ISSN:0974-276X

Journal of
Proteomics & Bioinformatics

The International Open Access
Journal of Proteomics and Bioinformatics Studies

Available online at
OMICS Publishing Group
www.omicsonline.com

Digital Object Identifier: http://dx.doi.org/10.4172/jpb.1000133

Journal of Proteomics & Bioinformatics - Open Access

www.omicsonline.com

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jpb.1000133

JPB/Vol.3 Issue 5

Evaluation of Computational Methods for Secreted Protein Prediction in Different Eukaryotes

Xiang Jia Min*

Center for Applied Chemical Biology, Department of Biological Sciences, Youngstown State University, Youngstown, OH 44555, USA

Abstract

Secreted proteins play important biological roles in eukaryotes. Computational identification of all secreted proteins, *i. e.* the secretome, from predicted proteome of completely sequenced genomes is an essential step in functional annotation. To develop screening methods for secreted proteins in different kingdoms of eukaryotes, we have evaluated the prediction accuracies of SignalP, Phobius, TargetP, and WolfPsort used individually or in combination with TMHMM and PS-Scan. Prediction accuracy was represented by Mathews' Correlation Coefficient (MCC). The tools show different strength for predicting secreted proteins in different kingdoms of eukaryotes. When individual tools were used, we found that the tools having the highest accuracy were WolfPsort for fungi (73.1%), Phobius for animals (82.8%), SignalP for plants (55.4%), and Phobius for protists (42.1%), respectively. Except using Phobius, combining the prediction tools with TMHMM significantly improved the prediction accuracy in all data sets. Based on the measured accuracies, it is recommended that using the following methods for secretome prediction in different eukaryotes: SignalP/TMHMM/WolfPsort/Phobius/PS-Scan for fungi (83.4%), Phobius/WolfPsort/PS-Scan for animals (86.7%), SignalP/TMHMM/Phobius/TargetP/PS-Scan for plants (73.2%), and combining all the tools for protists (52.8%).

Keywords: Secreted proteins; Secreteome; Prediction; Signal peptide; Method

Introduction

Eukaryotic cells make thousands of different proteins. Among them some proteins function only after they are secreted from the cell. These secreted proteins play critical biological roles. For example, fungi secrete enzymes to break down potential food sources (Tsang et al., 2009); plant secreted proteins are primarily part of cell wall proteome (Jamet et al., 2008); human secreted proteins are involved in cellular immunity and communication and provides useful information for the discovery of novel biomarkers such as for cancer diagnosis (Hathout 2007; Xue et al., 2008).

Almost all eukaryotic secreted proteins contain a signal peptide at the N-terminus that directs proteins to the rough ER and the Golgi complex (Blobel and Dobberstein, 1975; von Heijne, 1990). The signal peptide, typically 15 – 30 amino acids long, is cleaved off during translocation across the membrane. While some proteins without an N-terminal signal peptide can be found in the ER and the Golgi, over 90% of human secreted proteins contain classical N-terminal signal peptides (Scott et al., 2004). A number of computational tools have been developed for predicting subcellular locations of proteins including secretion to the extracellular space (Emanuelsson et al., 2007). Among the computational tools, SignalP (Bendtsen et al., 2004) in conjunction with TMHMM (http://www.cbs.dtu.dk/services/TMHMM/), TargetP (Emanuelsson et al., 2000; Emanuelsson et al., 2007), and Phobius (Kall et al., 2004; Kall et al., 2007) were widely applied to genomewide prediction of all secreted proteins in an organism, i. e. putative secretomes; and some of the predicted secreted proteins were experimentally validated using proteomics approaches (Wymelenberg et al., 2005; Tsang et al., 2009). For example, secretomes are analyzed in the nematode *Nippostrongylus brasiliensis* (Harcus et al., 2004) and the vetigastropod *Haliotis asinina* (Jackson et al., 2006) using SignalP, human, puffer fish, and pig using TargetP (Klee et al., 2004), and zebrafish using SignalP/TargetP/Phobius/pTarget (Klee, 2008). A number of fungal secretomes including *Candida albicans* (Lee et al., 2003), *Phanerochaete chrysosporium* (Wymelenberg et al., 2005), and *Aspergillus* niger (Tsang et al., 2009) were predicted using SignalP/TMHMM/TargetP method.

The quality of secretome prediction depends primarily on the accuracy of the computational tools. The prediction performances of some earlier versions of these tools mentioned above have been evaluated (Menne et al., 2000). Klee and Ellis, (2005) evaluated SignalP (version 2.0 and 3.0), TargetP (version 1.01), Phobius, PrediSi, and ProtComp using mammalian data extracted from SwissProt database and proposed to use TargetP, SignalP 2.0 maximum Y-score, and SignalP 3.0 maximum S-score to increase accuracy (Klee and Ellis, 2005). Chen et al. (2003) compared CJ-SPHMM/TMHMM with PSORT and found that combing these tools improved the prediction accuracy and subsequently used this method to construct a secreted protein database for human, mouse and rat (Chen et al., 2003; Chen et al., 2005). O'Toole et al. (2005) compared SignalP/TMHMM with Phobius using manually curated *Saccharomyces cerevisiae* data in the SGD database (http://www.yeastgenome.org/) and found Phobius predicted secreted proteins with higher sensitivity and specificity than using SignalP/TMHMM (O'Toole et al., 2005). However, Tsang et al. (2009) recently showed that the prediction of secreted proteins using SignalP/TMHMM results in higher specificity than Phobius in *A. niger* experimental data (Tsang et al., 2009). In addition, they also observed that

***Corresponding author:** Xiang Jia Min, Center for Applied Chemical Biology, Department of Biological Sciences, Youngstown State University, Youngstown, OH 44555, USA, E-mail: xmin@ysu.edu

Received March 15, 2010; **Accepted** April 21, 2010; **Published** April 21, 2010

Citation: Min XJ (2010) Evaluation of Computational Methods for Secreted Protein Prediction in Different Eukaryotes. J Proteomics Bioinform 3: 143-147. doi:10.4172/jpb.1000133

the combined results from the Phobius and SignalP/TMHMM protocols yielded higher specificity than these protocols were used separately (Tsang et al., 2009).

Taken together, there is a need for comparative assessment of the prediction accuracies of computational tools, particularly the combination of these tools, for organisms in specific kingdoms of eukaryotes. In order to develop computational methods for secretome prediction from species in different kingdoms of eukaryotes, we have evaluated the accuracies of SignalP, Phobius, TargetP, and a new version of PSort – WolfPsort (Horton et al., 2007), and their synergistic performances in conjunction with TMHMM and PS-Scan (http://www.expasy.org/tools/scanprosite/) using manually curated data of fungi, animals, plants and protists in Swiss-Prot database. PS-Scan was included to remove ER targeted proteins (de Castro et al., 2006).

Data and Methods

Manually annotated secreted (positives) and nonsecreted (negatives) proteins in eukaryotes were extracted from UniProt-SwissProt database (UniProt release 15.4; June 2009) (http://www.ebi.ac.uk/uniprot/database/download.html). The positive dataset was extracted from entries annotated as "secreted" or "extracellular" in the subcellular location. The negative dataset includes all entries with a subcellular location in membrane, endoplasmic reticulum (ER), cytoplasm, nucleus, mitochondrion, etc. Entries annotated both as "secreted" or "extracellular" and also located in one or more other subcellular locations were eliminated. Entries in the subcellular location annotated as "Putative", "Potential", "Possible", or "By similarity" were also eliminated. Protein sequences not starting with methionine (M) or having less than 50 amino acids were eliminated. The dataset was subdivided into four datasets consisting of 241 secreted proteins and 5,992 nonsecreted proteins in fungi, 5,568 secreted proteins and 19,048 nonsecreted proteins in animals including human (metazoan), 216 secreted proteins and 7,528 nonsecrected proteins in plants, 32 secreted proteins and 1,979 nonsecreted proteins in protists, respectively. The data are available upon request from the author.

Four secreted protein predictors including SignalP (version 3.0) (Bendtsen et al., 2004), Phobius (Kall et al., 2004; Kall et al., 2007), TargetP (Emanuelsson et al., 2000), and WolfPsort (Horton et al., 2007; Sprenger et al., 2006) were selected as they were favourably evaluated previously (Klee and Ellis, 2005) and were widely used in the work mentioned above (see the references in Introduction). To improve the accuracy for secreted protein prediction, we also included TMHMM and PS-Scan. TMHMM was used to identify membrane proteins and was often used in tandem with SignalP (Emanuelsson et al., 2007; Tsang et al., 2009). PS-Scan was used to scan ER targeting sequence (Prosite: PS00014) (de Castro et al., 2006; O'Toole et al., 2005). Standalone-for-Linux version of SignalP (version 3.0) (http://www.cbs.dtu.dk/services/SignalP/), Phobius (http://phobius.binf.ku.dk/), TargetP (http://www.cbs.dtu.dk/services/TargetP/), WolfPsort (http://wolfpsort.org/), TMHMM (http://www.cbs.dtu.dk/services/TMHMM), and PS-Scan (http://www.expasy.org/tools/scanprosite/) were obtained from the authors or downloaded from available websites. The default parameters were used for all the programs, except for processing protist data set as no protist specific parameters were available, we found TargetP performed better using the non-plant parameters and WolfPsort performed better using the fungal parameters. For SignalP prediction, only entries that are predicted to have a "mostly likely cleavage site" by SignalP-NN algorithm and a "signal peptide" by SignalP-HMM algorithm are considered to be true "positives" (Bendtsen et al., 2004) using the N-terminal 70 amino acids. For predicting membrane proteins using TMHMM, the entries having membrane domains not located within the N-terminus (the first 70 amino acids) were treated as real membrane proteins since our analysis showed treating entries having a single transmembrane domain located in the N-terminus significantly decreased the prediction accuracy. The performance of each individual prediction tools and various combinations of these tools was measured using prediction sensitivity (Equation 1), specificity (Equation 2) and Mathews' Correlation Coefficient (MCC) (Equation 3) (Mathews, 1975; Baldi et al., 2005; Menne et al., 2000).

$$\text{Sensitivity (\%)} = TP/(TP + FN) \times 100 \qquad (1)$$

$$\text{Specificity (\%)} = TN/(TN + FP) \times 100 \qquad (2)$$

$$\text{MCC (\%)} = (TP \times TN - FP \times FN) \times 100 / ((TP + FP)(TP + FN)(TN + FP)(TN + FN))^{1/2} \qquad (3)$$

Where TP is the number of true positives, FN is the number of false negatives, FP is the number of false positives, and TN is the number of true negatives. When multiple tools were used, only the entries predicted to be positives by all the tools were taken as true positives.

Results

The performances of each individual prediction tool, combining with THMM, and combining with PS-Scan were evaluated in four datasets separately. The sensitivity, specificity, and Mathews' Correlation Coefficient were reported in Table 1 – 4. Because the sensitivity and the specificity were almost always in opposite trend, we used the MCC value to represent prediction accuracy of each individual tool or the combination of the tools.

Prediction of fungal data

Prediction accuracies for fungal secreted proteins are shown in Table 1. For individual tools, based on the MCC values, the performance is in the order of WolfPsort > Phobius > SignalP > TargetP. When TMHMM was used in tandem after these individual tools, the accuracies were substantially improved in SignalP/TMHMM (11.4% increment in MCC) and TargetP/TMHMM (14.9% increment in MCC). However, using TMHMM only slightly improved the performance of WolfPsort, and no improvement was observed in Phobius. Phobius has incorporated an algorithm to eliminate membrane proteins (Kall et al., 2004; Kall et al., 2007), thus the result was expected. The highest accuracy was achieved by combining SignalP, WolfPsort, and Phobius for signal peptide prediction, TMHMM for eliminating membrane proteins, and PS-Scan for removing ER targeting proteins (Table 1). However, adding TargetP to the above pipeline slightly reduced the accuracy, thus it is not recommended for processing fungal data. Our results are consistent with experimental validation in *A. niger* that reported higher specificity of secreted proteins was generated when combing SignalP/TMHMM with Phobius than when used individually, and TargetP erroneously assigned some secreted proteins to mitochondrion (Tsang et al., 2009).

	TP	FP	TN	FN	Sn (%)	Sp (%)	MCC (%)
SignalP	232	329	5663	9	96.3	94.5	61.2
Phobius	226	203	5789	15	93.8	96.6	68.8
TargetP	228	583	5409	13	94.6	90.3	48.6
WolfPsort	230	167	5825	11	95.4	97.2	73.1
SignalP/TMHMM	228	168	5824	13	94.6	97.2	72.6
Phobius/TMHMM	224	200	5792	17	92.9	96.7	68.6
TargetP/TMHMM	224	265	5727	17	92.9	95.6	63.5
WolfPsort/TMHMM	227	135	5857	14	94.2	97.7	75.8
SignalP/TMHMM/WolfPsort	226	86	5906	15	93.8	98.6	81.6
SignalP/TMHMM//WolfPsort/Phobius	222	69	5923	19	92.1	98.8	83.1
SignalP/TMHMM/WolfPsort/Phobius/PS-Scan	**222**	**67**	**5925**	**19**	**92.1**	**98.9**	**83.4**
SignalP/TMHMM/WolfPsort/Phobius/TargetP/PS-Scan	218	66	5926	23	90.5	98.9	82.6

TP: true positives; FP: false positives; TN: true negatives; FN: false negatives. Sn: sensitivity; Sp:specificity; MCC: Mathews' correlation coefficient.

Table 1: Prediction accuracies of secreted proteins in fungi.

	TP	FP	TN	FN	Sn (%)	Sp (%)	MCC (%)
SignalP	5307	4108	14940	261	95.3	78.4	63.5
Phobius	5157	1167	17881	411	92.6	93.9	82.8
TargetP	5313	5412	13636	255	95.4	71.6	56.5
WolfPsort	5135	1762	17286	433	92.2	90.7	77.3
SignalP/TMHMM	5217	1383	17665	351	93.7	92.7	81.6
Phobius/TMHMM	5148	1142	17906	420	92.5	94.0	82.9
TargetP/TMHMM	5222	1369	17679	346	93.8	92.8	81.8
WolfPsort/HMM	5093	1084	17964	475	91.5	94.3	82.8
Phobius/WolfPsort	4959	555	18493	609	89.1	97.1	86.4
Phobius/WolfPsort/PS-Scan	**4956**	**531**	**18517**	**612**	**89.2**	**97.2**	**86.7**
Phobius/WolfPsort/TMHMM	4952	544	18504	616	88.9	97.1	86.5
Phobius/WolfPsort/TMHMM/SignalP	4952	544	18504	616	88.9	97.1	86.5
Phobius/WolfPsort/TMHMM/TargetP	4934	505	18543	634	88.6	97.3	86.7
Phobius/WolfPsort/TMHMM/TargetP/PS-Scan	**4931**	**482**	**18566**	**637**	**88.6**	**97.5**	**86.9**
Phobius/WolfPsort/TMHMM/TargetP/PS-Scan/SignalP	4931	482	18566	637	88.6	97.5	86.9

TP: true positives; FP: false positives; TN: true negatives; FN: false negatives. Sn: sensitivity; Sp:specificity; MCC: Mathews' correlation coefficient.

Table 2: Prediction accuracies of secreted proteins in animals.

Prediction of animal data

Prediction accuracies for secreted proteins in animals are shown in Table 2. For individual tools, based on the MCC values, the performance is in the order of Phobius > WolfPsort > SignalP > TargetP. When TMHMM was used in tandem after these individual tools, the accuracies were substantially improved in SignalP/TMHMM (18.1% increment in MCC) and TargetP/TMHMM (25.3% increment in MCC). Using TMHMM following WolfPsort was also improved the prediction accuracy (5.5% increment in MCC) using TMHMM. However, when Phobius and TMHMM were combined, there was no improvement in prediction accuracy. Combining the top two prediction tools, Phobius and WolfPsort, and then PS-Scan further improved the prediction accuracy to a MCC value of 86.7%. Though adding TargetP and TMHMM to the pipeline further increased the accuracy to 86.9% (Table 2), for most of the applications, the method of combining Phobius/WolfPsort/PS-Scan would be adequate for predicting secreted proteins in animals. Certainly there is a trade-off between adding more tools for improving the accuracy and the efficiency for using less number of tools for processing animal data.

Prediction of plant data

Prediction accuracies for secreted proteins in plants are shown in Table 3. The performance of the individual tools is in the order of SignalP > WolfPsort > TargetP > Phobius. When individual tools are used, none of them had an accuracy of > 60%. Phobius has the lowest accuracy and thus was not suitable for predicting plant secreted proteins. WolfPsort had the highest specificity and the lowest sensitivity. When TMHMM was used in tandem after these individual tools, the accuracies were slightly improved in SignalP/TMHMM (7.6% increment in MCC), TargetP/TMHMM (9.8% increment in MCC), and WolfPsort/TMHMM (3.8%). The highest accuracy was achieved by combining SignalP, Phobius, TargetP, TMHMM and PS-Scan (Table3). This method thus is recommended for predicting secreted proteins in plants. Adding WolfPsort to the above pipeline reduced the accuracy as it significantly reduced the sensitivity and only slightly improved the specificity.

Prediction of protist data

Comparing with the number of curated secreted proteins in other eukaryotes, there are relatively much less data in protists. TargetP was trained for plants and non-plants, and we found it performed better using the non-plant parameter for secreted protein prediction in protists. WolfPsort was trained for fungi, animals, and plants. We found it performed best with fungal parameter for protist prediction. Prediction accuracies for secreted proteins in protists are shown in Table 4. Overall, all the programs performed poorly for secreted protein prediction in protists as all the MCC values are lower than 50%. However, based on the MCC values, the performance of the tools can still be assessed, and the order of the accuracies was Phobius > WolfPsort > SignalP > TargetP. SignalP/TMHMM performed best when used in tandem. The highest accuracy was achieved by combining SignalP, Phobius, TargetP, WolfPsort, TMHMM, and PS-Scan (Table3). It should be noted that even combining all these tools, the MCC value was only 52.8%, much lower than MCC values obtained for other eukaryotes. Clearly, more experimental data need to be collected for improving the prediction tools. Our evaluation was consistent with recent findings that reported SignalP failed to predict a significant portion of secreted proteins in *Plasmodium falciparum*, a protozoan parasite that causes malaria in humans (van Ooij et al., 2008).

Discussion

Among the tools including SignalP, Phobius, TargetP, and WolfPsort, evaluated in this work for secreted protein prediction, the predictive accuracy of all these tools was much lower than those reported in the original publications, as was also found by Klee and Ellis, (2005). However, this comparison study has determined which tool, among the widely used prediction tools, had the highest accuracy for prediction of secreted proteins in different kingdoms of eukaryotes; and based on these results, the optimal computational method of combing different tools for each kingdom of eukaryotes was proposed. The data shows that when a single prediction tool was used or combined with TMHMM, different tools have different strength in processing different kingdoms of organisms. For fungal prediction WolfPsort/TMHMM was most accurate, and also SignalP/TMHMM performed better than Phobius (Table 1), even though Phobius was originally designed to perform better than SignalP/TMHMM (Kall et al., 2004; Kall et al., 2007). However, Phobius had the highest accuracy for processing animal data, though it was only marginally better when other tools used in tandem with TMHMM (Table 2). For plant secreted protein prediction, SignalP/TMHMM had the highest accuracy and Phobius was the worst predictor (Table 3). Among the tools evaluated for protist data processing, no any single tool and even combined with THHMM had a MCC accuracy value higher than 50% (Table 4). This was primarily due to the limit of available secreted protein data in protists that prevented adequate training for these tools.

Combing multiple tools, however, often substantially improves the accuracy of secreted protein prediction (Chen et al., 2003; Klee and Ellis, 2005; Tsang et al., 2009). Adding SignalP/Phobius/PS-Scan to WolfPsort/TMHMM increased 7.9% in accuracy in fungi, adding WolfPsort/PS-Scan to Phobius increased 3.9% in accuracy in animals, adding Phobius/TargetP/PS-Scan to SignalP/TMHMM increased 10.2% in plants (Table 1-3). Even in protists, when all the tools combined, we see 8.7% improvement in accuracy over SignalP/TMHMM method (Table 4). However, adding more prediction tools without discretion may decrease the prediction accuracy. In this work we found that adding TargetP in the method for fungi or WolfPsort in the prediction methods for plants reduced the prediction accuracy (Table 1, Table 3).

In summary, among the tools evaluated in this work including SignalP, Phobius, TargetP, and WolfPsort, there was no single individual tool suitable for accurately predicting secreted proteins in all eukaryotes. If a single tool is used, WolfPsort is recommended for processing fungal data, Phobius for processing animal data, SignalP for processing plant data. TMHMM used in tandem with SignalP or WolfPsort significantly improved the prediction accuracies in all data sets. PS-Scan was useful to remove ER target proteins. More experimental data need to be collected for protists. The accuracy, particularly the specificity, however, is much improved when the two or more tools are differentially combined for predicting secreted proteins in different eukaryotes.

	TP	FP	TN	FN	Sn (%)	Sp (%)	MCC (%)
SignalP	199	364	7164	17	92.1	95.2	55.4
Phobius	188	638	6890	28	87.0	91.5	41.9
TargetP	198	442	7086	18	91.7	94.1	51.3
WolfPsort	108	70	7458	108	50.0	99.1	53.9
SignalP/TMHMM	197	237	7291	19	91.2	96.9	63.0
Phobius/TMHMM	188	636	6892	28	87.0	91.6	42.0
TargetP/TMHMM	195	256	7272	21	90.3	96.6	61.1
WolfPsort/TMHMM	106	45	7483	110	49.1	99.4	57.7
SignalP/TMHMM/TargetP	195	149	7379	21	90.3	98.0	70.6
SignalP/TMHMM/TargetP/PS-Scan	195	134	7394	21	90.3	98.2	72.3
Phobius/TargetP/TMHMM	183	122	7406	33	84.7	98.4	70.4
SignalP/TMHMM/WolfPsort	106	35	7493	110	49.1	99.5	59.9
SignalP/TMHMM/Phobius	188	183	7345	28	87.0	97.6	65.2
SignalP/HMM/Phobius/TargetP	183	113	7415	33	84.7	98.5	71.5
SignalP/TMHMM/Phobius/TargetP/PS-Scan	**183**	**100**	**7428**	**33**	**84.7**	**98.7**	**73.2**
SignalP/TMHMM/Phobius/TargetP/WolfPsort/PS-Scan	102	29	7499	114	47.2	99.6	59.8

TP: true positives; FP: false positives; TN: true negatives; FN: false negatives. Sn: sensitivity; Sp:specificity; MCC: Mathews' correlation coefficient.

Table 3: Prediction accuracies of secreted proteins in plants.

	TP	FP	TN	FN	Sn (%)	Sp (%)	MCC (%)
SignalP	30	154	1825	2	93.8	92.2	37.3
Phobius	30	118	1861	2	93.8	94.0	42.1
TargetP	31	259	1720	1	96.9	86.9	29.8
WolfPsort	28	118	1861	4	87.5	94.0	39.3
SignalP/TMHMM	29	98	1881	3	90.6	95.0	44.1
Phobius/TMHMM	30	113	1866	2	93.8	94.3	42.9
TargetP/TMHMM	30	143	1836	2	93.8	92.8	38.6
WolfPsort/TMHMM	27	99	1880	5	84.4	95.0	41.0
SignalP/TMHMM/Phobius	29	85	1894	3	90.6	95.7	46.7
SignalP/TMHMM/Phobius/TargetP	29	70	1909	3	90.6	96.5	50.4
SignalP/TMHMM/Phobius/TargetP/WolfPsort	26	48	1931	6	81.3	97.6	52.4
SignalP/TMHMM/Phobius/TargetP/WolfPsort/PS-Scan	**26**	**47**	**1932**	**6**	**81.3**	**97.6**	**52.8**

TP: true positives; FP: false positives; TN: true negatives; FN: false negatives. Sn: sensitivity; Sp:specificity; MCC: Mathews' correlation coefficient.

Table 4: Prediction accuracies of secreted proteins in protists.

Citation: Min XJ (2010) Evaluation of Computational Methods for Secreted Protein Prediction in Different Eukaryotes. J Proteomics Bioinform 3: 143-147. doi:10.4172/jpb.1000133

Acknowledgements

We thank Gary Walker and Chest Cooper for their helpful comments on the paper. This work was supported by the Youngstown State University Research Council grant (2009-10 #04-10), Research Professorship, and the STEM Dean's reassigned time for research.

References

1. Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16: 412-424. » CrossRef » PubMed » Google Scholar
2. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783-795. » CrossRef » PubMed » Google Scholar
3. Blobel G, Dobberstein B (1975) Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J Cell Biol 67: 835-851. » CrossRef » PubMed » Google Scholar
4. Chen Y, Yu P, Luo J, Jiang Y (2003) Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. Mamm Genome 14: 859-865. » CrossRef » PubMed » Google Scholar
5. Chen Y, Zhang Y, Yin Y, Gao G, Li S, et al. (2005) SPD - a web-based secreted protein database. Nucleic Acids Res. 33: D169-173. » CrossRef » PubMed » Google Scholar
6. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, et al. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 34: W362-365. » CrossRef » PubMed » Google Scholar
7. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2: 953-971. » CrossRef » PubMed » Google Scholar
8. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300: 1005-1016. » CrossRef » PubMed » Google Scholar
9. Harcus YM, Parkinson J, Fernandez C, Daub J, Selkirk ME, et al. (2004) Signal sequence analysis of expressed sequence tags from the nematode Nippostrongylus brasiliensis and the evolution of secreted proteins in parasites. Genome Biol 5: R39. » CrossRef » PubMed » Google Scholar
10. Hathout Y (2007) Approaches to the study of the cell secretome. Expert Rev Proteomics 4: 239-248. » CrossRef » PubMed » Google Scholar
11. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, et al. (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35: W585-587. » CrossRef » PubMed » Google Scholar
12. Jackson DJ, McDougall C, Green K, Simpson F, Wörheide G, et al. (2006) A rapidly evolving secretome builds and patterns a sea shell. BMC Biol 4: 40. » CrossRef » PubMed » Google Scholar
13. Jamet E, Albenne C, Boudart G, Irshad M, Canut H, et al. (2008) Recent advances in plant cell wall proteomics. Proteomics 8: 893-908. » CrossRef » PubMed » Google Scholar
14. Käll L, Krogh A, Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338: 1027-1036. » CrossRef » PubMed » Google Scholar
15. Käll L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. Nucleic Acids Res 35: W429-432. » CrossRef » PubMed » Google Scholar
16. Klee EW, Carlson DF, Fahrenkrug SC, Ekker SC, Ellis LB (2004) Identifying secretomes in people, pufferfish and pigs. Nucleic Acids Res 32: 1414-1421. » CrossRef » PubMed » Google Scholar
17. Klee EW, Ellis LB (2005) Evaluating eukaryotic secreted protein prediction. BMC Bioinformatics 6: 256. » CrossRef » PubMed » Google Scholar
18. Klee EW (2008) The zebrafish secretome. Zebrafish 5: 131-138. » CrossRef » PubMed » Google Scholar
19. Lee SA, Wormsley S, Kamoun S, Lee AF, Joiner K, et al. (2003) An analysis of the Candida albicans genome database for soluble secreted proteins using computer-based prediction algorithms. Yeast 20: 595-610. » CrossRef » PubMed » Google Scholar
20. Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta 405: 442-451. » CrossRef » PubMed » Google Scholar
21. Menne KM, Hermjakob H, Apweiler R (2000) A comparison of signal sequence prediction methods using a test set of signal peptides. Bioinformatics 16: 741-742. » CrossRef » PubMed » Google Scholar
22. O'Toole N, Min XJ, Storms R, Butler G, Tsang A (2006) Sequence-based analysis of fungal secretomes. In: Arora DK, Berka RM and Singh GB (eds) Applied Mycology Biotechnology: Bioinformatics 6: 277-296. » CrossRef » PubMed » Google Scholar
23. Scott M, Lu G, Hallett M, Thomas DY (2004) The Hera database and its use in the characterization of endoplasmic reticulum proteins. Bioinformatics 20: 937-944. » CrossRef » PubMed » Google Scholar
24. Sprenger J, Fink JL, Teasdale RD (2006) Evaluation and comparison of mammalian subcellular localization prediction methods. BMC Bioinformatics 7: S3. » CrossRef » PubMed » Google Scholar
25. Tsang A, Butler G, Powlowski J, Panisko EA, Baker SE (2009) Analytical and computational approaches to define the Aspergillus niger secretome. Fungal Genetics Biol 46: S153-S160. » CrossRef » PubMed » Google Scholar
26. van Ooij C, Tamez P, Bhattacharjee S, Hiller NL, Harrison T, et al. (2008) The malaria secretome: from algorithms to essential function in blood stage infection. PLoS Pathog 4: e1000084. » CrossRef » PubMed » Google Scholar
27. von Heijne G (1990) The signal peptide. J Membr Biol 115: 195-201.
28. Wymelenberg AV, Sabat G, Martinez D, Rajangam AS, Teeri TT, et al. (2005) The Phanerochaete chrysosporium secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. J Biotechnol 118: 17-34. » CrossRef » PubMed » Google Scholar
29. Xue H, Lu B, Lai M (2008) The cancer secretome: a reservoir of biomarkers. J Transl Med 6: 52. » CrossRef » PubMed » Google Scholar