FEBS Letters 571 (2004) 74–80 FEBS 28608

# hOLF44, a secreted glycoprotein with distinct expression pattern, belongs to an uncharacterized olfactomedin-like subfamily newly identified by phylogenetic analysis ☆

Ling-Chun Zeng[a,b], Feng Liu[b], Xin Zhang[b], Zhi-Dong Zhu[b], Zhi-Qin Wang[b], Ze-Guang Han[b], Wei-Jun Ma[a,b,*]

[a] *Health Science Center, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences and Shanghai Second Medical University, 225 South Chongqing Road, Shanghai 200025, China*
[b] *Chinese National Human Genome Center at Shanghai, 351 Guo Shou-Jing Road, Shanghai 201203, China*

Received 14 May 2004; revised 23 June 2004; accepted 24 June 2004

Available online 4 July 2004

Edited by Takashi Gojobori

**Abstract Secreted proteins are indispensable for the development and differentiation of multicellular organisms. Cloning and characterization of novel or hypothetical genes encoding these proteins are therefore inviting great incentives. Using bioinformatics tools and experimental approaches, we isolated and characterized a human secreted glycoprotein, hOLF44, which contains a highly conserved olfactomedin-like (OLF) domain in the C-terminal. However, phylogenetic analysis revealed that hOLF44 is not clustered into any of the OLF subfamilies containing characterized members, and obviously falls into a newly identified uncharacterized OLF subfamily. Western blot analysis showed that hOLF44 protein is robustly secreted from the transfected COS-7 cells. Expression levels of *hOLF44* mRNA are abundant in placenta, moderate in liver and heart, whereas fairly weak in other tissues examined. Immunohistochemical study on human term placenta demonstrated that hOLF44 is mainly localized extracellularly surrounding the syncytiotrophoblastic cells and very rarely expressed in the maternal decidua layer. These results suggest that hOLF44 may have matrix-related function involved in human placental and embryonic development, or play a similar role in other physiological processes. The further functional characterization of hOLF44 may provide insights into a better understanding of the newly identified OLF subfamily.**


*Keywords:* Secreted protein; Olfactomedin; Signal peptide; hOLF44; Placenta

## 1. Introduction

Secreted proteins play an essential role in the development, differentiation and maintenance of multicellular organisms. Isolation and characterization of human function-unknown secreted proteins may not only provide insights into the understanding of the crucial biological processes that these molecules govern, but also offer new opportunities for the development of drug therapies. In particular, secreted proteins possess properties that lend themselves to be utilized as therapeutic agents or targets [1]. Thus, from a pharmaceutical perspective, secreted proteins indicate a great potential in the treatment of major diseases, and thus systematic screening and cloning of genes encoding these proteins are of particular interest. Several methods were therefore developed to screen secreted proteins. Tashiro et al. [2] invented one effective method, signal sequence trap (SST), by exploiting the ability of the signal peptide to lead proteins outside of the plasma membrane. However, some drawbacks, including considerable time-consuming, high cost and high redundancy of known genes, are still wrapped with the SST method. With the availability of the avalanche of DNA and protein sequence information, a computer based prediction method, SignalP [3], has emerged, and this method is, in principle, more suitable for us to trap novel or hypothetical secreted proteins compared with the SST method.

In order to isolate novel or hypothetical genes encoding secreted proteins involved in intercellular signaling, we utilized SignalP [3] to predict the putative secreted proteins from in-house EST (expressed sequence tag) database generated by our transcriptome projects [4,5]. Many in silico protein sequences possessing a putative signal peptide were obtained through bioinformatics analysis. Interestingly, five of them were found to contain a large olfactomedin-like (OLF) domain, and their amino acid sequences are 46–99% similar to that of an OLF protein playing a crucial role in neural development [6,7]. Additionally, some of the characterized OLF proteins are also involved in various developmental processes. This finding motivated us to categorize the five OLF sequences we trapped, but simple highest-hit method gave incomprehensive results because of the complexity of the OLF family. Our further phylogenetic analysis on more OLF proteins clustered four of the five OLF sequences into a few OLF subfamilies containing characterized members, whereas the other one, named hOLF44, fell into a newly identified OLF subfamily, in which all the members have not been characterized yet. The distinction of

☆ Nucleotide sequence and protein sequence of *hOLF44* have been deposited to GenBank database with Accession Nos. AY464015 and AAR88262, respectively.

* Corresponding author. Fax: +86-21-63852655.
*E-mail address:* wjma@sibs.ac.cn (W.-J. Ma).

*Abbreviations:* OLF, olfactomedin-like; SST, signal sequence trap; EST, expressed sequence tag; ECM, extracellular matrix

0014-5793/$22.00 © 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2004.06.059

hOLF44 promoted us to make a further study of it. In this paper, we report the identification, isolation and preliminary characterization of a secreted glycoprotein, hOLF44, which belongs to an uncharacterized OLF subfamily newly identified by phylogenetic analysis.

## 2. Materials and methods

### 2.1. Trapping for human novel or hypothetical secreted proteins

SignalP, a computer on-line program designed for searching signal sequences and their cleavage sites [3], was employed to predict the putative secreted proteins. We defined a positive SignalP hit as the simultaneous presence of three criteria: (a) signal peptide predicted by SignalP-NN; (b) signal peptide predicted by SignalP-HMM; and (c) signal peptide cleavage site located within 10–50 amino acids from the N-terminal.

### 2.2. Sequence alignment and phylogenetic analysis

Almost all currently available OLF protein sequences from the major species extensively sequenced were retrieved from Smart database [8]. After the redundancy (alternately spliced transcripts of the same gene) was eliminated, the full-length OLF protein sequences were aligned using ClustalW program [9] with the default settings. Phylogenetic trees were constructed from the resulting multiple sequence alignments using minimum-evolution [10] and neighbor-joining [11] methods, as implemented in the MEGA2 program [12]. The Poisson correction model was used for distance matrix calculations and sites containing missing data or alignment gaps were removed in a complete-deletion fashion. Reliability of the phylogenetic trees was examined by bootstrap [13] test with 1000 replications using the same computer package. The two methods gave virtually identical results for our datasets. Only the minimum-evolution tree is shown in this paper, as the minimum-evolution method is known to be more efficient than most other distance methods for phylogenetic inference [10]. Since the protein repertoire between mouse and rat is almost identical, only the mouse OLF proteins were presented in the phylogenetic tree in this paper.

### 2.3. Cloning of hOLF44 gene

The in silico *hOLF44* gene was obtained from in-house EST database and relevant public EST database using in silico cloning method. The coding region of *hOLF44* was amplified from total RNA of human adult liver by RT-PCR using the following two primers based on the in silico *hOLF44* gene: 5′-GGAATTCACTCTCCAGGCTGCCATG-3′ and 5′-GGGGTACCAACCTCCTCCTCTTTCTTCCTC-3′. The PCR product, designated as hOLF44-ORF, was then cloned into pcDNA3.1-A (Invitrogen) with a c-Myc tag in the C-terminal and verified by DNA sequencing. The resulting construct was designated as pcDNA3.1-hOLF44-Myc.

### 2.4. In vitro transcription and translation of hOLF44

The construct pcDNA3.1-hOLF44-Myc was transcribed and translated in the presence of [$^{35}$S]methionine using TNT coupled transcription/translation system (Promega). The synthesized proteins were separated by 12% SDS–polyacrylamide gel electrophoresis (PAGE) and signal was visualized by autoradiography.

### 2.5. Northern blot analysis

Northern blot analysis was performed on Human 12-lane Multiple Tissue Northern Blot (Clontech). hOLF44-ORF was used as a template to generate the probe which was labeled with [$\alpha$-$^{32}$P]dCTP. Prehybridization, hybridization and washing were performed according to Clontech's protocol. Hybridization signals were visualized by autoradiography. The same blot was stripped and re-hybridized with a 2 kb human *β-actin* probe to verify that all lanes contain comparable amounts of mRNA.

### 2.6. Cell transfection and Western blot analysis

Cell transfection experiment was performed with Lipofectamine according to the manufacturer's protocol (Invitrogen). Both culture medium and cell lysate of the transfected COS-7 cells were enriched using TALON metal affinity resins (Clontech). The enriched samples were separated on 12% SDS–PAGE and transferred onto PVDF membrane (Amersham Biosciences). The membrane was blocked and then incubated with c-Myc monoclonal antibody (Clontech), followed by incubation with goat horseradish peroxidase-conjugated anti-mouse antibody (Invitrogen). The immune complexes were visualized by chemiluminescence kit (ECL, Amersham Biosciences). For deglycosylation experiment, the enriched hOLF44 protein sample from culture medium was subject to deglycosylation with PNGase F according to the manufacturer's protocol (New England BioLabs).

### 2.7. Immunofluorescence microscopy

COS-7 cells growing on glass coverslips were transfected with the construct pcDNA3.1-hOLF44-Myc. After incubation at 37 °C for 60 h, the transfected cells were fixed in 2% paraformaldehyde (PFA) at 4 °C for 30 min. Then, the fixed cells were blocked with 5% BSA and incubated with c-Myc monoclonal antibody (Clontech) at 4 °C overnight, followed by goat fluorescein isothiocyanate (FITC)-conjugated anti-mouse IgG (Invitrogen). Coverslips were mounted and then examined by immunofluorescence light microscopy.

### 2.8. Antibody generation and immunohistochemistry

The cDNA fragment encoding the N-terminal (residues 22–271) of hOLF44 was subcloned into pGEX-5x-1 vector resulting in glutathione *S*-transferase (GST)-hOLF44 construct. GST-hOLF44 fusion proteins were produced and purified using Glutathione Sepharose 4B (Amersham Biosciences) according to the manufacturer's protocol. Rabbits were immunized with standard procedures using 2 mg of GST-hOLF44 fusion proteins as antigen for injection. The antisera were purified using Protein G Sepharose 4 Fast Flow (Amersham Biosciences) and then the anti-GST antibody was removed according to the manufacturer's instructions. The resulting purified polyclonal antibody was designated as anti-hOLF44. Rabbit polyclonal pre-immune serum was obtained prior to immunization and parallelly processed. The resulting purified pre-immune antibody was used as a negative control.

Human term placenta was midsagittally excised and routinely fixed in 4% PFA. After dehydration with ethanol and xylene, the fixed samples were embedded in paraffin wax using standard procedures. Paraffin blocks containing samples were then sagittally sectioned with a microtome at 5 μm and mounted on subbed slides. After dewaxing and rehydration, sections were incubated in freshly made 3% $H_2O_2$/methanol to remove any endogenous peroxidase activity. After digestion with 4% pepsin, the sections were blocked and then incubated with anti-hOLF44 polyclonal antibody at 4 °C overnight, followed by incubation with goat horseradish peroxidase-conjugated anti-rabbit IgG (Dako Japan Ltd.). The signals were detected with Diaminobenzidine (DAB) substrate kit (Vector Laboratories, Inc.). Sections were then counterstained with methyl green to identify nuclei.

## 3. Results

### 3.1. hOLF44 falls into an uncharacterized OLF subfamily newly identified by phylogenetic analysis

In order to categorize the five OLF sequences we trapped (clone number corresponding to in-house ESTs: AV702260, AV722944, AV722100, AV723866 and AV713200, respectively), we first used the common highest-hit homology method to analyze them, but it gave incomprehensive results because of the complexity of the OLF family. Thus, we performed phylogenetic analysis on available full-length OLF proteins from the major species extensively sequenced. The resulting phylogenetic tree showed that four of the five OLF sequences fell into a few OLF subfamilies containing well-characterized members. For example, AV702260 and AV722944 were clustered into the latrophilin subfamily; AV722100 and AV723866 fell into the noelin subfamily (Fig. 1). However, AV713200 was not clustered into any of the OLF subfamilies containing characterized members, and fell into a newly identified OLF subfamily, in which all the members have not been

characterized at all (Fig. 1). Thus, AV713200 may represent a distinct OLF member and characterization of it becomes more attractive, and we named it hOLF44 after its predicted molecular weight of 44 kDa. Nucleotide sequence and protein sequence of *hOLF44* have been deposited to GenBank database with Accession Nos. AY464015 and AAR88262, respectively.

### 3.2. Isolation, cloning and structural analysis of hOLF44

The coding region of *hOLF44* was cloned by RT-PCR method and the sequencing result conformed to the in silico *hOLF44* sequence and relevant sequences deposited in public database. The full-length cDNA sequence of *hOLF44* consists of approximately 1.8 kb and the longest open reading frame encodes a protein with a predicted molecular weight of 44 kDa. The start codon is surrounded by a homologous Kozak consensus sequence (GCCATGG) [15]. *hOLF44* gene locus was mapped within chromosome 1p13.1 by searching the human genome database in NCBI, and the genomic structure is composed of three exons and two introns that cover a minimum of 2.9 kb. All these exon-intron splice junctions conform to the GT/AG rule [16].

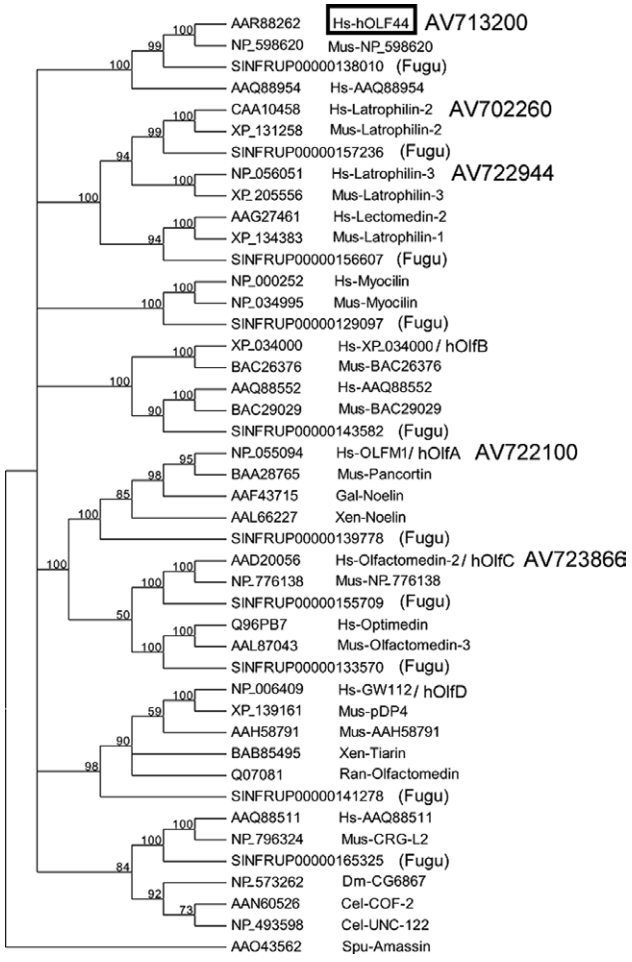


Fig. 1. hOLF44 belongs to a newly identified uncharacterized OLF subfamily. The phylogenetic tree was constructed from the multiple alignments of the full-length OLF protein sequences using minimum-evolution [10] method. Numbers above the branches represent the bootstrap values [13] based on 1000 replicates. The proteins are named using an abbreviated species name followed by a gene label. Accession numbers are listed at the left of their corresponding proteins. hOLF44 is boxed. Previously reported human OLF sequences, hOlfA-hOlfD [14], are depicted at the right of their corresponding protein. The five OLF proteins we trapped are labeled with the corresponding clone numbers of in-house ESTs (AV713200, AV702260, AV722944, AV722100 and AV723866, respectively). Hs, *Homo sapiens*; Mus, *Mus musculus*; Gal, *Gallus gallus*; Xen, *Xenopus laevis*; Ran, *Rana catesbeiana*, Dm, *Drosophila melanogaster*; Cel, *Caenorhabditis elegans*; Spu, *Strongylocentrotus purpuratus*. For more details on tree generation, see Section 2.

Bioinformatic analysis of the predicted protein of *hOLF44* revealed that it has a signal peptide in the N-terminal (residues 1–21), two potential N-glycosylation sites (residues 177 and 248) in the middle region, and a large OLF domain in the C-terminal (residues 135–401). To better characterize the structural property of hOLF44, we performed sequence and secondary structure comparisons between hOLF44 and other well-characterized OLF members, including olfactomedin, myocilin, noelin, tiarin, amassin and pDP4. Secondary structure predictions were performed using sequence alignments as input for Jpred [17]. The result showed that the C-terminal (OLF domain) of hOLF44 mainly contains β-strands motif (though two α-helixes are predicated), whereas the N-terminal of it consists of α-helixes (Fig. 2A). Although the sequence similarity between hOLF44 and the other well-characterized OLF members is not very high (approximately 45%), they obviously share similar structural motifs and the key residues that stabilize the OLF domain structure are sufficiently conserved during evolution (Fig. 2A). Furthermore, their domain architecture was found to be strikingly similar in the presence, size and orders (Fig. 2B).

### 3.3. hOLF44 is a secreted glycoprotein

To determine whether hOLF44 is secreted, the construct pcDNA3.1-hOLF44-Myc was transiently transfected into COS-7 cells, both the culture medium and the cell lysate were analyzed by Western blot using c-Myc antibody. For negative and positive controls, vector pcDNA3.1-A and the construct pcDNA3.1-hGH-Myc (hGH, human growth hormone) were parallelly processed. A specific protein band of approximately 60 kDa was detected in both culture medium and cell lysate (Fig. 3A, lanes 3 and 6). The size of the detected band was showed to be much larger than that of the predicted protein molecular weight of hOLF44. Since hOLF44 contains two potential N-glycosylation sites (residues 177 and 248), we next tested whether this phenomenon is due to glycosylation by treatment of the enriched hOLF44 protein sample with a deglycosylating enzyme PNGase F. After treatment with PNGase F, the size of the detected band was approximately 49 kDa, which was the predicted size of the Myc-tagged hOLF44 protein (Fig. 3B, lane 2). It is equivalent to that of in vitro translated protein from pcDNA3.1-hOLF44-Myc (Fig. 3B, lane 3). Taken together, these data strongly demonstrate that hOLF44 is a secreted glycoprotein. Additionally, the experimental results provided proof for the aforementioned bioinformatics predictions.

### 3.4. hOLF44 mRNA was abundantly expressed in placenta

To find clues to the functional involvement in vivo, the mRNA expression pattern of *hOLF44* was investigated by Northern blot analysis, which was carried out on Human 12-lane Multiple Tissue Northern Blot from Clontech. The result showed that to some extent *hOLF44* mRNA was differentially expressed in multiple tissues (Fig. 4). Approximately 1.8 kb

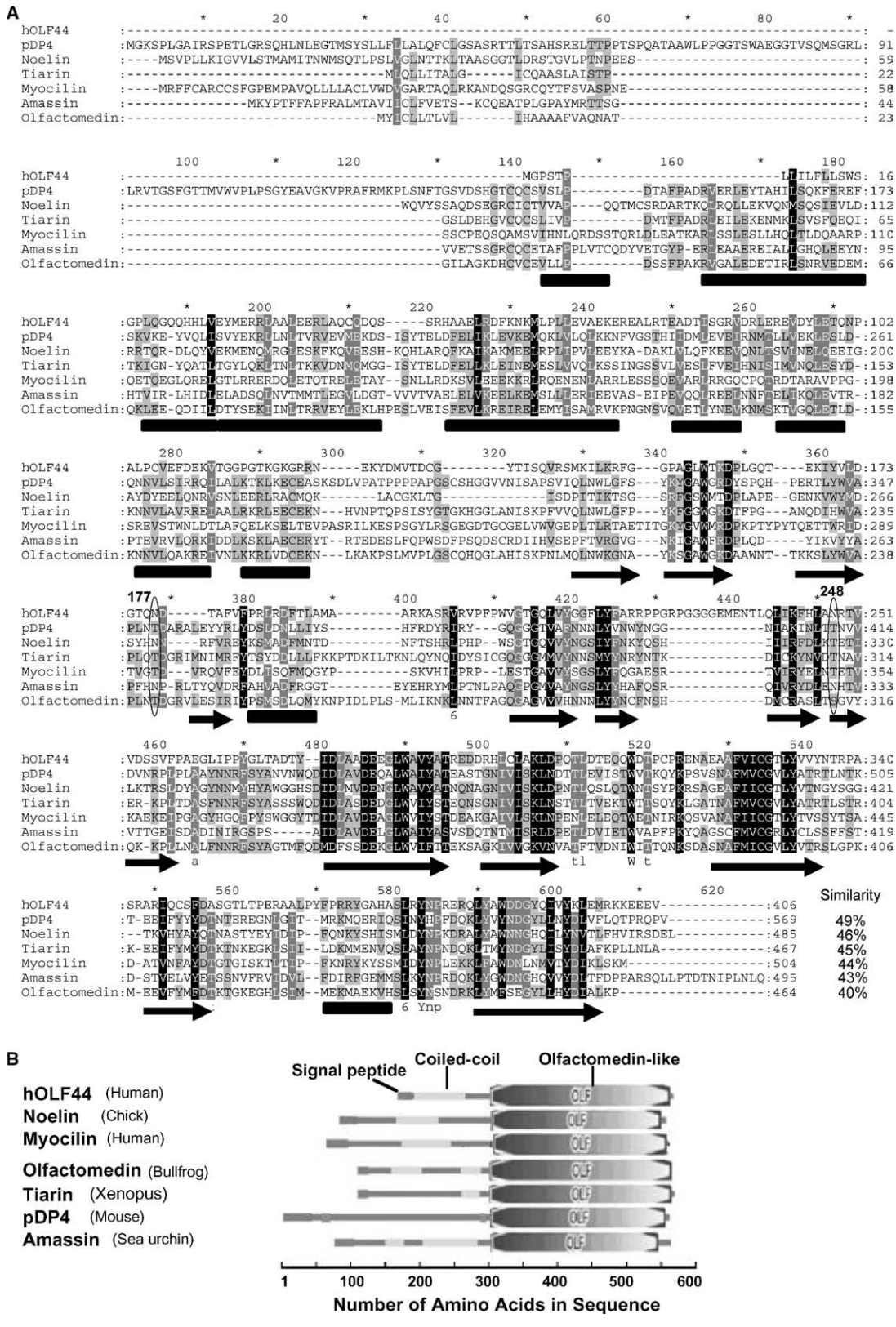


Fig. 2. Sequence and domain architecture comparisons between hOLF44 and other well-characterized OLF members. (A) Full-length sequence comparison between hOLF44 and other well-characterized OLF members. Identical amino acid residues are dark-shaded and similar residues are shaded at various degrees. Ellipses indicate the conserved glycosylation sites (residues 177 and 248). The percentage indicates the sequence similarities relative to hOLF44. The predicted secondary structural motifs are depicted with arrows for β-strands and cylinders for α-helixes. (B) Domain architecture comparison among these seven full-length OLF members. From N-terminal to C-terminal, there is a signal peptide sequence, a long coiled-coil motif and the OLF domain.

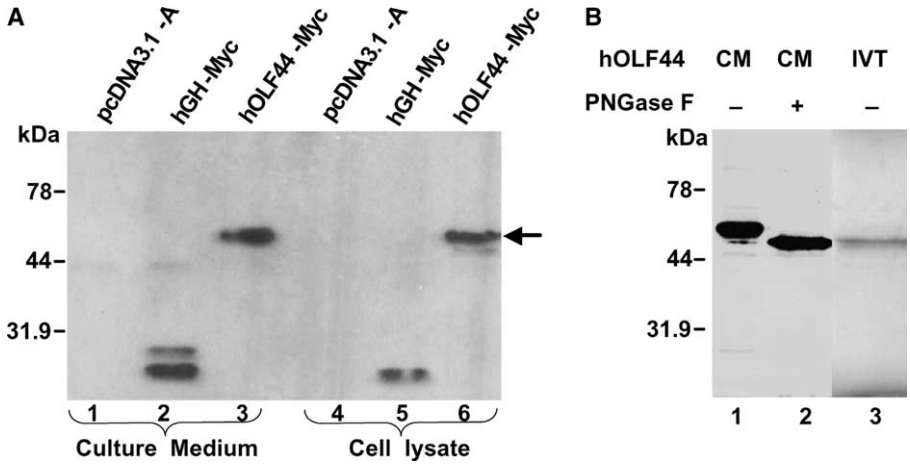


Fig. 3. Secretion and deglycosylation analyses of hOLF44. (A) Myc-tagged hOLF44 protein is robustly secreted from the transfected COS-7 cells (lane 3). pcDNA3.1-A and pcDNA3.1-hGH-Myc are the negative and positive controls, respectively. Lanes 1, 2 and 3 are the enriched samples from culture medium; lanes 4, 5 and 6 are the enriched samples from the corresponding cell lysate. An arrow at the right indicates the position of the expressed Myc-tagged hOLF44 protein. Proteins were detected with c-Myc antibody. (B) Deglycosylation of the enriched hOLF44 protein sample from culture medium (CM) of transfected COS-7 cells. After treatment with PNGase F (lane 2), the size of detected Myc-tagged hOLF44 protein band is reduced and is very close to that of the in vitro translated protein (IVT) from pcDNA3.1-hOLF44-Myc (lane 3). Proteins were detected with c-Myc antibody. CM, culture medium; IVT, in vitro translated protein.

transcript of *hOLF44* was detected, predominantly in placenta. Moderate expression was observed in liver and heart, and fairly weak signals were also detected in skeletal muscle, small intestine, lung and kidney. Very weak signals were observed in colon, thymus, spleen and brain, whereas in peripheral blood leukocyte, no signal could be detected at this level of sensitivity. Additionally, fairly weak signals of smaller bands were detected in heart, skeletal muscle and placenta, and they may represent alternative splicing variants.

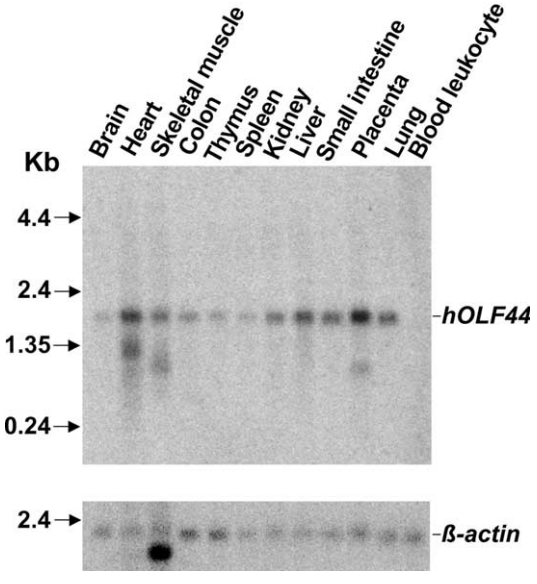


Fig. 4. *hOLF44* mRNA is differentially expressed in various tissues and is particularly abundant in placenta. Human 12-Lane Multiple Tissue Northern Blot from Clontech was hybridized with $\alpha$-$^{32}$P-labeled probe for *hOLF44* and the 2 kb human *β-actin* cDNA as control for total mRNA quantification. Size standards and the order of tissues source are marked at the left and top of the blot, respectively.

### 3.5. Cellular and subcellular localization of hOLF44 protein

To clarify the localization of hOLF44, we utilized two methods to resolve this issue, including immunofluorescence for Myc-tagged hOLF44 protein and immunohistochemical staining for the endogenous hOLF44 protein. COS-7 cells were transfected with the construct pcDNA3.1-hOLF44-Myc, and the immunofluorescence assay with c-Myc antibody showed that Myc-tagged hOLF44 protein was preferentially detected in the perinuclear regions (Fig. 5A), most likely in the endoplasmic reticulum (ER), which is the common feature of secreted proteins. Furthermore, we also investigated the localization of endogenous hOLF44 protein in human placenta where its transcripts exhibit the strongest expression according to our Northern blot analysis. Immunohistochemical staining with rabbit anti-hOLF44 primary antibody on human term placenta sections (sagittally sectioned at 5 μm) showed that strong signals were detected extracellularly surrounding the syncytiotrophoblastic cells (ST) on the fetal side (Fig. 5C), whereas very rare signal was detected in the maternal decidua (MD) layer. With purified pre-immune antibody for the negative control, no such signal was detected (Fig. 5B). Additionally, our immunohistochemical studies on cryosections of human adult normal and cancerous livers also revealed that hOLF44 was localized extracellularly surrounding hepatocytes (data not shown).

## 4. Discussion

In this study, we isolated and preliminarily characterized human hOLF44, a secreted glycoprotein differing from other previously reported human OLF proteins. The members in the OLF family all contain a highly conserved OLF domain at the C-terminal, whereas the N-terminal of them is more variable [18]. Another common feature of the most characterized OLF members is that their mRNA expression patterns are highly selective. For example, myocilin/TIGR, a secreted protein that

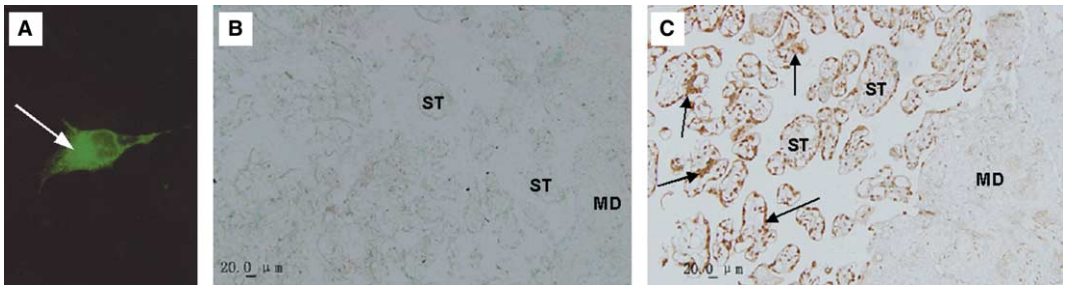


Fig. 5. Immunofluorescence and immunohistochemical detection of hOLF44 protein. (A) Subcellular localization of Myc-tagged hOLF44 protein. COS-7 cells were transfected with pcDNA3.1-hOLF44-Myc. Immunostaining was carried out using c-Myc antibody as the primary antibody and FITC-conjugated goat anti-mouse IgG as the secondary antibody. Arrow indicates the Myc-tagged hOLF44 protein stained with green immunofluorescence in the perinuclear regions. (B and C) Immunohistochemical detection of endogenous hOLF44 protein in human term placenta sections, sagittal view. (B) Purified pre-immune antibody incubation is presented as negative control. (C) Section was incubated with rabbit anti-hOLF44 polyclonal antibody as the primary antibody and arrows indicate the representative positive staining. All sections were counterstained with methyl green. Magnifications: A, 400×; B and C, 200×.

has been associated with human primary open angle glaucoma [19], is selectively expressed in heart, skeletal muscle, brain and retina [14,20]. The previously reported human OLF sequences, *hOlfA–hOlfD/hGC-1/GW112* and *optimedin*, all show distinct expression patterns with *hOlfD/hGC-1/GW112* in colon, small intestine and prostate [14,21,22] and *hOlfB* in pancreas and prostate [14]. Since *hOlfA*, *hOlfC* and *optimedin* all fall into a relevant subfamily (Fig. 1), they share similar expression patterns in brain [14,20]. However, *hOLF44* is differentially expressed in multiple tissues with particularly abundant expression in placenta and moderate expression in liver and heart. Additional weak signals of smaller bands were detected in heart, skeletal muscle and placenta. This may imply that there exists alternative splicing variants in different tissues. However, this phenomenon needs further experimental confirmation. So in this case, *hOLF44* differs greatly from any of the aforementioned human OLF members belonging to other OLF subfamilies.

Although the biological functions of most OLF members were not well elucidated, some characterized members were found to be involved in the formation of the extracellular matrix (ECM). Olfactomedin is an ECM protein surrounding the bullfrog olfactory neurons [23,24]. Two recent reports on invertebrate OLF members found that amassin functions as a cell adhesion protein in sea urchin coelomocytes [25] and UNC-122 may act as a structural component of the neuromuscular junctions in *Caenorhabditis elegans* [26]. Accordingly, hOLF44 is localized extracellularly surrounding the syncytiotrophoblastic cells, which are the major structural components of human placental villous chorion and play a crucial role in the development and maintenance of placenta [27,28]. Additional immunohistochemical staining on human liver sections shows that hOLF44 is localized extracellularly surrounding hepatocytes. These observations imply that hOLF44 may form ECM structure by interacting with other components. Additionally, hOLF44 is heavily glycosylated and this glycosylation may play a role in the folding, stability, trafficking and/or function of this protein.

Furthermore, several OLF members, including noelin [6,7], tiarin [29] and pDP4 [30], are all implicated in developmental processes. Since hOLF44 contains a highly conserved OLF domain and shares similar structural properties with them (Fig. 2), it may also play a similar role in particular developmental processes. Based on its particular mRNA expression pattern and ability to form ECM structure in human placenta and liver, we suggest that hOLF44 may act as a matrix-related component involved in placental and embryonic development or play a similar role in liver and other tissues.

In conclusion, hOLF44 is a secreted glycoprotein and falls into a newly identified uncharacterized OLF subfamily. Our findings suggest that hOLF44 may serve as a matrix-related effector in human placental and embryonic development or play a similar role in other physiological processes. Our experimental results also demonstrate that bioinformatics tools, to some extent, are reliable for guiding functional studies of uncharacterized genes. Further studies in mouse, where genetic ablation is possible would promise to uncover important clues related to the biological function(s) of hOLF44.

*Acknowledgements:* We are very grateful to Ying-wei Chen, Wen-wei Yu and Jun-yu Yao for their kind help in the preparation of tissue sections. We also thank Xian-Ru Xu, Jian Qu and Zhi-Hong Cheng for their technical assistance. This work was supported by State 863 High Technology R&D Project of China and grant from Shanghai Commission of Science and Technology.

## References

[1] Walsh, G. (2003) Eur. J. Pharm. Biopharm. 55, 3–10.
[2] Tashiro, K., Tada, H., Heiker, R., Shirozu, M., Nakano, T. and Honjo, T. (1993) Science 261, 600–603.
[3] Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Protein Eng. 10, 1–6.
[4] Hu, R.M., Han, Z.G., Song, H.D., Peng, Y.D., Huang, Q.H., Ren, S.X., Gu, Y.J., Huang, C.H., Li, Y.B., Jiang, C.L., Fu, G., Zhang, Q.H., Gu, B.W., Dai, M., Mao, Y.F., Gao, G.F., Rong, R., Ye, M., Zhou, J., Xu, S.H., Gu, J., Shi, J.X., Jin, W.R., Zhang, C.K., Wu, T.M., Huang, G.Y., Chen, Z., Chen, M.D. and Chen, J.L. (2000) Proc. Natl. Acad. Sci. USA 97, 9543–9548.
[5] Xu, X.R., Huang, J., Xu, Z.G., Qian, B.Z., Zhu, Z.D., Yan, Q., Cai, T., Zhang, X., Xiao, H.S., Qu, J., Liu, F., Huang, Q.H., Cheng, Z.H., Li, N.G., Du, J.J., Hu, W., Shen, K.T., Lu, G., Fu, G., Zhong, M., Xu, S.H., Gu, W.Y., Huang, W., Zhao, X.T., Hu, G.X., Gu, J.R., Chen, Z. and Han, Z.G. (2001) Proc. Natl. Acad. Sci. USA 98, 15089–15094.
[6] Barembaum, M., Moreno, T.A., LaBonne, C., Sechrist, J. and Bronner-Fraser, M. (2000) Nat. Cell Biol. 2, 219–225.
[7] Moreno, T.A. and Bronner-Fraser, M. (2001) Dev. Biol. 240, 340–360.
[8] Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R., Ponting, C. and Bork, P. (2002) Nucl. Acids Res. 30, 242–244.

[9] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucl. Acids Res. 22, 4673–4680.
[10] Rzhetsky, A. and Nei, M. (1993) Mol. Biol. Evol. 10, 1073–1095.
[11] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[12] Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001) Bioinformatics 17, 1244–1245.
[13] Felsenstein, J. (1985) Evolution 39, 783–791.
[14] Kulkarni, N.H., Karavanich, C.A., Atchley, W.R. and Anholt, R.R. (2000) Genet. Res. 76, 41–50.
[15] Kozak, M. (1984) Nucl. Acids Res. 12, 857–872.
[16] Shapiro, M.B. and Senapathy, P. (1987) Nucl. Acids Res. 15, 7155–7174.
[17] Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998) Bioinformatics 14, 892–893.
[18] Karavanich, C.A. and Anholt, R.R. (1998) Mol. Biol. Evol. 15, 718–726.
[19] Stone, E.M., Fingert, J.H., Alward, W.L., Nguyen, T.D., Polansky, J.R., Sunden, S.L., Nishimura, D., Clark, A.F., Nystuen, A., Nichols, B.E., Mackey, D.A., Ritch, R., Kalenak, J.W., Craven, E.R. and Sheffield, V.C. (1997) Science 275, 668–670.
[20] Torrado, M., Trivedi, R., Zinovieva, R., Karavanova, I. and Tomarev, S.I. (2002) Hum. Mol. Genet. 11, 1291–1301.
[21] Shinozaki, S., Nakamura, T., Iimura, M., Kato, Y., Iizuka, B., Kobayashi, M. and Hayashi, N. (2001) Gut 48, 623–629.
[22] Zhang, J., Liu, W.L., Tang, D.C., Chen, L., Wang, M., Pack, S.D., Zhuang, Z. and Rodgers, G.P. (2002) Gene 283, 83–93.
[23] Snyder, D.A., Rivers, A.M., Yokoe, H., Menco, B.P. and Anholt, R.R. (1991) Biochemistry 30, 9143–9153.
[24] Yokoe, H. and Anholt, R.R. (1993) Proc. Natl. Acad. Sci. USA 90, 4655–4659.
[25] Hillier, B.J. and Vacquier, V.D. (2003) J. Cell Biol. 160, 597–604.
[26] Loria, P.M., Hodgkin, J. and Hobert, O. (2004) J. Neurosci. 24, 2101–2191.
[27] Enders, A.C. (1965) Obstet. Gynecol. 25, 378–386.
[28] Cross, J.C., Werb, Z. and Fisher, S.J. (1994) Science 266, 1508–1518.
[29] Tsuda, H., Sasai, N., Matsuo-Takasaki, M., Sakuragi, M., Murakami, Y. and Sasai, Y. (2002) Neuron 33, 515–528.
[30] Rosenbauer, F., Wagner, K., Zhang, P., Knobeloch, K.P., Iwama, A., and Tenen, D.G. (2004) Blood 2004 Feb 12 [Epub ahead of print].