

Computational Analyses of TBC Protein Family in Eukaryotes

Xinjiao Gao, Changjiang Jin, Yu Xue* and Xuebiao Yao*

Laboratory of Cellular Dynamics, Hefei National Laboratory for Physical Science at Microscale; and the University of Science and Technology of China, Hefei, Anhui 230027, China

Abstract: Tre-2/Bub2/Cdc16 domain-containing proteins (TBC proteins) participate in wide range cellular processes. With computational approaches, 137 non-redundant TBC proteins from five model organisms were identified and classified into 13 subfamilies base on molecular evolutionary tree. This phylogenetic analysis provides useful functional annotation of newly-identified TBC proteins and guides for further experimentation.

Keywords: TBC domain, phylogenetic tree, USP6, polarized exocytosis, cell cycle, cytokinesis.

INTRODUCTION

The TBC (Tre-2/Bub2/Cdc16) domain was originally identified as a cluster of conserved motifs presented in the human tre-2 oncogene and the yeast cell cycle regulators Bub2 and Cdc16 [1]. TBC domain-containing proteins (TBC proteins) are conserved from yeast to human and function as GTPase activating proteins for Rab proteins specifically (RabGAPs). The crystal structure of TBC domain of yeast Gyp1 has been solved, and together with other kinetics analyses proposes that TBC domains have an arginine finger catalytic mechanism [2,3].

Numerous TBC proteins have been experimentally identified or computationally predicted from yeast to human [4]. These proteins play important roles in cell division [5-9], polarized exocytosis [10], cellular polarity [11], Endoplasmic reticulum (ER) to Golgi transport of proteins [12] and vesicular transport [2,13]. The first TBC protein Bub2 with GAP activity was identified in budding yeast, as a spindle checkpoint during mitosis [5]. Contrast to other spindle checkpoint proteins (e.g. Bub1, Mad1, and Mad2, etc), Bub2 interacts with Bfa1 and localized on spindle pole body (SPB) during cell cycle. The chief regulator of mitosis exit network (MEN) is the small GTPase Tem1, which modulated by Bub2/Bfa1 [7]. Furthermore, Bfa1 phosphorylated by Cdc5 (Polo kinase in mammalians) dynamically regulates the GAP activity of Bub2/Bfa1 [8]. Thus, Bub2/Bfa1 play an essential regulatory role in mitosis exit process through an alternative pathway [6-9]. Both Bub2 and Bfa1 are conserved in fission yeast with the orthologs of Cdc16 and Byr4, with conserved functions during cytokinesis [14]. Later several other TBC proteins were identified experimentally, including Gyp1, Mdr1/Gyp2, Msb3/Gyp3, Msb4/Gyp4, Gyp5, Gyp6, and Gyp7. These TBC-GAP proteins are implicated in numerous non-mitotic processes. The GyPs were identified as GAPs of Ypt proteins. For example, Gyp1 localized on Golgi apparatus as the GAP of Ypt1 [2,15,16], and involved in ER to

Golgi transport process [12]. The Msb3 and Msb4 interact with Cdc42 and Spa2 and regulate the cellular polarity [11]. And Gyp5 and Gyp7 were implicated in the control of polarized exocytosis [10]. In mammals, there were also some TBC proteins reported. During cell cycle, the human Rab-GAP1/GAPCenA is localized on centrosome, as a Rab6 GTPase activating and interacting protein, and plays a role in microtubule nucleation [17]. And human oncogene Evi5 is localized on centrosome and binds with tubulins in interphase cells [18]. During cytokinesis, Evi5 is translocated to midbody, interacting with the chromosomal passenger complex of INCENP-Aurora-B-Survivin [19]. The Tre-2 oncogene, also called as Usp6, is a hominoid-specific gene and functional as a ubiquitin carboxyl-terminal hydrolase [20]. Aberration of either Evi5 or USP will induce a variety of cancers [20,21].

Although experimental efforts have provided enormous knowledge on TBC proteins, many TBC proteins are automatically predicted as hypothetical proteins accumulated in public databases, without any functional annotation. Thus, the newly-predicted/hypothetical TBC proteins should be functional annotated *in silico* for further experimental consideration. And the full content of the TBC superfamily still remain to be dissected and integrated.

In this manuscript, we present a computational analysis for identification and integration of TBC proteins in five model organisms, including budding yeast, nematode, fruit fly, mouse and human. Totally, there have been 137 unique TBC genes/proteins identified, with 11 from budding yeast, 20 from nematode, 23 from fruit fly, 39 from mouse and 44 from human. Then the molecular evolutionary tree has been constructed with Neighbor-Joining (NJ) model. Based on the phylogenetic and functional results, we classify the TBC proteins into 13 sub-families. The detailed information for each sub-family is provided below. And our analysis provides useful functional annotation of newly-identified TBC proteins and guide for further experimentation.

MATERIALS AND METHODS

Data Preparation

The InterPro database has annotated all protein sequences from UniProt database with functional domains and motifs.

*Address correspondence to this author at the Laboratory of Cellular Dynamics, Hefei National Laboratory for Physical Science at Microscale; and the University of Science and Technology of China, Hefei, Anhui 230027, China; Tel: +86(551)360-7821; Fax: +86(551)360-7141; E-mail: xueyu@ustc.edu.cn; yaorb@ustc.edu.cn

The TBC domain in InterPro database is recorded as RabGAP/TBC (IPR000195). Totally, there are 1,554 proteins annotated as TBC domain-containing proteins.

In this work, we retrieve all the protein sequences of TBC proteins of five model organisms, including *Saccharomyces cerevisiae* (budding yeast), *Caenorhabditis elegans* (nematode), *Drosophila melanogaster* (fruit fly), *Mus Musculus* (mouse) and *Homo sapiens* (human). And 348 TBC proteins are obtained, with 22 from budding yeast, 28 from nematode, 52 from fruit fly, 106 from mouse and 140 from human.

Redundant Cleaning

The sequence data from UniProt are highly redundant, e.g., some proteins are transcribed from a same gene differing only by a few amino acids, and many of them are alternative splicing isoforms. To obtain a non-redundant data set of TBC proteins, we developed a simple but faster computational program.

All the sequences of TBC proteins were mapped on the entire genomes of budding yeast, nematode, fruit fly, mouse and human by BLAT [22], which is useful for localization positioned of a protein or DNA sequence on its genome. All coordinates of the TBC proteins were obtained. If the coordinates of two proteins are the same or their coordinates are overlapped, the shorter one is discarded. Totally 137 unique TBC proteins were identified.

Multiple Sequence Alignment and Phylogenetic Analyses

Since all the TBC proteins contain the conserved TBC domains, we retrieved these TBC domains of the TBC proteins. The sequences of the TBC domains were aligned with ClustalX (Version 1.83) with default parameters as previously described [23].

Then the molecular evolutionary NJ tree with Poisson Correction (PC) was constructed with MEGA program (Version 3.1) [24]. And the Bootstrap testing was also deployed to promise the accuracy of the topology of the phylogenetic tree.

Based on the phylogenetic and functional analyses, we classified 137 TBC proteins into 13 subfamilies.

RESULTS AND DISCUSSION

137 Unique TBC Proteins in Five Model Organisms

Previously computational analysis has predicted 24 and 52 TBC gene in fruit fly and human, separately [4]. During database updated, many entries will also be updated, and many novel TBC proteins will be discovered. Since without genomic localization, several proteins may turn out to be the same gene as redundant data. Also, InterPro database automatically annotate 1,554 proteins containing TBC domains in all available organisms.

In this work, we focused on the TBC proteins identification among five model organisms, including *Saccharomyces cerevisiae* (budding yeast), *Caenorhabditis elegans* (nematode), *Drosophila melanogaster* (fruit fly), *Mus Musculus* (mouse) and *Homo sapiens* (human). First, we scanned all

the protein sequences of the proteome of the five organisms with InterProScan [25]. The sequences with RabGAP/TBC (IPR000195) hits were preserved for further analysis. And totally, there were 363 TBC protein sequences reserved.

Since the data set was highly redundant with many proteins transcribed from the same genes, we localized all the protein sequence of TBC proteins onto their genomes, respectively. The BLAT program was employed [22]. And the coordinates of the TBC proteins were gained. We compared the coordinates of the TBC proteins pairwise for each organism. If the coordinates of two TBC proteins were the same or overlapped, the shorter sequence was discarded.

Finally 137 unique TBC proteins were computationally identified including 11 from budding yeast, 20 from nematode, 23 from fruit fly, 39 from mouse and 44 from human.

There are several popular databases developed for functional domain annotation with various algorithms, including PROSITE, Pfam, SMART and InterPro, etc. We searched these databases for annotated TBC proteins, with 300, 302, 186 and 348 hits in PROSITE, Pfam, SMART and InterPro databases, respectively (Table 1). Although the total numbers of TBC proteins are much more than our results, there are too many redundant records in these databases. And also, these databases are not integrated with several entries missed. For example, a typical TBC protein HsTbc1d13 (Q9NVG8) can not be detected by PROSITE.

Phylogenetic Analysis of TBC Superfamily Members

Although numerous TBC proteins were identified in yeast and mammals, the functional and evolutionary relationships among these proteins are still elusive and remain to be elucidated. To address the question, we performed a phylogenetic analysis for TBC superfamily and classify the TBC proteins into several sub-families, based on functional and evolutionary results.

Firstly, we retrieved all the TBC domain sequences from the TBC proteins. Then the domain sequences were aligned by ClustalX (Version 1.83) with default parameters as previously described [23]. The molecular evolutionary NJ tree with PC was constructed with MEGA program (Version 3.1) [24]. The Bootstrap evaluation was also employed to avoid irresponsible topologies. We also tried Minimum Evolution (ME) model to construct the evolutionary tree, with little difference with NJ tree. The NJ tree is un-root and shown in Fig. 1.

The Sub-Families Classification of TBC Proteins

Based on phylogenetic analyses and existed functional annotations, we classified the 137 unique TBC proteins into 13 subfamilies shown in Fig. 1. Each group has the special subcellular localizations and plays different roles in various cellular processes.

The TBC-1 sub-family contains 8 proteins. The Tbc1d10a of human (Swissprot ID: Q9BXI6) or mouse (P58802), also named as EPI64 (EBP50-PDX interactor of 64 kDa), was identified as a GAP specific for the small GTPase Rab27A [26]. The DTYLA mutation of TBC domain of Tbc1d10a/EPI64 will eliminate its function to in-

Table 1. The TBC Proteins Annotated in PROSITE, Pfam, SMART and Interpro Database and Non-Redundant Data Set of Our Work. There are 348 Proteins in Five Model Organisms Automatically Annotated with TBC Domain in InterPro Database, 300 in PROSITE Database, 290 in Pfam Database and 186 in SMART Database. And Our Analysis Provided a Number of 137 Unique TBC Proteins Listed in Table

Toolkits	total	S.cerevisiae	C.elegans	D.melanaogaster	M.Musculus	H.Sapiens
PROSITE	300	10	25	42	108	115
Pfam	302	13	25	34	123	107
SMART	186	10	22	28	56	70
InterPro	348	22	28	52	106	140
Our results	137	11	20	24	39	43

duce the melanosome aggregation [26]. The Tbc1d10a/EPI64 also interacts with the PDZ domain of EBP50, playing an important roles in microvillar morphology. Since the proteins in TBC-1 group are highly conserved, we propose these proteins might have similar functions.

The TBC-2 group is the most interesting sub-family in our study. The human USP6/Tre-2 oncogene is a hominoid-specific gene, with DNA-binding activity. The USP6/Tre-2 is functional as a ubiquitin carboxyl-terminal hydrolase in CaM-dependent manner *in vivo* [27], controlled a signaling pathway leading to actin cytoskeleton remodeling and regulated the Arf6-dependent plasma membrane recycling system [13]. And its aberration will induce a variety of cancers [20]. In our analysis, we also found several other hominoid-specific genes, including HsTbc1d3, HsTbc1d3p2, HsMGC51025. These genes have no orthologs in both mouse and rat and might be originated after human-rodents divergence, with distinct coordinates on human genome (see in Table 2). The Usp6nl is similar with the N-terminal of USP6, without the peptidase domain. And the HsTbc1d3, HsTbc1d3p2, HsMGC51025 are also lack of peptidase domain. Thus, in TBC-2 sub-family, only the human USP6/Tre-2 is a ubiquitin carboxyl-terminal hydrolase. And the functions of other components still remain to be dissected.

The TBC-4 sub-family comprises 19 proteins, which specifically activate the GTPase activity of Rabs and probably function in the cross-talk of membrane traffic and cell adhesion. Yeast Gyp5p, another GAP of Ypt1p, accelerates the intrinsic GTPase activity of Ypt1p, but Gyp5p displayed different subcellular localization from Gyp1p as previously described [12]. In the progress of cytokinesis, Gyp5p partly co-localizes at the bud emergence site, the bud tip and the bud neck and takes part in control of polarized exocytosis [10]. In mammals, human Evi5 is an oncogene [21] and localizes on centrosome and binds with alpha- and beta-tubulins during interphase [18]. During cytokinesis, Evi5 translocates to the midbody, interacting with the centromere passenger protein complex of INCENP-Aurora-B-Survivin [19]. Also, Evi5 is a specific GAP of Rab11 [28]. Another component RabGAP1/GAPCenA is localized on centrosome, as a Rab6 GTPase activating and interacting protein, participates in the coordination of microtubule and Golgi dynamics

during the cell cycle, and plays a role in microtubule nucleation [17]. Taken together, the members of TBC-4 group might be implicated in cellular polarization and cell cycle process.

The TBC-5 is the most divergent group in our results. The HsGrtp1 is a growth hormone Regulated TBC Protein [29], and the detailed function is unclear. But in budding yeast, the Msb3 and Msb4 interact with Cdc42 and Spa2 and regulate the cellular polarity [11]. These two proteins both located at the budding site and were the cellular components of bud tip. And Mdr1 located in cytosol to regulated the protein trafficking [30]. Also, the HsTbc1d2 (Q9BYX2) is differentially expressed in prostate normal and cancer cells, proposing a potential role in tumorigenesis. We propose the proteins of TBC-5 group might be implicated in cellular polarity and tumorigenesis in metazoans.

The TBC-7 is a small group and specific for yeast. The Ypt6p is the preferred substrate of Gyp6p, which activates the GTPase activity of Ypt6p [15]. A complete loss of GAP activity and Ypt6p binding were resulted from short N-terminal and C-terminal deletions of Gyp6p [31]. Although the Ypt6p-specific Gyp6p is the smallest of the eight known Ypt/Rab GAPs from yeast with 458 amino acids, there is no ortholog found in other organisms based on sequence alignment. And the function of Gyp8 is unknown. But with its high similarity with Gyp6, we propose Gyp8 is also a GAP of Ypt6 with similar function of Gyp6 [15].

The core protein of TBC-9 sub-families Gyp1, with function as GTPase regulators, which play key roles in internalized membrane material recycling and vesicle trafficking. Hydrolysis of Ypt1p-bound GTP is negatively regulated by Gyp1p to balance vesicle transport between endoplasmic reticulum and Golgi apparatus. The Gyp1 is localized on Golgi apparatus as the GAP of Ypt1 [2,15,16].

In TBC-10 group, the yeast Gyp7p is conserved from budding yeast to human based on sequence alignment. It accelerates the GTPase activity of preferred substrate Ypt7, compared to other proteins, such as Ypt31, Ypt32, Ypt1, Ypt6 and Sec4. Also, the Gyp7 was implicated in the control of polarized exocytosis [10]. Moreover, the human Tbc1d15 of TBC-10 group was identified as a specific GAP of Rab7 [32].

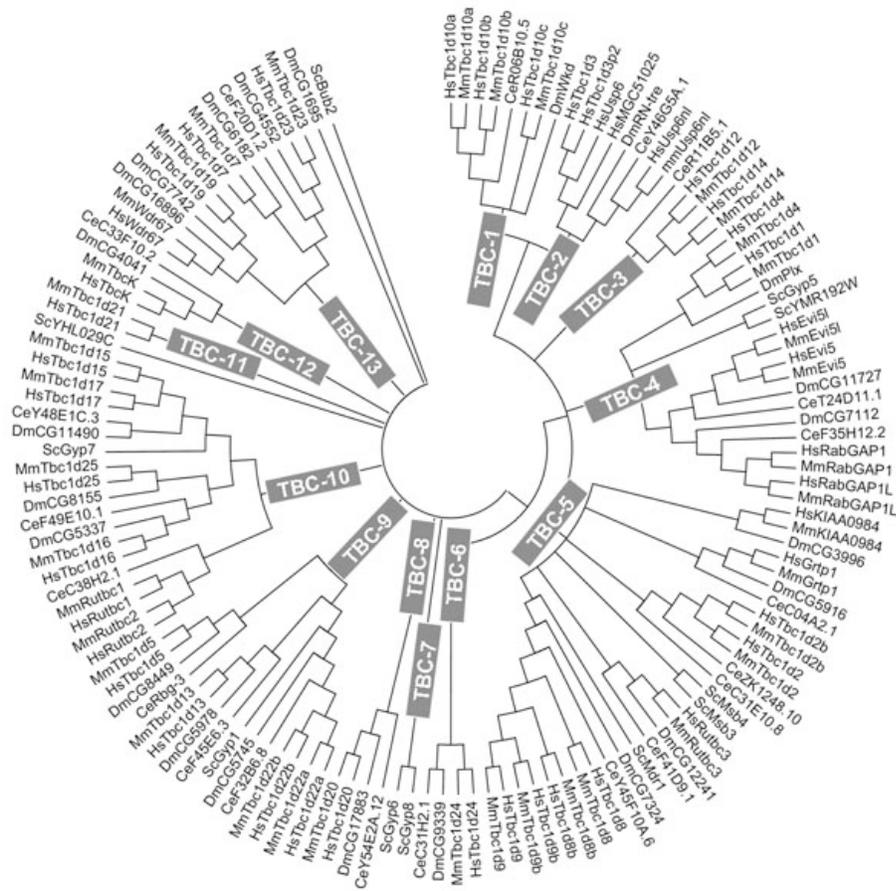


Fig. (1). Evolutionary analysis and classification of TBC superfamily. The 137 TBC proteins from five model organisms were employed, and the Neighbor-Joining (NJ) tree with bootstrap testing was constructed. Based on the phylogenetic and functional analyses, we classified the TBC proteins into 13 subfamilies.

Table 2. The Hominoid-Specific Genes of Usp6nl in TBC-2 Group. The HsUSP6, HsTbc1d3, HsTbc1d3p2 and HsMGC51025 are Generated by Gene Duplication of Usp6nl, After Human-Rodents Divergence

TBC-2	Swissprot ID	Chromosome No.	Strand	Start	End
MmUsp6nl	Q80XC3	2	++	6308329	6359011
HsUsp6nl	Q92738	10	+ -	11544449	11609570
HsUSP6	P35125	17	++	4974387	5016994
HsTbc1d3	Q8IZP1	17	+ -	33541609	33550662
HsTbc1d3p2	Q6PD72	17	+ -	57698795	57706258
HsMGC51025	Q86UD7	17	++	15579388	15586064

All of the TBC-12 proteins contain a protein kinase (PK) domain. We propose these proteins could have kinase activity to phosphorylate their substrates.

Intriguingly, although Usp6 and Usp6nl of TBC-2 group were shown moderate similarity with their distant homologs Msb3 and Msb4 of TBC-5 sub-family in budding yeast, however, they can totally replace Msb3 and Msb4 in yeast [33].

The functions of TBC-3, TBC-6, TBC-8, TBC-11 and TBC-13 groups are still elusive from yeast to human. And further experimental identification will shed a light on the functions of these proteins. Also, several TBC proteins could not be classified into any sub-families and regarded as orphan proteins, e.g., Bub2 and YHL029C of budding yeast.

In our results, only sixteen TBC proteins are well annotated with subcellular localizations. Interestingly, these pro-

teins are all localized on the plasma membrane or membrane of cellular apparatus. Most of these TBC proteins could interact with several other proteins. Thus, TBC proteins appear to function as adaptors which bridge between membrane proteins. Actually, our analyses propose that most of the TBC proteins can active a variety of small Rab GTPases, which are important in vesicular membrane trafficking.

An open question is why most of the GAPs of the largest subfamily of Rab proteins have TBC domains? And does the TBC domain play an important role on activating the GTPase activity of Rabs? Some researches have supported the hypothesis that the complete TBC domain is necessary for the function of RabGAPs. The conformational change from GTP-state to GDP-state is controlled by RabGAPs [2,3]. The function of TBC domain may be better revealed by analysis all the TBC proteins in the future.

CONCLUSION

In this work, we have identified 137 unique TBC proteins from five model organisms, with 11 from budding yeast, 20 from nematode, 23 from fruit fly, 39 from mouse and 44 from human, respectively. The protein sequences of TBC domains were retrieved and aligned altogether. Then we constructed an evolutionary Neighbor-Joining tree with Poisson Correction and Bootstrap testing. Based on the phylogenetic and functional information, we classified the 137 TBC proteins into 13 sub-families. The *in silico* study of these TBC proteins will lay the groundwork for a more thorough investigation of the functions of this superfamily. The functional relevance of several TBC sub-families, including TBC-1, -2, -4, -5, -7, -9, 10 and -12 were computationally predicted, which provides useful for further experimentation.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of China (Grant Nos. 39925018, 30121001, 30270293 & 90508002), the Chinese Academy of Sciences (Grant No. KSCX2-2-01), the Chinese 973 Project (Grant No. 2002CB713700), the Chinese 863 Project (Grant No. 2001AA215331), Chinese Minister of Education (Grant No. 20020358051), and American Cancer Society (Grant No. RPG-99-173-01).

ABBREVIATIONS

TBC	= Tre-2/Bub2/Cdc16 (domain)
TBC proteins	= TBC domain-containing proteins
GAPs	= GTPase activating proteins
RabGAPs	= GTPase activating proteins for Rab proteins specifically
NJ	= Neighbor-Joining model
PC	= Poisson Correction
ER	= Endoplasmic reticulum

REFERENCES

- [1] Richardson, P.M. and Zon, L.I. (1995) *Oncogene*, 11, 1139-48.
- [2] Rak, A. Fedorov, R. Alexandrov, K. Albert, S. Goody, R.S. Gallwitz, D. and Scheidig, A.J. (2000) *The EMBO journal*, 19, 5105-13.
- [3] Pan, X. Eathiraj, S. Munson, M. and Lambright, D.G. (2006) *Nature*, 442, 303-6.
- [4] Bernards, A. (2003) *Biochimica et Biophysica Acta*, 1603, 47-82.
- [5] Hoyt, M.A. Totis, L. and Roberts, B.T. (1991) *Cell*, 66, 507-17.
- [6] Frascini, R. D'Ambrosio, C. Venturetti, M. Lucchini, G. and Piatti, S. (2006) *The Journal of cell biology*, 172, 335-46.
- [7] Cooper, J.A. and Nelson, S.A. (2006) *The Journal of cell biology*, 172, 331-3.
- [8] Geymonat, M. Spanos, A. Walker, P.A. Johnston, L.H. and Sedgwick, S.G. (2003) *The Journal of biological chemistry*, 278, 14591-4.
- [9] Gardner, R.D. and Burke, D.J. (2000) *Trends in cell biology*, 10, 154-8.
- [10] Chesneau, L. Dupre, S. Burdina, A. Roger, J. Le Panse, S. Jacquet, M. and Cuif, M.H. (2004) *Journal of cell science*, 117, 4757-67.
- [11] Tcheperegine, S.E. Gao, X.D. and Bi, E. (2005) *Molecular and cellular biology*, 25, 8567-80.
- [12] De Antoni, A. Schmitzova, J. Treppe, H.H. Gallwitz, D. and Albert, S. (2002) *The Journal of biological chemistry*, 277, 41023-31.
- [13] Martinu, L. Masuda-Robens, J.M. Robertson, S.E. Santy, L.C. Casanova, J.E. and Chou, M.M. (2004) *Molecular and cellular biology*, 24, 9752-62.
- [14] Furge, K.A. Wong, K. Armstrong, J. Balasubramanian, M. and Albright, C.F. (1998) *Current biology*, 8, 947-54.
- [15] Strom, M. Vollmer, P. Tan, T.J. and Gallwitz, D. (1993) *Nature*, 361, 736-9.
- [16] Du, L.L. and Novick, P. (2001) *Molecular biology of the cell*, 12, 1215-26.
- [17] Cuif, M.H. Possmayer, F. Zander, H. Bordes, N. Jollivet, F. Couedel-Courteille, A. Janoueix-Lerosey, I. Langsley, G. Bornens, M. and Goud, B. (1999) *The EMBO journal*, 18, 1772-82.
- [18] Faitar, S.L. Dabbekeh, J.T. Ranalli, T.A. and Cowell, J.K. (2005) *Genomics*, 86, 594-605.
- [19] Faitar, S.L. Sossey-Alaoui, K. Ranalli, T.A. and Cowell, J.K. (2006) *Experimental cell research*, 312, 2325-35.
- [20] Paulding, C.A. Ruvolo, M. and Haber, D.A. (2003) Proceedings of the National Academy of Sciences of the United States of America, 100, 2507-11.
- [21] Roberts, T. Chernova, O. and Cowell, J.K. (1998) *Human molecular genetics*, 7, 1169-78.
- [22] Kent, W.J. (2002) *Genome research*, 12, 656-64.
- [23] Thompson, J.D. Gibson, T.J. Plewniak, F. Jeanmougin, F. and Higgins, D.G. (1997) *Nucleic acids research*, 25, 4876-82.
- [24] Kumar, S. Tamura, K. and Nei, M. (2004) *Briefings in bioinformatics*, 5, 150-63.
- [25] Quevillon, E. Silventoinen, V. Pillai, S. Harte, N. Mulder, N. Apweiler, R. and Lopez, R. (2005) *Nucleic acids research*, 33, W116-20.
- [26] Itoh, T. and Fukuda, M. (2006) *The Journal of biological chemistry*, 281, 31823-31.
- [27] Shen, C. Ye, Y. Robertson, S.E. Lau, A.W. Mak, D.O. and Chou, M.M. (2005) *The Journal of biological chemistry*, 280, 35967-73.
- [28] Dabbekeh, J.T. Faitar, S.L. Dufresne, C.P. and Cowell, J.K. (2007) *Oncogene*, 26, 2804-8.
- [29] Lu, C. Kasik, J. Stephan, D.A. Yang, S. Sperling, M.A. and Menon, R.K. (2001) *Endocrinology*, 142, 4568-71.
- [30] Huh, W.K. Falvo, J.V. Gerke, L.C. Carroll, A.S. Howson, R.W. Weissman, J.S. and O'Shea, E.K. (2003) *Nature*, 425, 686-91.
- [31] Will, E. and Gallwitz, D. (2001) *The Journal of biological chemistry*, 276, 12135-9.
- [32] Zhang, X.M. Walsh, B. Mitchell, C.A. and Rowe, T. (2005) *Biochemical and biophysical research communications*, 335, 154-61.
- [33] Bizimungu, C. De Neve, N. Burny, A. Bach, S. Bontemps, F. Portetelle, D. and Vandenbol, M. (2003) *Biochemical and biophysical research communications*, 310, 498-504.