were larger than 3 kb in length. Human *FAT3* gene was found consisting of 26 exons (Fig. 1A).

Because XM\_926199 and XM\_936538 were human *FAT3* predicted sequences with the artificial first exon, complete CDS of human *FAT3* was determined by assembling exonic regions (Fig. 1B). Genetyx program revealed that nucleotide position 67-13740 was the coding region. Human *FAT3* gene was found to encode a 4557-amino-acid *FAT3* protein (Fig. 1B).

**Complete CDS of human *FAT4*.** Preliminary alignment of FAT family members revealed that human *FAT4* RefSeq (NM\_024582.2) was a partial CDS, lacking N-terminal one thirds of the coding region. BLAST programs using mouse *Fat4* RefSeq (NM\_183221.2) revealed that human *FAT4* gene was located within AC079835.5, AC098865.2 and AC092629.2 genome sequences. The 5'-UTR of FAT family transcripts were

interrupted by the first intron; however, putative first exon corresponding to the 5'-UTR of *FAT4* gene was not identified in this study due to the absence of EST or cDNA. Complete CDS of human *FAT4* was determined by assembling AC079835.5 genome sequence (nucleotide position 102306-107480) and NM\_024582.2 RefSeq (nucleotide position 70-9669) (Fig. 2A). Genetyx program revealed that nucleotide position 1-14775 was the coding region. *FAT4* gene was found to encode a 4924-aa *FAT4* protein.

**Complete CDS of human *FAT2*.** Although human *FAT2* RefSeq (NM\_001447.1) spanned the entire coding region, its 5'-end did not extend to the first exon corresponding to 5'-UTR. BLAST programs revealed that DA102144.1 EST spanned to the missing first exon. Therefore, full-length complete CDS of human *FAT2* was determined by assembling DA102144.1 EST and NM\_001447.1 RefSeq (Fig. 2A).

**Comparative genomics analyses on FAT family.** *FAT1* gene at human chromosome 4q35.2 was linked to the *MTNR1A* gene, while *FAT3* gene at human chromosome 11q14.3-q21 was linked to the *MTNR1B* gene (Fig. 2B). Third *MTNR1* family gene linked *FAT2* or *FAT4* gene was not identified by using the BLAST programs. Based on these facts, it was concluded *FAT1-MTNR1A* and *FAT3-MTNR1B* loci were paralogous regions within the human genome (Fig. 2B).

Full-length complete CDS of human *FAT1*, *FAT2* and *FAT3* (Fig. 2A) were used as query sequences for the BLAST programs to identify genome clones corresponding to *FAT* family genes. The 5'-flanking promoter regions of human *FAT1*, *FAT2* and *FAT3* genes were identified within AC107050.3, AC011374.6 and AP003124.3 genome sequences, respectively. Repetitive sequence was located within the 5'-promoter region of *FAT3* gene. TCF/LEF-binding sites within the 5'-promoter region of *FAT1*, *FAT2* and *FAT3* genes were then searched for based on manual inspection. One TCF/LEF-binding site was located about 1100 bp upstream of the transcription start site of *FAT1* gene, and also about 800 bp upstream of the transcription start site of the *FAT2* gene.

**Comparative proteomics analyses on the FAT family.** *FAT1*, *FAT2*, *FAT3* and *FAT4* are type I transmembrane proteins. Extracellular region of *FAT1* and *FAT3* consisted of 33 Cadherin repeats, one Laminin G (LamG) domain and two EGF domains. Extracellular region of *FAT2* consisted of 32 Cadherin repeats, one LamG domain and two EGF domains. Extracellular region of *FAT4* consisted of 34 Cadherin repeats, two LamG domains and three EGF domains. Phylogenetic analyses revealed that *FAT1*, *FAT2* and *FAT3* were divergent from *FAT4* (Fig. 2C).

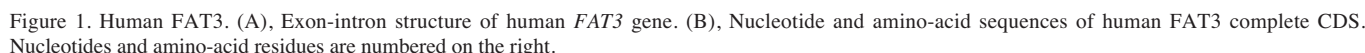
Although extracellular region was relatively well conserved among FAT family members, cytoplasmic region was divergent. Based on the alignment of C-terminal part of FAT family members, cytoplasmic VCSVxPxLP and SDYxS motifs were identified as novel motifs conserved among *FAT1*, *FAT2* and *FAT3* orthologs (Fig. 2D).

**Expression profile of human FAT family members.** *In silico* expression analyses were performed to investigate the

**B**

CCGACCGGAGAGCGCCCTCGTAGAGCCGCTGGATCGCGACGAGGACGAGGAGTGGAAATATGATGATGATGATTAATATATGGACACTGCTGTGGGCACACGCGCTCTGCTCTGTTGGCTCATCTCTGCTCTTTCAAGCTTTTGGCCACT 150  
MDIIMGHCVGTRFPACCLLFLKLLAT  
CTCTCCAGGGGCTGCCAGGAGCTGGACCCCTGGGCTTCCACTTCCACACATTCATTTATAATGCTACCTGCTAGAGCTCAGACGAGGAGCTACGCTCAGACGAGGAGTAGAATGGCATCAGCTTTAATGATCATCTGGGAT 300  
VBSQGLLPGTGTLPLFTHTHSIYNATVYENSAAATYVNSQSRMGITLLIDLBSWD  
ATCAAAATCTGCTCTCGGAGACGAGGAAGCTTTCAAAGCAGAGGAAGTATCATCTACAGATTTGCTTTTCCAGAAATAGAACTAAAGAGTGGCAATTCCTGCATATTAATAGGAAATAGCAAGATAATTTATTATTGATA 450  
IXYRIVSGDZGDEKGFFTKAEVVIADDFCLRLIRTKCGKGNSSAILNREIQDNYLLI  
GTAAAGATCTCTGTCAGAGGAGAGGATTTGAAGACATGGACCAAAGTGAATACACAGGTTTATGATATGAATGATCGAGACCTTTGCTTTACCACCACCACTACTCTGTACCATACGAGAAAGACACCTCTAAGGACTAGTGTGCC 600  
VKGSVRGEGEDLLEAWTKVNLQVLLDMNDLRLPLFSTPTTSYVTIAEASSSTPLRTSVA  
CAGGTACTGCACAAAGCCAGATATGGTTCGAATGGGAATCTACTACTACTTTAAAAATAAAGTTGATCTCTTTTCACTACCCGCCAGAGTGCTGTATCTCTTAAGTGCTGATTAATATGATGAAAAGATAGTGTATGAT 750  
QVATAADADIQSNGEFFYYFKNKVVDLFLSVHPTSSGVVISLSGLRLNYDEKKNRYD 228  
CTGAAATTTTGGCTGTGGACCGGGAAATGAACCTGTATGGCAACATGGAATGACGACTGCAAGAGCTTTATTTCCACTTTCAGGCGCATAAATGAACATCCATGATGACTCATGCTATGCTCTCTCTCTGGAAAAA 900  
LZELLAVDRGMKLTLYGMNMGVBSSTAKLYVHIERINEMHAPTTHVTVHVPFSLKEK  
GAGCCACATCATGCACTGGTGCAGATTTGATGAGCTTAGATGATGAGGACGAATGGAGAGATGCAATCTGTTTCATTGTGGCTGGGGATCTTTAGATCAGTTTCTCTGGCTTAAAGGAGGAATGGTGTGAATAGTACAGAAATAGAGAG 1050  
EPTYAVAVDLDLDDGAGXIESVSI VAGDPLDCTFFFLAKGEGKWLXGYIKIKE  
AGGAAGCAGATTTACTGTGGAGAGCTTTCCCTATGTGCTACATCTTCAAAGCAAGACAGGAGTCTCTCAAAAATTTCCAGCATTAAGAGCGATCTACATGTCGCAACCCACAGAGACACTGTCOCCTATGATTTGAAAAA 1200  
RKQIDWNESFPYGYNTLLQAKDKGSPQKCSALKAVIIGNPTRDTPVIRPFEK  
GAAGTGTAGATGTGAGCATAGTGAATTTTCCCTCTGCTGTGTGTTCTGATATGAAATTAAGTCTGTAACCGATAGATGTGGAATACAAATATCTCTGTTGAGGAGATGAGTGTATCTTTAAATTAATRCCTCGCTCGGCTG 1350  
XVYDVSISIEFSPFPVVVAIVKLSPEPIDVEYATKLSPLGEGDAVATFYKINPFRSGCL 1425  
ATTGTTACAGACGGCCATTAATCTGTTAAAGAGAGGTTTAAATCGAGGTGACAAACAGAGAGAGAGATTTAAAGACAGAGCTCACATCAGCATAGAGATCCAAATGACCCACAGAGATTTACAGACAGATGTATGAT 1500  
IVTARPLNTVYKKEVYKLEVTNKNKRGDLKAQVYITISIEDANDHTPEFQQLPYD 1575  
GCTTTATGAAATGAAGTGTCCGAGTGGAAACAGGCTCTCAACAGTTTCAGCTCTGATAGGATAGAGAGAGAAATGGTATCATCAGCTATAGTCTAGCTGAGCTGAATTTGTATACATTTGTCATTAATCACTTACAGGTGTATT 1650  
AYVNESVVPVGTSTSVLTVSASBDKKGKNGTYITYSIASLNLPLFVINYNQFTGVYI 1625  
AGCACAACTGAAGAACTGGATTTTGAATCCTCCCGCAAAATTTACAGATTCACTTTAGAGGCTCTGACTGGGGTTCCACCATACCGCCCATGAAGTGAGTCAATGTGACTATTGCAATAGGAATGTCAACAGACACAGCCCTCTCTT 1800  
S T X E L D F E S S P E I Y R F I V R A S D W G S P T R H E S V N W T I R I G N W N D N S P L F 1575  
GAAAAAGTCTGCCAGGAGCTATTTCATATGACTTCCAGTGTGGTGTGCATACAGACGAGCTCTCAGCGCTGATATGATGAATCTGAACTTGAATCTTTGCGAAGATGAATCTGGCTCTTTTATTAAACCCA 1900  
EKVACQCGDSYVISYDFFVVGXHTIVASAIIDIDELKLVKIKIISGNKELGFFYILNP 2100  
GATTCGTGTTTACGCGTAAAAATCACTGCAGAAATCTGCGCATTAATAATGGCAATTTGGCCTCAGAAATACAGCAACTGATGGAGAGATCTTCCAGACCCCATCTATTACATCTTCACTGTACATGTGGAAAGTGTCTTCA 628  
DSGVLLQKKLSKLTNSGQIKNGNFPALRLITATTDGKNLADPMSINISV LHQKVSS 628  
AAGGCTCTCAAGTCCAGAGAACTCGTGTGCTCAAAAGCTGCCAGAGAACTCAATTAAGCGAAAGCAAAAGCAAACTGAACTGTGAGTGTGATTTTCACTTAATAGACAGAGGACATATTTCAGACAGTCT 2250  
KSFSGRCRTTRVAKQKLAKEKLLIKAKANGKLNLEDDGFLDIFYTISINRQGPFIYFDKS 728  
TTTCTCTGATGTGTGTAAGAGAGATCTCCGAGTGTGCTCAACATCTGGAAGATTAAGCGCTATGATGCGCACTCTGCTGCTCAATGGAAGTGCTATTACAACTACAGATGGAATACAGATATGCTCTTAATATTGATATG 2400  
F P S D V A V K E D L P V G A N I L K I K A T I D A D S G F N G K V L F T T I S D G N T D S C F N I D M 2750  
GAGACTGGGCACTTAAGTCTTTATGCCATGGATGGAGAACACAGACTCTTATCTCCTTAATATCACCATCTATGACTTAGTGTATATCCAGAAATCTGTCAAGAGACTGTGACCATCAATGTGGAGGATCTAATGACATAGC 2750  
ETGQLKVLMPMDRREHTDLYLLNITITIDLGNPQKSSWRLTTLINVDANDNATAGC 828  
CCAGTTTATTATGAAGCATGTATCTAGCTTAACATTTCTGAAAGTTCAGGCATGTGATCGAAATCATCAAGTGGAGGCCAGAGACAAAGCATTAGTGTCTAATGTGAAAGTGAATCTCACTCACTGTTCAGACAGATACACAGACTTGGCC 878  
PVFIQDSDSYISVNI L E S S G I G T E I Q V E A R D K D L G S N G E V T Y S V L T D T Q Q F A 2700  
ATCAATAGCTCAACTGGAATCTTTATGTGAGCGACAGCTGTGGACGGGAATCAAAAGCAAAATCTCTTGAATAAGAACAGAGGACAGAGGAGAGAGTGTGACAGCGCTTTTTCAGTGTGTCAATCTTAAAGTCTTTTATGATGAT 2850  
INSSTGTIVYVADQLDRLRESKANYSLLKIEARDKAKESGQQLFVSVTLKVLFLD 2925  
GTCAATGCTGCTCCGACGCTTCATTCGACAGTACTATGTGTGAAGGTTCTTGAAAGATCTCCTCTTGGACATCTGATCTGCTGCTGGAGCCGATGATCAGATCTTGGAGTGGGGGTCAAGTGCGCATCTTTTGGTCAATGAC 3000  
VNDCTSPAFIPISYSVVKVLEDLPVGTVIANLETHDPLDGLGLGQVRYVSLVND 3078  
TATAATGGGATTTGAAATAGATAAAGCAAGTGTGCCATCCGCTTGAGCAAGAGCTTGATATGAGAAACAGAGCTCTATAAGCTTACTGTGCGGGCCAAAGACAGAGGGCGGCTGTCTCTGTGCATCTGTTTCTCTTTGATGAT 3125  
YNGRFEXIDKASGAIRLSKELDYEKQQLYINLTVRAKDKGRPVLSLSVSVFE 3105  
GTGAAAGTGTGATGTCAATGAAACCTCCACACTGCTATTTCOCAGAGTTTCTGTGTGATGTGTGAAGGAAACCTCAGCATTTGGAACAAAGCTGCTGCTGAGTCTGCTGAGCATGAAGACTCCGGAAGGGATGGAGAGAT 3300  
VEVVDVNVKNMLHTPYFPDFAVVGVSVKKNRSRIGTISVLLQVTAARDKEDSGRDKI 3307  
CAGTACTCCATCAGGAGTGCAAGTGTCTTGGAAAGTTCAGTATAGACAGCAGAGATGGGGTTCATCCTCCGCGAGACATTTCTGATGGGAGACAAATGGGTCATCTGGCTTAACAGTGTATGCCACAGACAGGGGCGTTGTCCACTC 3450  
QYISIRIDGSGGLGRFSIDDESQVITTAADILDRCTEMGSSWLTATVIAKPLDAEKQRSI 3425  
TACTCCAGCAATAGAGTGTATCATGGAATGTGAAGATGTGAATGACAAATCCCGCGTGCAGCTCAGAACATATATATTCTGTTGTGATGGAACCTCCCAAGAGCATATCTGTCATTAGATCCAGGCTGGAAGATCTGTCAGTCACT 1168  
YSTITCTGATSYIEVEDVDNDNAPLTSKPIIYYPVVMKNESPKDVSIVIQIAEKDPPDSBS 1120  
TCCAAATGAAAACGTGACATACAGATTAACAAGTGAATCTCCAGAAATTTTTCGCTCAATCAATAACAAACAGTGTGATTACACAACTCTCAAGAGAAATGGATGGAGAACAGAGGACAGACTTTCTCGAGGTGCTGTGACAGAT 3750  
SNKLLTYRTIRTSGNPNQNFIAINIKTGLLITTSRKLLDRKEQQAKEHFLVETVTD 1228  
GGTGGCTCTCTCAAAACGATCAACATTTGGTGGTGTCTGAGTCTGATGAAATGACACAAAGCCGAGTCCGACAGAGGTTCTACAGCAAGCTCGAGAGAGTGGAGCAAGAGAGCCGATTCACAGGCT 3900  
GGPSPKQGSTIWWVVQVLDENDNKPQFPFEKVIYQIKLPERADRKKRGEPIYRA 1278  
TTTGCACTGATGAGAGTGAGGGCCCAACCGAGAAATCTGCATAGATTTGTGGATGGGAATGATGACAGAAAGATCTTTATTGAGCCATAAAGTGGATGTTCTTCTTCAAGAACAGCTTTACAGACAGCGATTAAGACATCTTAAG 4050  
FAFDRDRDEQPNNAEISIVDGNDDGKFFLVIDPKTKMVSBRKQFTTAGSVIDILT 4320  
ATAAGCACTGGAGAAATGGGGCCACAGAAATCTCCAGGGCGCGCTCCAGATGAATGAGATTAAGAAACACCCGCTCAGCATGACCTGAGTGAAGGCTTTATACTTCAAGCTACGGAAGTGAATGAGTGACT 1307  
ITAAAGDANGRFPQKSSSTARLHIEWIKKPPPSFIFLTFDEPFYVNFVMESDVT 1371  
GAAATTTAGGGGTGTGTCTGTGCGAGCCGATTAACACCCCTGTGTGTTGACATAGTTGGGGGAATTTGACAGCGGCTTTGATGACAGAGAGGGTGTGGGCAATTTGATCTGCGAAAGGCTTTGGATGACAGAGAGTGTCCATC 1450  
XIVGVTVSGQFANPTPLWFDIVGNGNFDLSAFDAEKGVGTIVIAKPLDAEKQRSI 3425  
TATAATTAGAGTGTGAAGTCCAGGATGGGCAAAATGTCTGTTTACTCAGGTATTATTCAAAAGTCTGGATATATGATATGAAATGGCCAGAAATCTCTCAGCGCAATACGATGTGACAAATTCGAGAGATGTCTCCAGACAGGGAG 1450  
YNMSVEVTDGTGNVAATQVFIKVLNDNDNGPEFSQPNYIDVTISEVDVLPDTE 1487  
ATCTCTCAAGATTAAGCCACAGATAGAGATGAGAGACAGAAAGCTGAGCTACACTGTTCTATGACAGCATGAGCTCCATCAGCATGAGAGAAATTCGGAATGACCTGAGCTGCGGCTCTATACCTGCCAGAGGCTGGACATGAGGCC 4650  
ILQIXEATDRDRGEEKHKLSTVHSIDSISIMRKNFRIDPSTGVLTYATERLDHEA 1521  
CAGGACAGACCATTTCAACATATATGTCAGAGATCAGAGATGCTTTTATCGAGAGAAATCTGGCCGAGTCACTTGTGAATGTGGAGATCTCAATAGTCTCTATTATTCACACCCCATGTATGAGGCTGTGTGTTGAATCT 4800  
QDKHILNIMVVRDQEFYFYRRNRLARVINVVEDANDHSPIFYTNPLIEASVFS 1571  
GCTGCTCTGGGATCAGCTGTTCTGCAAGTGCAGGCTCTGGACAGAGCAAGAGAGAAATTCAGAACTATATACATAGAGAGAGAAACCTGGGAACATTTGAAGATGGAACGGCTGTACGAGCATGATTCAGATTTGCAAGAA 1650  
AALGSSAVLLQVLTALDKDKGNENAEELIYITIEAGNNTGNMFPYKIEKPVLLIITICKE 1492  
CCAGACATGACAGGATGGTCAGTTTGTCTATCCATCAAAAGTCACAGATCAGGGATCCCGCCCAATGTGTGCTACTGCAATTTGCGCATTTCCGTCAGCATGTGTGACAATTTCTACGCCCAAGTTTCATTCAGAAAGACTACCAAGCA 5100  
PDMT





cancer, lung cancer and brain tumors. *FAT2* mRNA was expressed in infant brain, cerebellum, and also in a variety of tumors, such as pancreatic cancer, diffuse type gastric cancer, ovarian cancer, esophageal cancer, skin squamous cell



Human gene	Alias	Chromosomal localization	Full-length complete CDS
<i>FAT1</i>	<i>CDHF7</i>	4q35.2	NM_005245.3
<i>FAT2</i>	<i>CDHF8</i>	5q39.1	Assembly of DA102144.1 and NM_001447.1
<i>FAT3</i>	<i>CDHF15</i>	11q14.3 - q21	Figure 1B
<i>FAT4</i>	<i>CDHF14</i>	4q28.1	Assembly of AC079835.5 and NM_024582.2

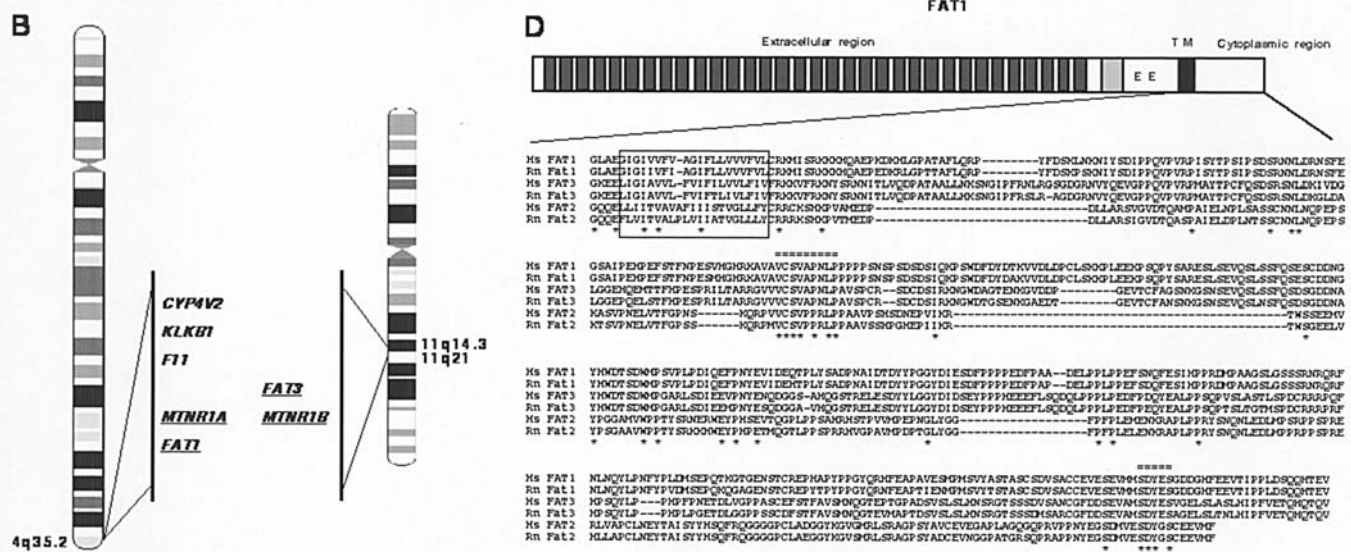


Figure 2. Comparative integromics on FAT family. (A), Human *FAT* gene family. Gene symbol, alias, chromosomal localization and full-length complete coding sequence of *FAT* family genes are listed. (B), Intra-species comparative genomics on *FAT1* and *FAT3* loci. *FAT1*-*MTNR1A* locus at 4q35.2 and *FAT3*-*MTNR1B* locus at 11q14.3-q21 are paralogous regions within the human genome. (C), Phylogenetic analysis on human *FAT* family members. *FAT1*, *FAT2* and *FAT3* are divergent from *FAT4*. (D), Schematic representation of *FAT3* and partial alignment of *FAT1*, *FAT2* and *FAT3* orthologs. Extracellular region, consisting of Cadherin repeats (dark gray box), Laminin G domain (light gray box) and EGF repeat (E), is well conserved among *FAT* family members; however cytoplasmic region is divergent. The regions around transmembrane domain and cytoplasmic region of *FAT1*, *FAT2* and *FAT3* orthologs are aligned. Hs, human; Rn, rat. Conserved amino-acid residues are shown by asterisks. Two novel cytoplasmic motifs VCSVxPxLP and SDYxS are shown by double overlines.

carcinoma, head and neck cancer. *FAT3* mRNA was expressed in ES cells, primitive neuroectoderm, fetal brain, infant brain, adult neural tissues and prostate. *FAT4* mRNA was expressed in fetal brain, infant brain, brain tumor and colorectal cancer.

## Discussion

Complete CDS of human *FAT3* and *FAT4* were determined for the comparative integromic analyses on *FAT* family members in this study. *FAT3* gene, consisting of 26 exons, was found to encode a 4557-aa protein with extracellular 33 Cadherin repeats, one LamG domain and two EGF domains (Fig. 1). *FAT4* gene was found to encode a 4924-aa protein with extracellular 34 Cadherin repeats, two LamG domains and three EGF domains. Extracellular region of *FAT* family members with Cadherin repeats, LamG domain and EGF domains was relatively well conserved; however, cytoplasmic region was divergent. Cytoplasmic VCSVxPxLP and SDYxS motifs were identified as novel motifs conserved among *FAT1*, *FAT2* and *FAT3* orthologs (Fig. 2D). Domain architecture comparison and phylogenetic analysis revealed that *FAT1*, *FAT2* and *FAT3* were divergent from *FAT4*.

Intra-species comparative genomics revealed that *FAT1*-*MTNR1A* locus at 4q35.2 and *FAT3*-*MTNR1B* locus at 11q14.3-q21 were paralogous regions within the human genome (Fig. 2B). Comparative proteomics analyses revealed that C-terminal PDZ-binding motif was conserved among

*FAT1* and *FAT3* orthologs. Together, these facts indicate that *FAT1* and *FAT3* are paralogs.

*FAT1* mRNA was expressed in ES cells, neural tissues, and a variety of tumors. *FAT2* mRNA was expressed in infant brain, cerebellum, and a variety of tumors. *FAT3* mRNA was expressed in ES cells, primitive neuroectoderm, fetal brain, infant brain, adult neural tissues, and prostate. *FAT4* mRNA was expressed in fetal brain, infant brain, brain tumor and colorectal cancer. *FAT* family members are implicated in the early embryogenesis and neurogenesis.

The canonical WNT signaling pathway cross-talks with FGF, Notch, Hedgehog and BMP signaling pathways during embryogenesis, chronic persistent inflammation and carcinogenesis (49-56). Because canonical WNT signaling activation leads to transcriptional activation of target genes through the TCF/LEF,  $\beta$ -catenin, Legless and Pygo complex, TCF/LEF-binding site within the promoter region of human *FAT* family genes were investigated. A single TCF/LEF-binding site was identified within *FAT1* and *FAT2* promoters; however, the relationship between the canonical WNT signaling activation and the expression of *FAT1* or *FAT2* in a variety of human tumors remains to be elucidated.

*Drosophila* Fat functions as a tumor suppressor (27). Tumor suppressor gene is inactivated due to epigenetic change (CpG hypermethylation) and genetic alterations (mutation or deletion) during multi-stage carcinogenesis (57,58). Epigenetic changes and genetic alterations of *FAT* family genes in a



variety of human tumors should be investigated in the future. FAT family members were revealed to be targets of systems medicine in the fields of oncology and neurology.

## References

- Adler PN: Planar signaling and morphogenesis in *Drosophila*. *Dev Cell* 2: 525-535, 2002.
- Vinson CR, Conover S and Adler PN: A *Drosophila* tissue polarity locus encodes a protein containing seven potential transmembrane domains. *Nature* 338: 263-264, 1989.
- Wolff T and Rubin GM: *Strabismus*, a novel gene that regulates tissue polarity and cell fate decisions in *Drosophila*. *Development* 125: 1149-1159, 1998.
- Heitzler P, Coulson D, Saenz-Robles MT, *et al*: Genetic and cytogenetic analysis of the 43A-E region containing the segment polarity gene *costa* and the cellular polarity genes *prickle* and *spiny-legs* in *Drosophila melanogaster*. *Genetics* 135: 105-115, 1993.
- Chae J, Kim MJ, Goo JH, *et al*: The *Drosophila* tissue polarity gene *starry night* encodes a member of the protocadherin family. *Development* 126: 5421-5429, 1999.
- Feiguin F, Hannus M, Mlodzik M and Eaton S: The ankyrin repeat protein Diego mediates Frizzled-dependent planar polarization. *Dev Cell* 1: 93-101, 2001.
- Katoh M: WNT/PCP signaling pathway and human cancer. *Oncol Rep* 14: 1583-1588, 2005.
- Saitoh T and Katoh M: Frequent up-regulation of *WNT5A* mRNA in primary gastric cancer. *Int J Mol Med* 9: 515-519, 2002.
- Saitoh T and Katoh M: Molecular cloning and characterization of human *WNT5B* on chromosome 12p13.3 region. *Int J Oncol* 19: 347-351, 2001.
- Katoh M and Katoh M: Comparative genomics on *WNT5A* and *WNT5B* genes. *Int J Mol Med* 15: 749-753, 2005.
- Kirikoshi H, Sekihara H and Katoh M: Molecular cloning and characterization of human *WNT11*. *Int J Mol Med* 8: 651-656, 2001.
- Katoh Y and Katoh M: Comparative genomics on *WNT11* gene. *Int J Mol Med* 15: 879-883, 2005.
- 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.
- Kirikoshi H, Koike J, Sagara N, Saitoh T, Tokuhara M, Tanaka K, Sekihara H, Hirai M and Katoh M: Molecular cloning and genomic structure of human *Frizzled-3* at chromosome 8p21. *Biochem Biophys Res Commun* 271: 8-14, 2000.
- Kirikoshi H, Sekihara H and Katoh M: Expression profiles of 10 members of *Frizzled* gene family in human gastric cancer. *Int J Oncol* 19: 767-771, 2001.
- Masiakowski P and Carroll RD: A novel family of cell surface receptors with tyrosine kinase-like domain. *J Biol Chem* 267: 26181-26190, 1992.
- Katoh M and Katoh M: Identification and characterization of rat *Ror1* and *Ror2* genes *in silico*. *Int J Mol Med* 15: 533-538, 2005.
- Katoh M and Katoh M: Comparative genomics on *ROR1* and *ROR2* orthologs. *Oncol Rep* 14: 1381-1384, 2005.
- Katoh M: Molecular cloning and characterization of *Strabismus 2* (*STB2*). *Int J Oncol* 20: 993-998, 2002.
- Katoh M: Structure and expression of *Strabismus 1* (*STB1*) gene on human chromosome 1q21-q23. *Int J Oncol* 20: 1197-1203, 2002.
- Katoh M: *Strabismus* (*STB*)/*Vang-like* (*Vangl*) gene family (Review). *Int J Mol Med* 10: 11-15, 2002.
- Katoh Y and Katoh M: Comparative genomics on *VANGL1* and *VANGL2* genes. *Int J Oncol* 26: 1435-1440, 2005.
- Katoh M and Katoh M: Identification and characterization of human *PRICKLE1* and *PRICKLE2* genes as well as mouse *Prickle1* and *Prickle2* genes homologous to *Drosophila* tissue polarity gene *prickle*. *Int J Mol Med* 11: 249-256, 2003.
- Katoh M and Katoh M: Identification and characterization of rat *Ankrd6* gene *in silico*. *Int J Mol Med* 15: 359-363, 2005.
- Lu X, Borchers AG, Jolicoeur C, *et al*: PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates. *Nature* 430: 93-98, 2004.
- Wu Q and Maniatis T: A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97: 779-790, 1999.
- Bryant PJ, Huettnner B, Held LL Jr, Ryerse J and Szidonya J: Mutations at the fat locus interfere with cell proliferation control and epithelial morphogenesis in *Drosophila*. *Dev Biol* 129: 541-554, 1988.
- Cho E and Irvine KD: Action of fat, four-jointed, dachsous and dachs in distal-to-proximal wing signaling. *Development* 131: 4489-4500, 2004.
- Fanto M and McNail H: Planar cell polarity from fly to vertebrates. *J Cell Sci* 117: 527-533, 2004.
- Castillejo-Lopez C, Arias WM and Baumgartner S: The fat-like gene of *Drosophila* is the true ortholog of vertebrate fat cadherins and is involved in the formation of tubular organs. *J Biol Chem* 279: 24034-24043, 2004.
- Ponassi M, Jacques TS, Ciani L and French Constant C: Expression of the rat homologue of the *Drosophila* fat tumour suppressor gene. *Mech Dev* 80: 207-212, 1999.
- Rock R, Schrauth S and Gessler M: Expression of mouse *dchs1*, *fjx1* and *fat-j* suggests conservation of the planar cell polarity pathway identified in *Drosophila*. *Dev Dyn* 234: 747-755, 2005.
- Tanoue T and Takeichi M: New insights into Fat cadherins. *J Cell Sci* 118: 2347-2353, 2005.
- Katoh M: Paradigm shift in gene-finding method: from bench-top approach to desk-top approach. *Int J Mol Med* 10: 677-682, 2002.
- Katoh M and Katoh M: Identification and characterization of human *HES2*, *HES3* and *HES5* genes *in silico*. *Int J Oncol* 25: 529-534, 2004.
- 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.
- Katoh Y and Katoh M: Identification and characterization of rat *Wnt1* and *Wnt10b* genes *in silico*. *Int J Oncol* 26: 841-845, 2005.
- Katoh M and Katoh M: Comparative genomics on *WNT8A* and *WNT8B* genes. *Int J Oncol* 26: 1129-1133, 2005.
- Katoh M: Comparative genomics on *WNT3-WNT9B* gene cluster. *Int J Mol Med* 15: 743-747, 2005.
- Katoh Y and Katoh M: Comparative genomics on *DKK1* orthologs. *Int J Oncol* 27: 275-279, 2005.
- Katoh Y and Katoh M: Comparative genomics on *DKK2* and *DKK4* orthologs. *Int J Mol Med* 16: 477-481, 2005.
- Katoh Y and Katoh M: Comparative genomics on *FGF16* orthologs. *Int J Mol Med* 16: 959-963, 2005.
- Katoh M and Katoh M: Comparative genomics on *FGF8*, *FGF17* and *FGF18* orthologs. *Int J Mol Med* 16: 493-496, 2005.
- Katoh M: Molecular evolution of *WNT2B* orthologs. *Int J Oncol* 26: 1135-1139, 2005.
- Katoh Y and Katoh M: Comparative genomics on *SFRP1* orthologs. *Int J Oncol* 27: 861-865, 2005.
- Katoh Y and Katoh M: Comparative genomics on HHIP family orthologs. *Int J Mol Med* 17: 391-395, 2006.
- 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.
- 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.
- Katoh M: *WNT* and *FGF* gene clusters. *Int J Oncol* 21: 1269-1273, 2002.
- 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.
- Garciaadiego-Cazares D, Rosales C, Katoh M and Chimal-Monroy J: Coordination of chondrocyte differentiation and joint formation by  $\alpha 5 \beta 1$  integrin in the developing appendicular skeleton. *Development* 131: 4735-4742, 2004.
- Katoh M: Epithelial-mesenchymal transition in gastric cancer. *Int J Oncol* 27: 1677-1683, 2005.
- Katoh M: WNT2B: Comparative integromics and clinical application. *Int J Mol Med* 16: 1103-1108, 2005.
- 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/ $\beta$ -catenin signaling. *Cell Commun Signal* 3: 12, 2005.
- Katoh Y and Katoh M: Hedgehog signaling pathway and gastric cancer. *Cancer Biol Ther* 4: 1045-1049, 2005.
- Katoh Y and Katoh M: WNT antagonist, *SFRP1*, is Hedgehog signaling target. *Int J Mol Med* 17: 171-175, 2006.
- Katoh M and Katoh M: Pharmacogenomics on gastric cancer. *Cancer Biol Ther* 3: 566-567, 2004.
- Katoh M and Katoh M: Bioinformatics for cancer management in the post-genome era. *Technol Cancer Res Treat* 5: 169-176, 2006.