VERSITA

CELLULAR & MOLECULAR BIOLOGY LETTERS
Volume 11, (2006) pp 161 – 170
http://www.cmbl.org.pl

DOI: 10.2478/s11658-006-0014-6

Received: 21 December 2005
Revised form accepted: 24 February 2006

© *2006 by the University of Wrocław, Poland*

A NOVEL RECENTLY EVOLVED GENE *C19orf24* ENCODES A NON-CLASSICAL SECRETED PROTEIN

XIN-RONG WANG[1,2], YU-BO ZHOU[2], FENG LIU[2], KE-SHENG WANG[2], YAN SHEN[2], JIAN-HUA LIU[1] and ZE-GUANG HAN[2]*
[1]School of Life Science and Technology, Shanghai Jiao Tong University, No.1954 Hua-Shan Rd., Shanghai 200030, China, [2]Chinese National Human Genome Center at Shanghai, No.351 Guo Shou-Jing Rd., Shanghai 201203, China

Abstract: Secreted proteins play important roles in many crucial biological processes, and can be new agents or targets for drug therapies. Here, we report on the isolation and characterization of a novel human non-classical secreted protein which is encoded by the hypothetical gene *C19orf24* (chromosome 19 open reading frame 24). It has no signal peptide, but can still secrete extracellularly despite the presence of the inhibitor brefeldin A (BFA), proving its non-classical secreted protein status. Via subcellular localization using *C19orf24 in vivo* and transfected pEYFP-Golgi plasmid in Hela cells, *C19orf24* was shown not to co-localize in the Golgi apparatus, which suggested that it secretes via a new and unknown pathway. Deglycosylation analysis with PNGase F verified that it has no N-glycosylation modification sites. Via the reverse transcription-PCR method, it was found to be expressed only in the human liver, and preferentially in normal tissue. In addition, *C19orf24* was shown to be a recently evolved gene, found only in *Homo sapiens* and *Pan troglodytes*. By calculating its synonymous and non-synonymous substitution rate (d_S/d_N), we found that it experienced a purifying selection, which suggests that *C19orf24* may have a special, irreplaceable biological function in the human organism.

Key words: *C19orf24*, Late evolution, Non-classical secreted protein, BFA

---

* Corresponding author: e-mail: hanzg@chgc.sh.cn; fax: (86-21) 50800402; tel: (86-21) 50800529-111

Abbreviations used: BFA – brefeldin A; *C19orf24* – chromosome 19 open reading frame 24; CL – cell lysate; CM – culture media; d_S/d_N – synonymous and non-synonymous substitution rate; GH – growth hormone; TRX – theoredoxin.

INTRODUCTION

Secreted proteins are of great importance, and are involved in many biological process, including the signalling pathway, morphogenesis, cell apoptosis, cell differentiation and cell defense [1]. Furthermore, some such proteins can become medicines, or targets in medical therapy, or be used as diagnostic markers [2-3]. As we know, most secreted proteins have a classic structure and behaviour, i.e. they have a signal peptide and secrete extracellularly through the ER/Golgi pathway, although there are some unconventional secreted proteins, for example fibroblast growth factor-1 (FGF-1), theoredoxin (TRX) and interleukin-1 β (IL-1β) [4-6]. These unconventional proteins have no signal peptide but can still secrete via other routes. Significantly, brefeldin A (BFA) cannot inhibit their secretion [4-8]. There is much interest in finding those proteins and exploring their biological functions.

Using Secretome P 1.0 and other software, we screened a novel leaderless secreted gene, *C19orf24*, which is located on human chromosome 19p13.3. We confirmed that its translated protein was a non-classical secreted protein via BFA secretion inhibition, subcellular localization, and deglycosylation analysis. In addition, this new gene's expression profile at the RNA level showed that it might play a role in liver function. *C19orf24* was only found in *Homo sapiens* and *Pan troglodytes*, and therefore may be a young gene which evolved recently and thus is not present in other species. Over 20 younger genes have thusfar been reported on, which prompted research on their origin and evolution [9-10]. The study of *C19orf24* could contribute to that pool of knowledge as it is potentially a young gene.

MATERIALS AND METHODS

Screening of the *C19orf24* gene

35446 protein sequences obtained from the public protein database were submitted to the SecretomeP 1.0 server (http://www.cbs.dtu.dk/services/SecretomeP) for the prediction of non-classical secreted proteins [11-12]. Non-classical secreted proteins should obtain a neural network score (NN-score) exceeding the normal threshold of 0.6, but at the same time be predicted not to contain a signal peptide. We selected 318 hypothetical protein sequences with NN-scores greater than or equal to 0.9 but without a signal peptide, and submitted them to the SOSUI software (http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html) to delete membrane proteins [13]. Then, 249 protein sequences were investigated by Proteome Analyst's subcellular localization server (http://pasub.cs.ualberta.ca:8080/pa/Subcellular) to get proteins with an extracellular probability of more than 30%. Finally, the reliability of each protein sequence was justified using the BLAST algorithm.

Sequence alignment and evolutionary ratio d_S/d_N computation

A *C19orf24*-coding sequence from *Homo sapiens* (AK000647) and its homologue from *Pan troglodytes* (XM_512233) were aligned using ClustalW software (http://www.ebi.ac.uk/clustalw) [15]. The alignment file was submitted to the Synonymous/Non-synonymous Analysis Program (http://www.hiv.lanl.gov/content/hiv-db/SNAP/ WEBSNAP/SNAP.html) for the evolutionary ratio d_S/d_N calculation [16].

Gene cloning and plasmid construction

The coding region of *C19orf24* was amplified from mixed cDNA from 9 human tissues by RT-PCR using the following two primers: forward 5'-GGGCTCGA GGCCATGAGGGAAGGACAG-3' and reverse 5'-GGGAATTCTTGTTCCCA GCCCGGC-3'. The PCR product *C19orf24*-ORF was then inserted into a pcDNA3.1-B vector (Invitrogen) at the *XhoI* and *EcoRI* sites with a c-Myc tag in the C-terminal, and verified by DNA sequencing. The resulting construct was designated as pcDNA3.1-*C19orf24*-myc. A similar method was used to create the *TRX*-myc clone using the following two primers: forward 5'-GGGGA TCCGCCAAGATGGTGAAGC-3' and reverse 5'-GCGGTACCGACTAATTC ATTAATGGTGG-3'. The PCR product *TRX*-ORF was then inserted into the pcDNA3.1-A vector (Invitrogen) at the *BamHI* and *KpnI* sites with a c-Myc tag in the C-terminal, and verified by DNA sequencing. The resulting construct was designated as pcDNA3.1-*TRX*-myc. The growth hormone (GH) plasmid pcDNA-3.1-*GH*-myc was from our own lab.

Cell transfection and Western blot analysis

COS-7 cells were subcultured twice before transfection. The cell transfection was performed with Lipofectamine according to the manufacturer's protocol (Invitrogen). After transfection for 24 h, the culture media was changed by DMEM with 10% FBS or DMEM with 10% FBS and 1 μg/ml BFA. The cell lysate was collected by 2×loading lysis buffer (50 mmol/L Tris-HCl, pH 6.8), 2% sodium clodecyl sulfate, 10% 2-mercaptoethanol, 10% glycerol and 0.002% bromphenol blue). The culture medium of the transfected COS-7 cells was enriched using 6×His-tag/Ni-NTA agarose resin (Qiagen). The treated samples were separated on 15% SDS-PAGE and transferred onto nitrocellulose membrane (Amersham Biosciences). The membrane was blocked and then incubated with c-Myc mouse monoclonal antibody (Invitrogen), followed by incubation with IRDye™800DX-conjugated affinity purified goat anti-mouse secondary antibody (Rockland). The signals on the membrane were detected using the Odyssey Infrared Imaging System (LI-COR Biosciences). For the deglycosylation, the protein sample from the cell lysate was subjected to deglycosylation with PNGase F according to the manufacturer's protocol (New England BioLabs).

Antibody generation
Full-length *C19orf24* was subcloned into a pGEX-4T-1 vector (Invitrogen) yielding a glutathione S-transferase (GST)-*C19orf24* construct. The GST-*C19orf24* fusion proteins were prepared and purified using Glutathione Sepharose 4B (Amersham Biosciences) according to the manufacturer's protocol. Rabbits were immunized via standard procedures using 2 mg of the GST-*C19orf24* fusion protein as the antigen for injection. The anti-sera were purified using Protein G Sepharose 4 Fast Flow (Amersham Biosciences), and then the anti-GST antibody was removed according to manufacturer's instructions. The resulting purified polyclonal antibody was called anti-C19orf24 Rabbit. Pre-immune serum was obtained prior to immunization and parallelly processed. The resulting purified pre-immune antibody was used as a negative control.

Immunofluorescence microscopy
Hela cells growing on glass coverslips were transfected with the construct pEYFP-Golgi plasmid. After incubation at 37ºC for 60 hours, the transfected cells were washed twice with PBS, then fixed in 4% paraformaldehyde (PFA) at 4ºC for 30 min, followed by permeation with 0.1% Triton X-100. After that, the fixed cells were blocked with 5% horse serum for 2 h at room temperature, and incubated with anti-C19orf24 rabbit polyclonal antibody (1:100) at 4ºC overnight, followed by incubation with fluorescence-conjugated anti-rabbit Cy3. Coverslips were mounted with Dapi (Sigma), and then examined with a LSM 5100 META laser confocal microscope (Zeiss).

RESULTS

***C19orf24* is a recently evolved gene, derived through a purifying selection**
C19orf24-coded protein has 132 aa, and its theoretical pI of 5.56 was computed with the Compute pI/Mw tool (http://au.expasy.org/tools/pi_tool.html) [17]. The NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/) predicted three phosphorylation sites: Ser^{72}, Thr^{10} and Thr^{79} (Fig. 1A) [18]. With BLAST EST or tBLASTn in a non-redundant database, *C19orf24* was only found in *Homo sapiens* and *Pan troglodytes*, and the identity between the two homologues was 97%. Three amino acids were different (Fig. 1C): the 48th aspartic acid to histidine, the 63rd tryptophan to arginine and the 77th glycine to arginine. Besides that, they were different in their gene structure; in *Pan troglodytes*, *C19orf24* lacked the middle untranslated region (Fig. 1B). The alignment file of two coding nucleic acid sequences was submitted to the Synonymous/Non-synonymous Analysis Program to calculate d_S/d_N, yielding 2.96, *i.e.* a d_N/d_S of 0.338. Therefore, *C19orf24* is not only a recently evolved gene but also experienced a purifying selection [19]. Although the origin of this new gene is as yet unknown due to the lack of primate resources, it must have an irreplaceable function.

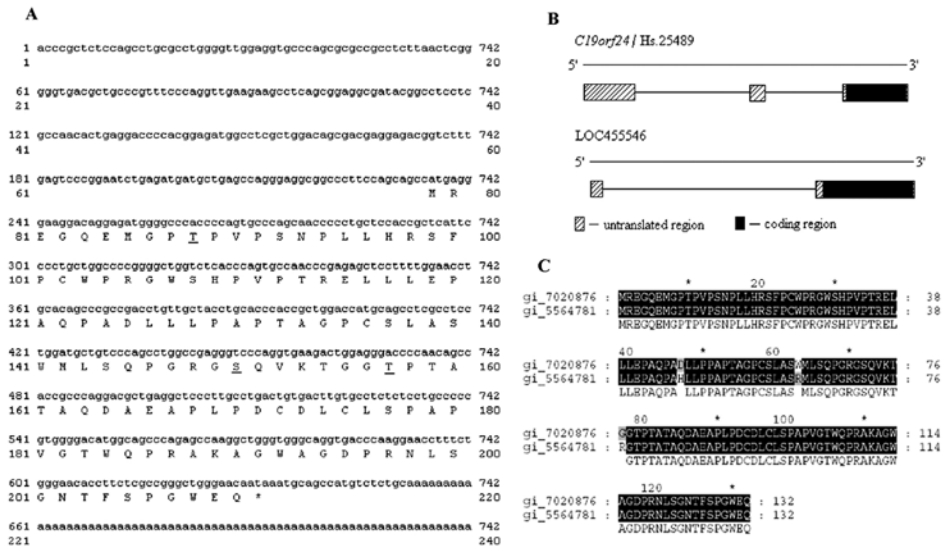

Fig. 1. Bioinformatics analysis of *C19orf24*. A - *C19orf24* cDNA sequence and the translated amino acid sequence. Putative phosphorylation sites were underlined. B - A sketch map of the gene structure of *C19orf24* from *Homo sapiens* (upper: Hs.25489) and *Pan troglodytes* (lower: LOC455546). C - The alignment of the two protein sequences: *Homo sapiens* (gi7020876), *Pan troglodytes* (gi55647817). The identity was 97%.

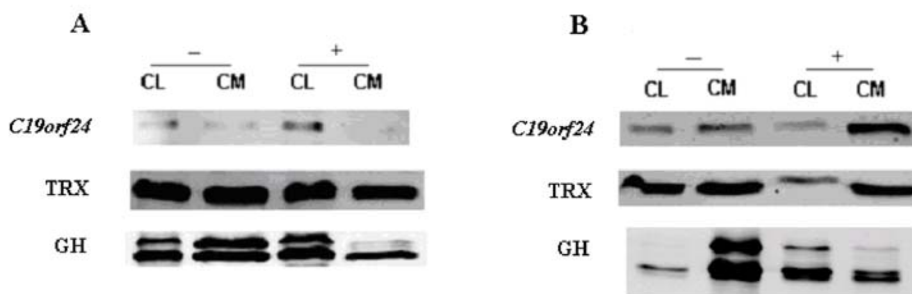

Fig. 2. The *C19orf24*-coded protein was secreted from COS-7 cells via a brefeldin A-insensitive pathway. A - COS-7 cells were transfected with 3 μg of plasmid DNA and 5 μl lipofactamine for each 35 mm dish. The culture medium was changed by DMEM/10%FBS (-) or DMEM/10%FBS with 1 μg/ml brefeldin A (+) after 24 h. The cell lysate (CL) and cell media (CM) were collected at 48 h. CM was purified with 6×His-tag/Ni-NTA, and CL was harvested with 2×loading lysis buffer, quantified by the BCA method, and subjected to Western blot analysis. TRX was the negative control and GH was the positive control. B - Methods of transfection and protein sample treatment were as in (A) except that CL and CM were collected at 72 h.

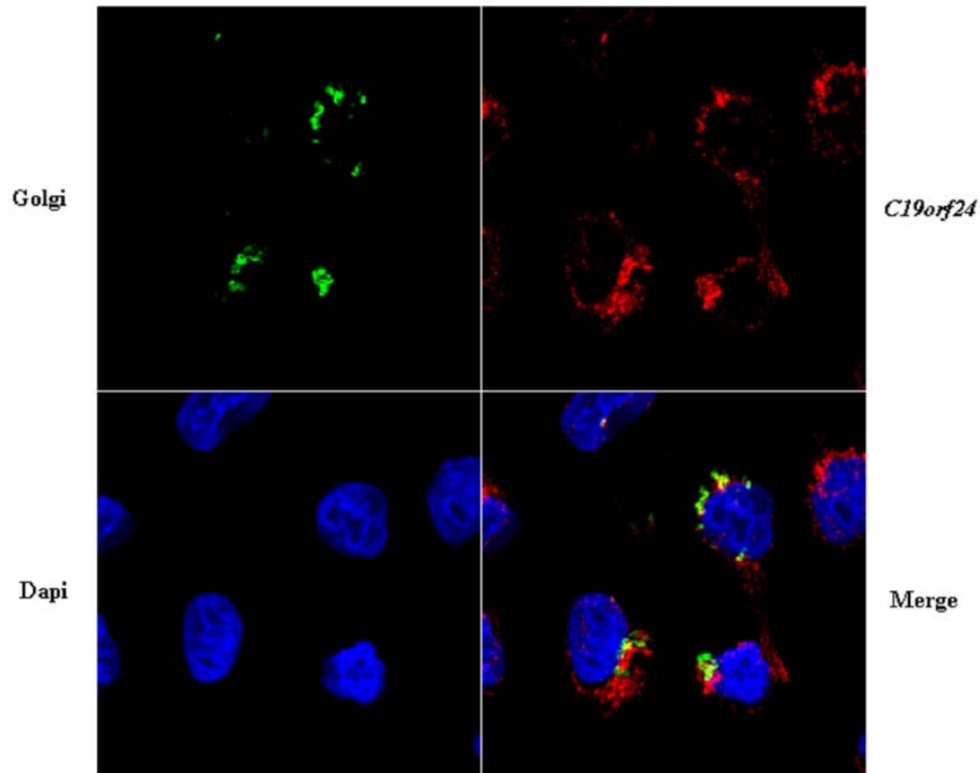

Fig. 3. *C19orf24* did not co-localize in the Golgi apparatus. Hela cells were transfected with 1 μg of an expression construct for pEYFP-Golgi and 2.5 μl lipofactamine. The intracellular distribution of the Golgi apparatus was detected by the intrinsic green fluorescence of the YFP tag. *C19orf24* localization was detected with an anti-*C19orf24* rabbit antibody and a Cy3-conjugated secondary antibody.

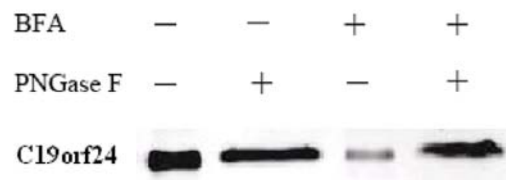

Fig. 4. The *C19orf24*-coded protein had no N-glycosylation modification sites. After transfection in COS-7 cells, the protein sample from the cell lysate was subjected to deglycosylation with PNGase F according to the manufacturer's protocol, and to Western blot detection. Transfected cell lysate without BFA and PNGase F treatment was the positive control.

C19orf24 encoded a non-classical secreted protein

C19orf24- coded protein was predicted to be a leaderless secreted protein since it satisfied the three conditions: NN>0.6, no signal peptide and 33.3% extracellular probability. To verify that prediction, we used BFA secretion inhibition. BFA is a fungal macrocyclic lactone, frequently used to inhibit both protein secretion through the ER/Golgi pathway and the maturation of N-linked core oligosaccharides [20-21]. pcDNA3.1-*C19orf24*-myc, pcDNA3.1-*GH*-myc and pcDNA3.1-*TRX*-myc plasmids were transfected into COS-7 to evaluate the effect of BFA. GH was the positive control and TRX was the negative control. 48 h after transfection, GH secretion was inhibited by nearly 50%, TRX secretion was unblocked and almost at its previous levels, while *C19orf24* had just began to secrete (Fig. 2A). Seventy two hours after transfection, *C19orf24* secretion had increased considerably, showing no inhibitory effect of BFA, while GH was obviously inhibited and TRX remained unaffected (Fig. 2B). Therefore, the *C19orf24*-coded protein was confirmed to be an unconventional secreted protein.

C19orf24 did not co-localize in the Golgi apparatus

In general, classical secreted proteins secrete through the ER/Golgi pathway, which means they can co-localize within the ER/Golgi. Non-classical proteins do not secrete along the ER/Golgi pathway, so they should not co-localize within the ER/Golgi. This idea was consistent with the results obtained. *In vivo*, C19orf24 localized in the cytoplasm, not in the Golgi apparatus (Fig. 3).

The _C19orf24_- coded protein had no N-glycosylation modification sites

The actual molecular weight of the *C19orf24*- coded protein was a little higher than the predicted molecular weight of 14kDa. Therefore, we guessed it may have been modified in some way. The most possible modification was glycosylation, since this process is favourable for the transport of the generally classical secreted proteins; however, what about ER/Golgi-independent secreted proteins? To answer this question, the *C19orf24*- coded protein sequence was first submitted to the NetNGlyc 1.0 server for N- glycosylation prediction (http://www.cbs.dtu.dk/services/NetNGlyc). The result indicated that the sequence might not contain a signal peptide. Proteins without signal peptides are unlikely to be exposed to the N-glycosylation machinery and thus may not be glycosylated *in vivo* even though they contain potential motifs. Secondly, a deglycosylation analysis with PNGase F indicated that the *C19orf24*-coded protein had no N-glycosylation modification sites (Fig. 4), regardless of whether the cell lysate was BFA treated, which further supported the thesis that the *C19orf24*- coded protein was a non-classical secreted protein. However, it may have undergone other modification.

C19orf24 was specifically expressed in human liver tissue and up-regulated in normal liver tissue

When the *C19orf24* gene expression profile was examined by RT-PCR for 35 cycles, it was surprisingly only found expressed in human liver tissue (Fig. 5A). We were eager to know how it was expressed in normal liver and cancerous liver tissue, so 10 pairs of tissue cDNA samples were randomly chosen from patients with liver cancer from our liver sample repository to make an evaluation. Primarily, *C19orf24* gene expression was up-regulated in normal liver tissue (Fig. 5B).

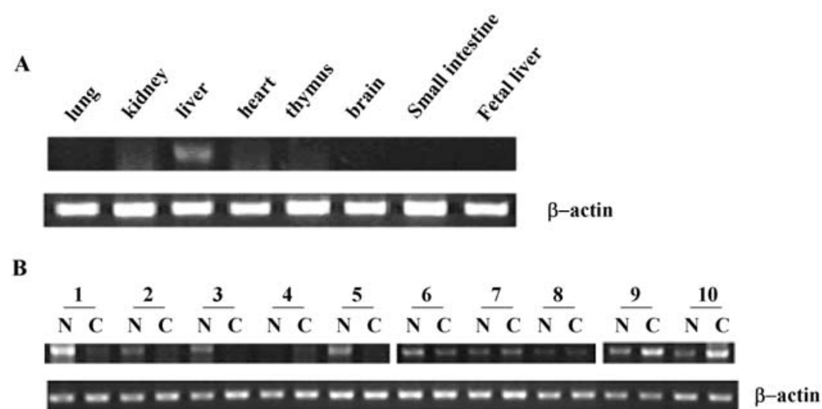

Fig. 5. The *C19orf24* gene expression profile at the RNA level. A - *C19orf24* was specifically expressed in human liver tissue by RT-PCR. RNA from different human tissues was extracted by Trizol (Invitrogen) and reverse transcripted into the cDNA sequence by M-MLV reverse transcriptase. 35 cycles were run in PCR. β-actin was the inter control. B - *C19orf24* was differently expressed in human liver cancer tissue. RNA from 10 pairs of liver cancer patients' samples were extracted by Trizol and reverse transcripted into the cDNA sequence by M-MLV reverse transcriptase. 35 cycles were run in PCR. β-actin was the inter control.

DISCUSSION

Non-classical secreted proteins were identified in eukaryotes approximately 15 years ago [11]. The discovery of IL-1 β and theoredoxin was the first step towards knowledge of proteins that could secrete despite their lack of a signal peptide. It was the combination of biology and information technology that made high throughput computation possible. Secretome P is the program that produces *ab initio* predictions of non-classical protein secretion, *i.e.* secretion that is not signal peptide triggered. This method queries a large number of other feature prediction servers to obtain information on various post-translation and localization aspects of the protein, which are integrated into the final secretion prediction. We used this program and other bioinformatics methods to screen the unconventional secreted protein under study.

The *C19orf24* gene was first predicted to encode a non-classical protein without a signal peptide, and then confirmed in experiments including BFA secretion inhibition, subcellular localization and deglycosylation analysis. However, the mechanism responsible for its secretion remains unknown. The RT-PCR result implies it could have some function in liver tissue, but further study is needed.
The *C19orf24* gene was only found in *Homo sapiens* and *Pan troglodytes*. Owing to the limited studiable resources, it was not possible to discover its origin and describe its phylogenesis. However, we inferred that it evolved recently and could be called young. The ratio of the rate of non-synonymous nucleotide substitution (d_N) to that of synonymous substitution (d_S) is a commonly used indicator of natural selection at the DNA sequence level, so we calculated its d_N/d_S. The data indicated that *C19orf24* had undergone a purifying selection. Most functional genes show a d_N/d_S of <1, because a substantial proportion of non-synonymous mutations are deleterious and are removed by purifying selection, whereas synonymous mutations are more or less neutral and are generally uninfluenced by selection [19]. Therefore, this gene would seem not to have changed its function from *Pan troglodytes* to *Homo sapiens*.

Acknowledgements. We are very grateful to the Shanghai Bioengineering center for their help in the preparation of anti-*C19orf24* antibody. We would also like to thank Yang Hong-Meng for technical help. This work was supported by State 863 High Technology R&D Project of China.

REFERENCES

1. Ladunga, I. Large-scale predictions of secretory proteins from mammalian genomic and EST sequences. **Curr. Opin. Biotechnol.** 11 (2000) 13-18.
2. Fukuda, M.E., Iwadate, Y., Machida, T., Hiwasa, T., Nimura, Y., Nagai, Y., Takiguchi, M., Tanzawa, H., Yamaura, A. and Seki, N. Cathepsin D is a potential serum marker for poor prognosis in glioma patients. **Cancer Res.** 65 (2005) 5190-5194.
3. Wu, C.C., Chien, K.Y., Tsang, N.M., Chang, K.P., Hao, S.P., Tsao, C.H., Chang, Y.S. and Yu, J.S. Cancer cell-secreted proteomes as a basis for searching potential tumor markers: nasopharyngeal carcinoma as a model. **Proteomics** 5 (2005) 3173-3182.
4. Mignatti, P., Morimoto, T. and Rifkin, D.B. Basic fibroblast growth factor, a protein devoid of secretory signal sequence, is released by cells via a pathway independent of the endoplasmic reticulum-Golgi complex. **J. Cell. Physiol.** 151 (1992) 81-93.
5. Rubartelli, A., Cozzolino, F., Talio, M. and Sitia, R. A novel secretory pathway for interleukin-1, a protein lacking a signal sequence. **EMBO J.** 9 (1990) 1503-1510.
6. Marcel, T., Sarah, H. and Steven, L.C. The nonclassic secretion of thioredoxin is not sensitive to redox state. **Am. J. Physiol. Cell Physiol.** 284 (2003) 1272-1279.

7. Walter, N. The mystery of nonclassical protein secretion: A current view on cargo proteins and potential export routes. **Eur. J. Biochem.** 270 (2003) 2109-2119.
8. Cleves, A.E. Protein transport: The nonclassical ins and outs. **Curr. Biol.** 7 (1997) 318-320.
9. Wang, W., Brunet, F.G., Nevo, E. and Long, M. Origin of Sphinx, a young chimeric RNA gene in Drosophila melanogaster. **Proc. Natl. Acad. Sci. USA** 99 (2002) 4448-4453.
10. Wang, W., Zhang, J., Alvarez, C., Llopart, A. and Long, M. The origin of the jingwei gene and the complex modular structure of its parental gene, yellow emperor, in D. melanogaster. **Mol Biol. Evol.** 17 (2000) 1294-1301.
11. Bendtsen, J.D., Kiemer, L., Fausbøll, A. and Brunak S. Non-classical protein secretion in bacteria. **BMC Microbiology** 5 (2005) 58.
12. Bendtsen, J.D., Jensen, L. J., Blom, N., Heijne, G.V. and Brunak S. Feature based prediction of non-classical and leaderless protein secretion. **Protein Eng. Des. Sel.** 17 (2004) 349-356.
13. Gomi, M., Sonoyama, M. and Mitaku S. High performance system for signal peptide prediction: SOSUI signal. **Chem. Bio Info. J.** 4 (2004) 142-147.
14. Lu, Z., Szafron, D., Greiner, R., Lu, P., Wishart, D.S., Poulin, B., Anvik, J., Macdonell, C. and Eisner, R. Predicting subcellular localization of proteins using machine-learned classifiers. **Bioinformatics** 4 (2004) 547-556.
15. Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. **Nucleic Acids Res.** 22 (1994) 4673-4680.
16. Korber, B. HIV signature and sequence variation analysis. In: **Computational Analysis of HIV Molecular Sequences** (Rodrigo, A.G. and Learn, G.H., Eds.) Kluwer Academic Publishers, Dordrecht, Netherlands (2000) 55-72.
17. Blom, N., Gammeltoft, S. and Brunak S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. **J. Mol. Biol.** 294 (1999) 1351-1362.
18. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. Protein identification and analysis tools on the ExPASy server. In: **The Proteomics Protocols Handbook** (Walker, J.M. Ed.) Humana Press, Totowa, New York (2005) 571-607.
19. Zhang, J.Z. Evolution of the human ASPM gene, a major determinant of brain size. **Genetics** 165 (2003) 2063-2070.
20. Philippe, D., Will, H., Guy B. and Denis L.B. Brefeldin A-induced prosomatostatin N-glycosylation in AtT20 cells. **Biochem. Biophys. Res. Commun.** 296 (2002) 618-624.
21. Klausner, R.D., Donaldson, J.G. and Lippincott-Schwartz J. Brefeldin A: insights into the control of membrane traffic and organelle structure. **J. Cell Biol.** 116 (1992) 1071-1080.