Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Databases and ontologies

Original Paper

PerMemDB: a database for eukaryotic peripheral membrane proteins

Katerina C. Nastou, Georgios N. Tsaousis, Stavros J. Hamodrakas and Vassiliki A. Iconomidou*

Section of Cell Biology and Biophysics, Department of Biology, National and Kapodistrian University of Athens, Panepistimiopolis, Athens 15701, Greece

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

The majority of all proteins in cells interact with membranes either permanently or temporarily. Peripheral membrane proteins form transient complexes with membrane proteins and/or lipids, via non-covalent interactions and are of outmost importance, due to numerous cellular functions in which they participate. In an effort to collect data regarding this heterogeneous group of proteins we designed and constructed a database, called PerMemDB. PerMemDB is currently the most complete and comprehensive repository of data for eukaryotic peripheral membrane proteins deposited in UniProt or predicted with the use of MBPpred – a computational method that specializes in the detection of proteins that interact non-covalently with membrane lipids, via membrane binding domains. The first version of the database contains 241173 peripheral membrane proteins from 1216 organisms. All entries have cross-references to other databases, literature references and annotation regarding their interactions with other proteins. Moreover, additional sequence annotation of the characteristic domains that allow these proteins to interact with membranes is available, due to the application of MBPpred. Through the web interface of PerMemDB, users can browse the contents of the database, submit advanced text searches and BLAST queries against the protein sequences deposited in PerMemDB. We expect this repository to serve as a source of information for the development of prediction algorithms regarding peripheral membrane proteins, in addition to proteome-wide analyses.

Availability: http://bioinformatics.biol.uoa.gr/permemdb
Contact: veconom@biol.uoa.gr
Supplementary information: http://83.212.109.111:8085/assets/Nastou_Supplement.xlsx

1 Introduction

One universal feature of all cells, upon which their structure and majority of functions rely, are membranes (Alberts, 2002). Membranes serve as permeable barriers for the entire cell or certain subcellular structures and are associated with the majority of cellular proteins (Stahelin, 2009). Signal transduction, molecular and ion transport, enzymatic activity and cell adhesion are among the most important functions performed by membrane proteins (Almen *et al.*, 2009; von Heijne, 2007). Membrane proteins can be classified into two broad categories based on the nature of membrane-protein interactions; integral and peripheral membrane proteins (Alberts, 2002; Escriba *et al.*, 2008).

Peripheral membrane proteins interact indirectly with membrane proteins or lipids via non-covalent interactions (Goni, 2002; Johnson and Cornell, 1999; Seaton and Roberts, 1996) and are critical due to the numerous cellular functions in which they participate (Lomize *et al.*, 2007; Smith, 2012). Peripheral proteins also possess domains that permit the specific or non-specific interaction with membrane lipids, to perform functions related to signal transduction and membrane trafficking (Cho and Stahelin, 2005; Hurley, 2006). These domains allow the identification and classification of these proteins (Moravcevic *et al.*, 2012) and have been exploited for the development of three computa-

tional methods for the detection of peripheral membrane proteins in proteomes (Bhardwaj *et al.*, 2010; Bhardwaj *et al.*, 2006; Nastou *et al.*, 2016). Among these three methods, MBPpred has the most extended library of pHMMs and detects proteins that possess 18 domains with experimentally validated interactions with membrane lipids (Nastou *et al.*, 2016).

To this day, two databases have been developed that contain data for specific subgroups of peripheral membrane proteins. The first one is OPM (Lomize *et al.*, 2006), a database dedicated to the optimization of the spatial arrangement of protein structures in lipid bilayers and contains data for a substantial number of peripheral proteins derived from PDB (Berman *et al.*, 2000). The other one, MeTaDoR (Bhardwaj *et al.*, 2007), is a decade-old online resource devoted specifically to peripheral proteins with membrane targeting domains and includes structural and sequence data. However, OPM contains only a collection of structural data on peripheral membrane proteins and MetaDoR's online interface has ceased functioning since 2014. At present there is no special-purpose biological database for peripheral membrane proteins available. This fact urges the need for a thorough data collection and more rigorous studies regarding this protein group.

In this study we addressed this need through the construction of PerMemDB, a database for peripheral membrane proteins in eukaryotes. This repository currently holds data on peripheral membrane proteins, deposited in UniProt (UniProt_Consortium, 2017) or predicted with the use of MBPpred (Nastou *et al.*, 2016) for all eukaryotic reference proteomes.

2 Methods

The development of PerMemDB, was based on three data collection approaches. Firstly, the available proteins from UniProt (UniProt_Consortium, 2017) with subcellular location "Peripheral membrane protein [SL-9903]" were isolated for all eukaryotic reference proteomes via programmatic access to the UniProt API (Nightingale *et al.*, 2017). Then, peripheral proteins with Membrane Binding Domains (MBDs) that interact directly with membrane lipids were retrieved, using MBPpred (Nastou *et al.*, 2016), a prediction method developed in our lab, that identifies these proteins from their sequence via a library of profile HMMs, specific for 18 MBDs. For this purpose, FASTA formatted files (Lipman and Pearson, 1985) for all eukaryotic reference proteomes were downloaded automatically from UniProt and used as input files for the stand-alone local version of MBPpred. Finally, in order to collect peripheral membrane proteins that interact indirectly with the membrane, all non-transmembrane interactors of transmembrane proteins (subcellular location: "Multi-pass membrane protein [SL-9909]" or "Single-pass membrane protein [SL-9904]" or "Single-pass type I membrane protein [SL-9905]" or "Single-pass type II membrane protein [SL-9906]" or "Single-pass type III membrane protein [SL-9907]" or "Single-pass type IV membrane protein [SL-9908]") for eukaryotic reference proteomes were collected programmatically and annotated, also, as peripheral. Fig. 1 shows the pipeline for the retrieval of all protein datasets. Seven final subsets of unique peripheral proteins from these three sources were created and stored in PerMemDB (Table 1).

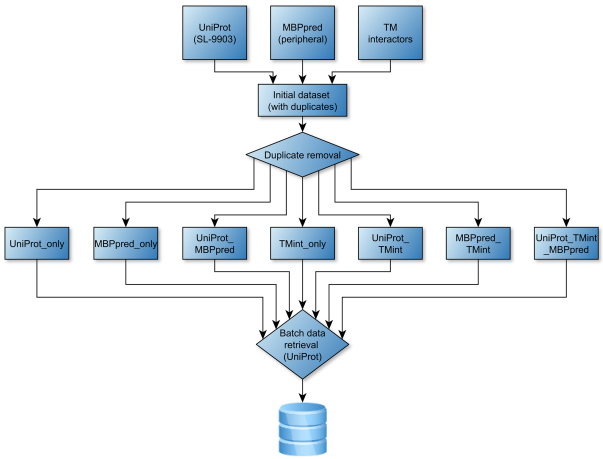


Fig. 1. The data retrieval pipeline. Proteins with subcellular location "Peripheral membrane protein" were retrieved from UniProt, "peripheral proteins" with MBDs were identified with the use of MBPpred and non-transmembrane interaction partners of transmembrane (TM) proteins were isolated from UniProt. Scripts were written in Perl and Python for the automated retrieval of all entries with the required information and of all fasta files that were used as input for MBPpred. An initial dataset of protein entries was created after the merge of all aforementioned lists. After a comparison of the three lists, seven non-overlapping datasets were created, that would provide the final set of proteins to be stored in the database. The contents of each dataset are described in Table 1. Python scripts were written to recover data from UniProt for all protein entries and were further manipulated in order to be stored in a relational database.

Table 1. The subsets of peripheral membrane proteins stored in PerMemDB, grouped by the source(s) from which they were isolated. A description of each subset is given in the last column.

Set	Name	Description
1	UniProt_only	Entries found only in UniProt, with SL-9903
2	MBPpred_only	Peripheral membrane proteins retrieved only with the use of MBPpred
3	TMint_only	Only non-transmembrane interactors of transmembrane (TM) proteins
4	UniProt_MBPpred	Entries with SL-9903 which were also retrieved with the use of MBPpred but were not found to interact with TM proteins
5	UniProt_TMint	Entries with SL-9903 which were also interactors of TM proteins, but were not detected with MBPpred
6	MBPpred_TMint	Peripheral membrane proteins retrieved with the use of MBPpred which were also interactors of TM proteins, but were not reported as peripheral in UniProt
7	UniProt_MBPpred_TMint	Entries with SL-9903, retrieved with MBPpred and found to interact with TM proteins

A web application for PerMemDB has been developed. A mySQL database system was used to store all protein data in a relational database and serves as the first layer of the application. The second layer is a Node.js application server that receives user queries to the database and returns data to the web browser. The web interface is based on modern technologies (HTML5, CSS3 and Javascript) and can be viewed from any screen size (desktop, tablet or mobile).

For each entry, basic information about the respective protein are provided, including the source type (Table 1), in addition to cross-references to major publicly available databases for diseases and drugs (DrugBank (Wishart *et al.*, 2018; Wishart *et al.*, 2006), OMIM (Hamosh *et al.*, 2005), Orphanet (Pavan *et al.*, 2017)), 3D structures (PDB (Berman *et al.*, 2000)), protein families (Pfam (Finn *et al.*, 2016), PROSITE (Sigrist *et al.*, 2013)), genes (EMBL (Stoesser *et al.*, 1999), HGNC (Yates *et al.*, 2017)), pathways (KEGG (Kanehisa and Goto, 2000), Reactome (Fabregat *et al.*, 2018)), interactions (IntAct (Orchard *et al.*, 2014), BioGrid (Stark *et al.*, 2006)) and proteins (UniProt (UniProt_Consortium, 2017), RaftProt (Shah *et al.*, 2015)). Moreover, a CytoscapeJS (Franz *et al.*, 2016) viewer is integrated for the visualization of the interactions between peripheral membrane proteins and their interaction partners (when information is available in UniProt).

3 Results

We have constructed PerMemDB, a relational protein database, which, in version 1.2 (Dec 2018), contains 241173 proteins originating from 1216 eukaryotes. The database can be either searched or browsed by organism. Each record contains sequence information and cross-references to many publicly available databases, with data spanning from protein family annotation to disease. Moreover, when available, information about the interaction partners of each peripheral protein in the database was retrieved from UniProt and is shown in an interaction network, that allows the quick identification of functional modules centering peripheral membrane proteins. There are 43226 entries isolated from UniProt (with subcellular location "Peripheral membrane protein"), 197961 were identified using MBPpred and 2304 were designated as indirectly interacting peripheral membrane proteins (TM interactors).

3.1 User interface and website features

The PerMemDB database has a user-friendly interface that offers convenient ways to gain access to its data. From the navigation bar at the top of every page, users can either perform searches or browse the database contents. Searching can be performed using various search terms (e.g. protein or gene name) and results may be refined by source type (UniProt, MBPpred or Transmembrane Interactor) or status (reviewed or unreviewed). While browsing PerMemDB, a user can have access to all entries for a specific eukaryotic reference proteome (Fig. 2). Results can be further filtered using the "Search" field at the top right of the page. When the green "Show selected Entries" button is pressed, the user is transferred to a new page with data regarding the proteome of the organism they have selected. Results, retrieved from either 'browsing by organism' or 'searching' the database, are displayed in tables. At the end of each row direct links to protein entry pages are given, which the user can follow by clicking on the respective buttons. Moreover, a BLAST (Altschul *et al.*, 1990) search tool is incorporated for running BLAST searches against the database proteins, using one or more FASTA formatted sequences as input.

Apart from unique entries, the entire database is available for download by pressing the 'Download' button at the top navigation bar. PerMemDB is currently available in four formats (text, XML, JSON and FASTA). Finally, a 'Home' page for a short description and database statistics, a 'Manual' page explaining the functionalities of PerMemDB and a 'Contact' page with author contact information and a submission form to retrieve data from the scientific community, are also available.

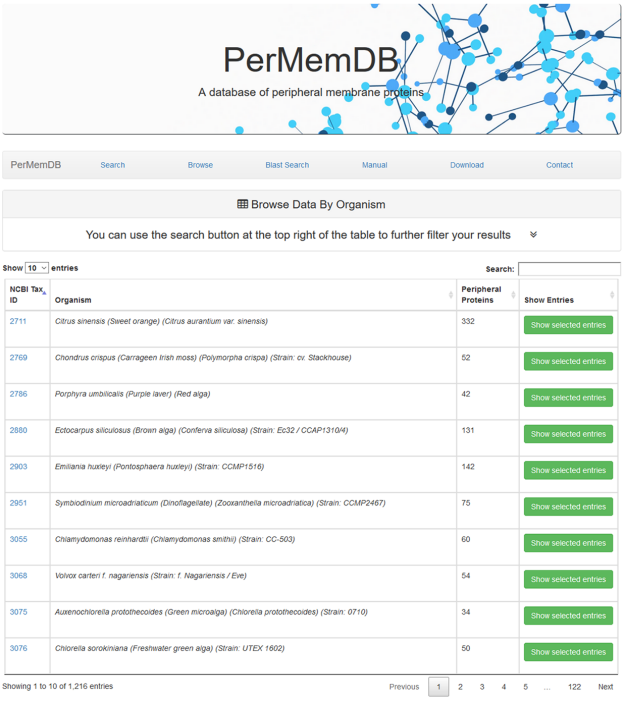


Fig. 2. The 'Browse' Page of PerMemDB. Users can browse data stored in PerMemDB for a specific organism, whose proteome is listed as a eukaryotic reference proteome in UniProt. Non-specific searches can be performed using the 'Search' option at the top right corner of the data table. If a user presses the green 'Show Selected Entries' button all proteins from the selected organism are shown in a new page.

Taking into consideration the biological and clinical significance of peripheral membrane proteins, our intention is to accurately represent all available information for this protein group through our repository. However, when the information source is all eukaryotic reference proteomes, this task can be challenging and some data may be falsely filtered out during the necessary automated retrieval process. In an effort to stay updated and be as comprehensive as possible, PerMemDB implements a user annotation feature. More specifically, in the contact page of our database, a form has been created dedicated explicitly to the submission of comments or data that has not been collected during the creation of the resource. Interaction with the users is of outmost importance to render this repository a useful tool for the scientific community. This process will allow us to incorporate valuable information regarding the sequence, the domains and the interaction networks of these proteins and thus, better annotate our entries and potentially improve our data retrieval protocol. It is our goal to implement all information gathered through this process in each database update, in addition to all data gathered using our automated pipeline.

3.2 Entry Pages

Database entries are generated dynamically via browsing, searching or through direct URL links. As shown in Fig. 3, on the top of each page, data are available for download in four formats (text, XML, JSON and FASTA). Tables displaying basic information (e.g. protein name) and additional information (e.g. sequence) about each entry are shown on the left of each page. On the right, CytoscapeJS (Franz *et al.*, 2016) is used to visualize the relationships between peripheral membrane proteins and their interaction partners. Links to RaftProt (Shah *et al.*, 2015) are available for proteins with experimental evidence regarding their presence on lipid rafts, membrane substructures that compartmentalize cellular processes (Korade and Kenworthy, 2008) and especially signal transduction processes. On the bottom part of each page direct links to several external databases are also provided.

Information regarding the position and sequence of Membrane Binding Domains (MBDs) for proteins retrieved with the use of MBPpred is provided in the "Source Type" field of each protein entry. Moreover, in the same field, links to UniProt are given for transmembrane proteins designated as interaction partners of indirectly interacting peripheral membrane proteins (Fig. 3).

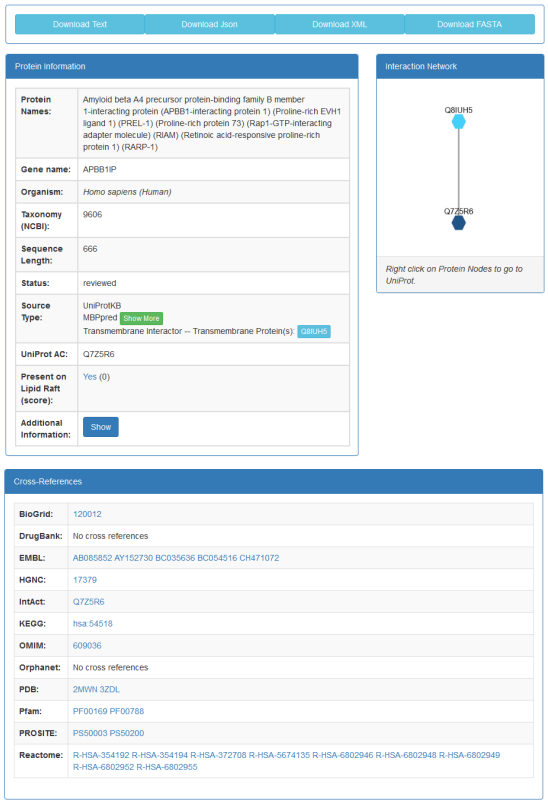


Fig. 3. Detailed view of a protein entry of PerMemDB. The user can view basic information about the protein through the "Protein Information" panel. Additional information is provided by pressing the blue "Show" button. For entries retrieved through MBPpred information regarding the Membrane Binding Domains (MBDs) of the proteins are provided by pressing the green "Show More" button in the "Source Type" field. For peripheral proteins that have been identified as "non-transmembrane interactors of transmembrane proteins", links to UniProt are provided for their transmembrane interaction partners (light blue button with UniProt AC). On the right, the relationships between peripheral membrane proteins and their interaction partners are shown when available. Upon clicking on the links at the bottom of the page users are transferred to the pages of the respective cross-references. All data can be downloaded in text, JSON, XML and FASTA format from the top of each entry page.

3.3 Analysis of Database Data

3.3.1 Quantification Analysis

In an effort to take a deeper look at the data stored in PerMemDB we performed a quantification analysis based on source type (Supplementary Table 1, http://83.212.109.111:8085/assets/Nastou_Supplement.xlsx). At first glance, it is evident that the majority of data stored in the database were derived from the prediction method MBPpred (ca. 80%). Data originating solely from UniProt account for ca. 17%, while a very small amount of data were designated as transmembrane interactors. Moreover, there is little cross-over between these three data sources. In particular, ~1% of peripheral membrane proteins are found both in UniProt and by running MBPpred, and only 76 proteins were retrieved from all three sources. Actually, this was not unexpected, since peripheral membrane proteins are generally understudied as a group, in comparison to other membrane protein groups. In addition, even though extensive efforts for the annotation of this diverse group of proteins in human and specific model organisms have been carried out, things are extremely different for all other eukaryotes (Supplementary Table 1, Fig. 4). Specifically, in the majority of eukaryotic reference proteomes in PerMemDB the number of predicted peripheral membrane proteins is more than tenfold that of the annotated proteins from UniProt. This underlines the fact that compared to generalized databases (like UniProt), PerMemDB presents a more complete coverage of available representatives in all eukaryotic reference proteomes. Thus, PerMemDB could be a useful resource for the computational analysis and the clarification of the functional nature of this protein group.

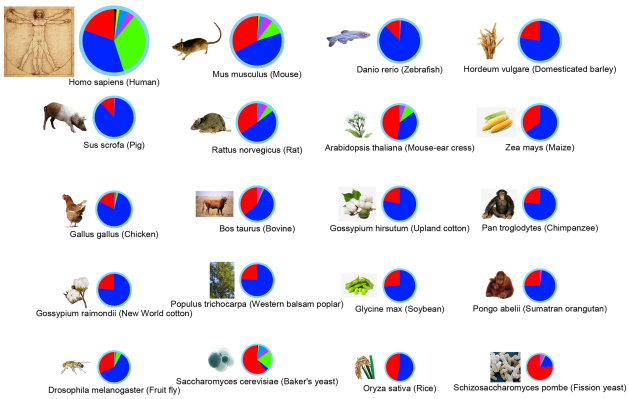


Fig. 4. Twenty proteomes with the most representatives in UniProt. Pie charts are used to depict the number of peripheral membrane proteins for each proteome. The size of each pie chart is analogous to the total number of peripheral proteins detected in it. Data are color coded based on the subset in which the peripheral proteins belong to (See Table 1). Blue: UniProt_only, Red: MBPpred_only, Green: TMint_only, Magenta: UniProt_MBPpred, Light_blue: UniProt_TMint, Yellow: UniProt_MBPpred, Gray: UniProt_MBPpred_TMint. This image was created with the use of Cytoscape 3.7.0 (Shannon *et al.*, 2003).

In Fig. 4 pie charts are used to depict the distribution of peripheral membrane proteins, based on data source, for the twenty proteomes with

the most representative proteins in UniProt. The size of each pie chart corresponds to the total number of peripheral proteins, while the different slices represent the categories shown in Table 1. It is evident by the overrepresentation of blue and red colors in these charts that the data were retrieved mostly either with the use of MBPpred or from UniProt "Subcellular Location". Only very well-annotated proteomes – human, mouse, rat, mouse-ear cress and baker's yeast – show diversity in regard to the source of peripheral membrane proteins

A quantification analysis was also performed to examine the distribution of Membrane Binding Domains (MBDs) for all proteins stored in PerMemDB, since, as mentioned above, ~80% of data originate from MBPPred. The two most abundant domains (that account for more than 40% of domains present in peripheral membrane proteins) are the Pleckstrin Homology (PH) domain and the C2 domain (Fig. 5). Both domains are found in proteins that are either cytoskeleton constituents (Musacchio *et al.*, 1993) or involved in cell signaling and enzymatic activities (Ingley and Hemmings, 1994; Zhang and Aravind, 2010), biological processes mainly performed by peripheral membrane proteins in cells (Lemmon, 2008). Recently characterized MBDs, like KA1 and Golph 3, are detected only in a small number of peripheral membrane proteins (Fig. 5).

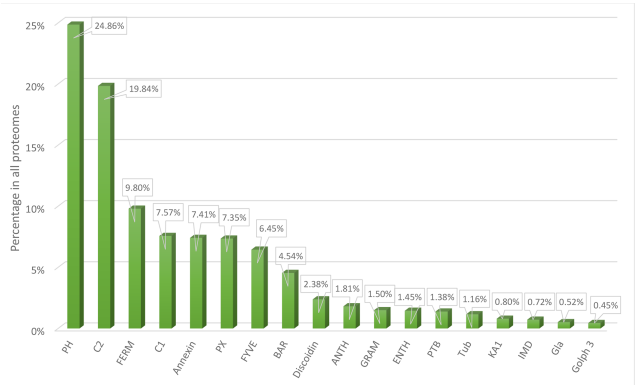


Fig. 5. The distribution of Membrane Binding Domains (MBDs) in all eukaryotic reference proteomes. The domains with the highest prevalence in peripheral membrane proteins are the well-studied PH and C2 domains, while recently identified domains like KA1 and Golph 3 are present in very small numbers

3.3.2 Functional enrichment analysis of ten selected proteomes

In an effort to investigate the functional roles of this diverse group of proteins, we analyzed the available data regarding the functions of these proteins for ten (10) selected proteomes (*Arabidopsis thaliana, Drosophila melanogaster, Danio rerio, Homo sapiens, Mus musculus, Rattus norvegicus, Saccharomyces cerevisiae, Gallus gallus, Bos taurus* and *Sus scrofa*). The functional enrichment tool incorporated in the Cytoscape stringApp (Doncheva *et al.*, 2018) was used for this analysis. Detailed results are available in Supplementary Tables 2-11 (http://83.212.109.111:8085/assets/Nastou_Supplement.xlsx).

GO term enrichment analysis (Ashburner *et al.*, 2000; The Gene Ontology, 2017) showed that proteins from these ten proteomes are located on membranes, vesicles or are involved in cytoskeleton organization, compartments where peripheral membrane proteins would be expected to be localized. Regarding their functions, they take part in catalytic activities and act mostly as kinases, which explains their tendency to participate in signal transduction processes. Moreover, considering their localization, it is only logical that these proteins participate in cell communication and vesicle-mediated transport. Thus, it is not surprising that peripheral membrane proteins are involved in many signaling pathways and endocytosis as indicated by the functional enrichment analysis of KEGG Pathways (Kanehisa and Goto, 2000). Finally, an enrichment analysis against data deposited in InterPro (Mitchell *et al.*, 2018) and Pfam (Finn *et al.*, 2016) revealed that these proteins contain, apart from known MBDs, other domains like PDZ (Lee and Zheng, 2010), SH3, SH2 (Buday, 1999) and RhoGAP (Amin *et al.*, 2016), which have all been observed to be present in peripheral membrane proteins, repeatedly.

3.3.3 Analysis of pathogenic mutations on human peripheral membrane proteins

PerMemDB contains peripheral membrane protein data collected from different perspectives. Many inter- and intra-species applications that can contribute towards the better understanding of this group can be implemented. Considering the clinical significance of many peripheral membrane proteins (Leth-Larsen *et al.*, 2010; Lukiw, 2013), we present here an application of PerMemDB for the study of the association between MBDs and disease variants in the subset of human proteins.

For this analysis, data on human genetic variations ("Polymorphisms" and "Disease" variants) were gathered from the UniProt database (release date: 16-01-2019). These variants were mapped on the sequences of human peripheral membrane proteins isolated from PerMemDB. A chi-square test was performed to get an estimate of differences in the emergence of "Polymorphisms" and "Disease" variants on regions that either do or do not contain MBDs.

The full dataset of missense Single Nucleotide Polymorphisms (SNPs) located on 341 out of the 3523 human peripheral membrane proteins consists of 2498 unique SNPs – 1471 "Disease" SNPs and 1027 "Polymorphisms". From those, 434 SNPs are located on MBDs and 2064 on other regions. Considering the fact, that these regions differ vastly in their length, all data had to be subjected to normalization, based on the length of each region.

Normalized data were subjected to chi-square testing to estimate the statistical difference between the frequency of "Disease" SNPs on MBDs and on other regions in peripheral membrane proteins. Results from this analysis show that there is a statistically significant difference (*p-value=0.029 < 0.05*) between the expected and observed "Disease" mutations in MBDs (Supplementary Table 12, http://83.212.109.111:8085/assets/Nastou_Supplement.xlsx). These preliminary results indicate the importance of 'structural' protein information, like the topological profile of peripheral membrane proteins with MBDs, extracted from PerMemDB, in the evaluation of the implications of genetic variations in this protein group.

4 Conclusions

PerMemDB is currently the only repository that contains data dedicated exclusively to peripheral membrane proteins. The collection of data using three different methods allows a complete and extensive recording of all proteins that belong to this group. The existence of such a dataset can be very useful for large-scale analyses and for the training of prediction methods. As shown above, with the mutation analysis on human proteins, data deposited in PerMemDB can be valuable for disease association applications. Furthermore, taking into consideration the fact that PerMemDB contains data on a wide range of eukaryotes, it can serve as a valuable tool for evolutionary analyses, either for the entire group of peripheral membrane proteins or for membrane binding peripheral proteins with specific domains. The database will be available for download

for those who would like to access an updated and annotated dataset of peripheral membrane proteins for their research purposes, both in human readable text format and XML or JSON formats for programmatic access to the data.

Our goal is to keep the database up-to-date with biannual updates. Moreover, we aim to add novel membrane binding protein families to the MBPpred algorithm – if they are described in the scientific literature – which will allow for a more comprehensive resource, in the future. It is our hope that this database will aid the scientific community, towards gaining a profound understanding of this important group of proteins.

PerMemDB is available at http://bioinformatics.biol.uoa.gr/permemdb.

Acknowledgements

The authors thank the National and Kapodistrian University of Athens for support.

Conflict of Interest: none declared.

References

Alberts, B. (2002) *Molecular biology of the cell*. Garland Science, New York.

Almen, M.S. *et al.* (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin, *BMC Biology*, **7**, 50.

Altschul, S.F. *et al.* (1990) Basic local alignment search tool, *J Mol Biol*, **215**, 403-410.

Amin, E. *et al.* (2016) Deciphering the Molecular and Functional Basis of RHOGAP Family Proteins: A SYSTEMATIC APPROACH TOWARD SELECTIVE INACTIVATION OF RHO FAMILY PROTEINS, *J Biol Chem*, **291**, 20353-20371.

Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat Genet*, **25**, 25-29.

Berman, H.M. *et al.* (2000) The Protein Data Bank, *Nucleic Acids Res*, **28**, 235-242.

Bhardwaj, N., Gerstein, M. and Lu, H. (2010) Genome-wide sequence-based prediction of peripheral proteins using a novel semi-supervised learning technique, *BMC Bioinformatics*, **11 Suppl 1**, S6.

Bhardwaj, N. *et al.* (2006) Structural bioinformatics prediction of membrane-binding proteins, *J Mol Biol*, **359**, 486-495.

Bhardwaj, N. *et al.* (2007) MeTaDoR: a comprehensive resource for membrane targeting domains and their host proteins, *Bioinformatics*, **23**, 3110-3112.

Buday, L. (1999) Membrane-targeting of signalling molecules by SH2/SH3 domain-containing adaptor proteins, *Biochim Biophys Acta*, **1422**, 187-204.

Cho, W. and Stahelin, R.V. (2005) Membrane-protein interactions in cell signaling and membrane trafficking, *Annu Rev Biophys Biomol Struct*, **34**, 119-151.

Doncheva, N.T. *et al.* (2018) Cytoscape stringApp: Network analysis and visualization of proteomics data, *J Proteome Res*.

Escriba, P.V. *et al.* (2008) Membranes: a meeting point for lipids, proteins and therapies, *J Cell Mol Med*, **12**, 829-875.

Fabregat, A. *et al.* (2018) The Reactome Pathway Knowledgebase, *Nucleic Acids Res*, **46**, D649-D655.

Finn, R.D. *et al.* (2016) The Pfam protein families database: towards a more sustainable future, *Nucleic Acids Res*, **44**, D279-285.

Franz, M. *et al.* (2016) Cytoscape.js: a graph theory library for visualisation and analysis, *Bioinformatics*, **32**, 309-311.

Goni, F.M. (2002) Non-permanent proteins in membranes: when proteins come as visitors (Review), *Mol Membr Biol*, **19**, 237-245.

Hamosh, A. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders, *Nucleic Acids Res*, **33**, D514-517.

Hurley, J.H. (2006) Membrane binding domains, *Biochimica et Biophysica Acta: Protein Structure and Molecular Enzymology*, **1761**, 805-811.

Ingley, E. and Hemmings, B.A. (1994) Pleckstrin homology (PH) domains in signal transduction, *J Cell Biochem*, **56**, 436-443.

Johnson, J.E. and Cornell, R.B. (1999) Amphitropic proteins: regulation by reversible membrane interactions (review), *Mol Membr Biol*, **16**, 217-235.

Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res*, **28**, 27-30.

Korade, Z. and Kenworthy, A.K. (2008) Lipid rafts, cholesterol, and the brain, *Neuropharmacology*, **55**, 1265-1273.

Lee, H.J. and Zheng, J.J. (2010) PDZ domains and their binding partners: structure, specificity, and modification, *Cell Commun Signal*, **8**, 8.

Lemmon, M.A. (2008) Membrane recognition by phospholipid-binding domains, *Nat Rev Mol Cell Biol*, **9**, 99-111.

Leth-Larsen, R., Lund, R.R. and Ditzel, H.J. (2010) Plasma membrane proteomics and its application in clinical cancer biomarker discovery, *Mol Cell Proteomics*, **9**, 1369-1382.

Lipman, D.J. and Pearson, W.R. (1985) Rapid and sensitive protein similarity searches, *Science*, **227**, 1435-1441.

Lomize, A.L. *et al.* (2007) The role of hydrophobic interactions in positioning of peripheral proteins in membranes, *BMC Structural Biology*, **7**, 44.

Lomize, M.A. *et al.* (2006) OPM: orientations of proteins in membranes database, *Bioinformatics*, **22**, 623-625.

Lukiw, W.J. (2013) Alzheimer's disease (AD) as a disorder of the plasma membrane, *Front Physiol*, **4**, 24.

Mitchell, A.L. *et al.* (2018) InterPro in 2019: improving coverage, classification and access to protein sequence annotations, *Nucleic Acids Res*.

Moravcevic, K., Oxley, C.L. and Lemmon, M.A. (2012) Conditional peripheral membrane proteins: facing up to limited specificity, *Structure*, **20**, 15-27.

Musacchio, A. *et al.* (1993) The PH domain: a common piece in the structural patchwork of signalling proteins, *Trends Biochem Sci*, **18**, 343-348.

Nastou, K.C. *et al.* (2016) MBPpred: Proteome-wide detection of membrane lipid-binding proteins using profile Hidden Markov Models, *Biochim Biophys Acta*, **1864**, 747-754.

Nightingale, A. *et al.* (2017) The Proteins API: accessing key integrated protein and genome information, *Nucleic Acids Res*, **45**, W539-W544.

Orchard, S. *et al.* (2014) The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases, *Nucleic Acids Res*, **42**, D358-363.

Pavan, S. *et al.* (2017) Clinical Practice Guidelines for Rare Diseases: The Orphanet Database, *PLoS One*, **12**, e0170365.

Seaton, B.A. and Roberts, M.F. (1996) Peripheral Membrane Proteins. In Merz, K.M. and Roux, B. (eds), *Biological Membranes: A Molecular Perspective from Computation and Experiment*. Birkhäuser Boston, Boston, MA, pp. 355-403.

Shah, A. *et al.* (2015) RaftProt: mammalian lipid raft proteome database, *Nucleic Acids Res*, **43**, D335-338.

Shannon, P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res*, **13**, 2498-2504.

Sigrist, C.J. *et al.* (2013) New and continuing developments at PROSITE, *Nucleic Acids Res*, **41**, D344-347.

Smith, A.W. (2012) Lipid-protein interactions in biological membranes: a dynamic perspective, *Biochim Biophys Acta*, **1818**, 172-177.

Stahelin, R.V. (2009) Lipid binding domains: more than simple lipid effectors, *J Lipid Res*, **50 Suppl**, S299-304.

Stark, C. *et al.* (2006) BioGRID: a general repository for interaction datasets, *Nucleic Acids Res*, **34**, D535-539.

Stoesser, G. *et al.* (1999) The EMBL Nucleotide Sequence Database, *Nucleic Acids Res*, **27**, 18-24.

The Gene Ontology, C. (2017) Expansion of the Gene Ontology knowledgebase and resources, *Nucleic Acids Res*, **45**, D331-D338.

UniProt_Consortium (2017) UniProt: the universal protein knowledgebase, *Nucleic Acids Res*, **45**, D158-D169.

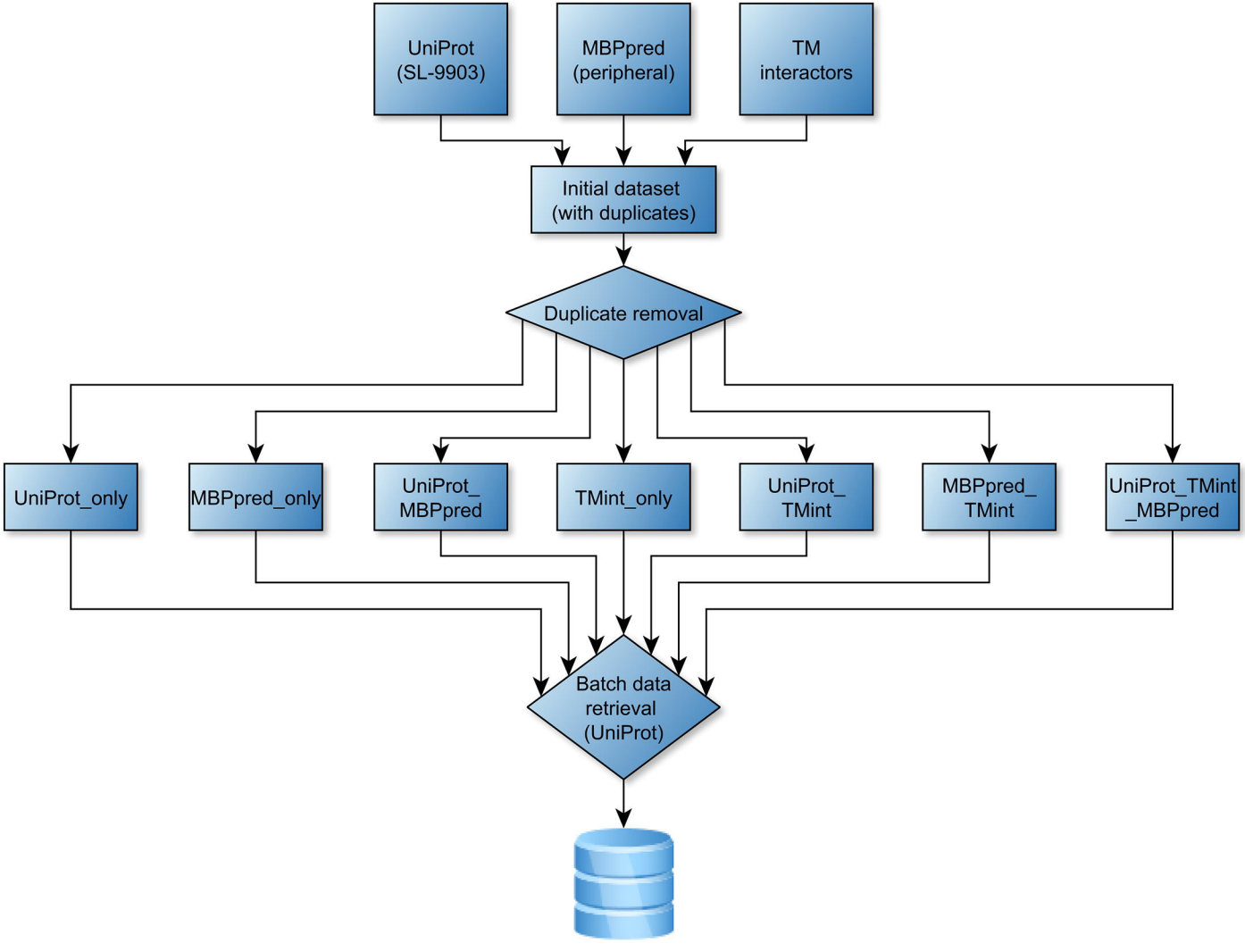
von Heijne, G. (2007) The membrane protein universe: what's out there and why bother?, *J Intern Med*, **261**, 543-557.

Wishart, D.S. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res*, **46**, D1074-D1082.

Wishart, D.S. *et al.* (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Res*, **34**, D668-672.

Yates, B. *et al.* (2017) Genenames.org: the HGNC and VGNC resources in 2017, *Nucleic Acids Res*, **45**, D619-D625.

Zhang, D. and Aravind, L. (2010) Identification of novel families and classification of the C2 domain superfamily elucidate the origin and evolution of membrane targeting activities in eukaryotes, *Gene*, **469**, 18-30.


UniProt
(SL-9903)
MBPpred
(peripheral)
TM
interactors
Initial dataset
(with duplicates)
Duplicate removal
UniProt_only
MBPpred_only
UniProt_
MBPpred
TMint_only
UniProt_
TMint
MBPpred_
TMint
UniProt_TMint
_MBPpred
Batch data
retrieval
(UniProt)

PerMemDB

A database of peripheral membrane proteins

PerMemDB | Search | Browse | Blast Search | Manual | Download | Contact

Browse Data By Organism

You can use the search button at the top right of the table to further filter your results

Show 10 entries

Search:

NCBI Tax ID	Organism	Peripheral Proteins	Show Entries
2711	*Citrus sinensis (Sweet orange) (Citrus aurantium var. sinensis)*	332	Show selected entries
2769	*Chondrus crispus (Carrageen Irish moss) (Polymorpha crispa) (Strain: cv. Stackhouse)*	52	Show selected entries
2786	*Porphyra umbilicalis (Purple laver) (Red alga)*	42	Show selected entries
2880	*Ectocarpus siliculosus (Brown alga) (Conferva siliculosa) (Strain: Ec32 / CCAP1310/4)*	131	Show selected entries
2903	*Emiliania huxleyi (Pontosphaera huxleyi) (Strain: CCMP1516)*	142	Show selected entries
2951	*Symbiodinium microadriaticum (Dinoflagellate) (Zooxanthella microadriatica) (Strain: CCMP2467)*	75	Show selected entries
3055	*Chlamydomonas reinhardtii (Chlamydomonas smithii) (Strain: CC-503)*	60	Show selected entries
3068	*Volvox carteri f. nagariensis (Strain: f. Nagariensis / Eve)*	54	Show selected entries
3075	*Auxenochlorella protothecoides (Green microalga) (Chlorella protothecoides) (Strain: 0710)*	34	Show selected entries
3076	*Chlorella sorokiniana (Freshwater green alga) (Strain: UTEX 1602)*	50	Show selected entries

Showing 1 to 10 of 1,216 entries

Previous 1 2 3 4 5 ... 122 Next

Download Text | Download Json | Download XML | Download FASTA

Protein Information

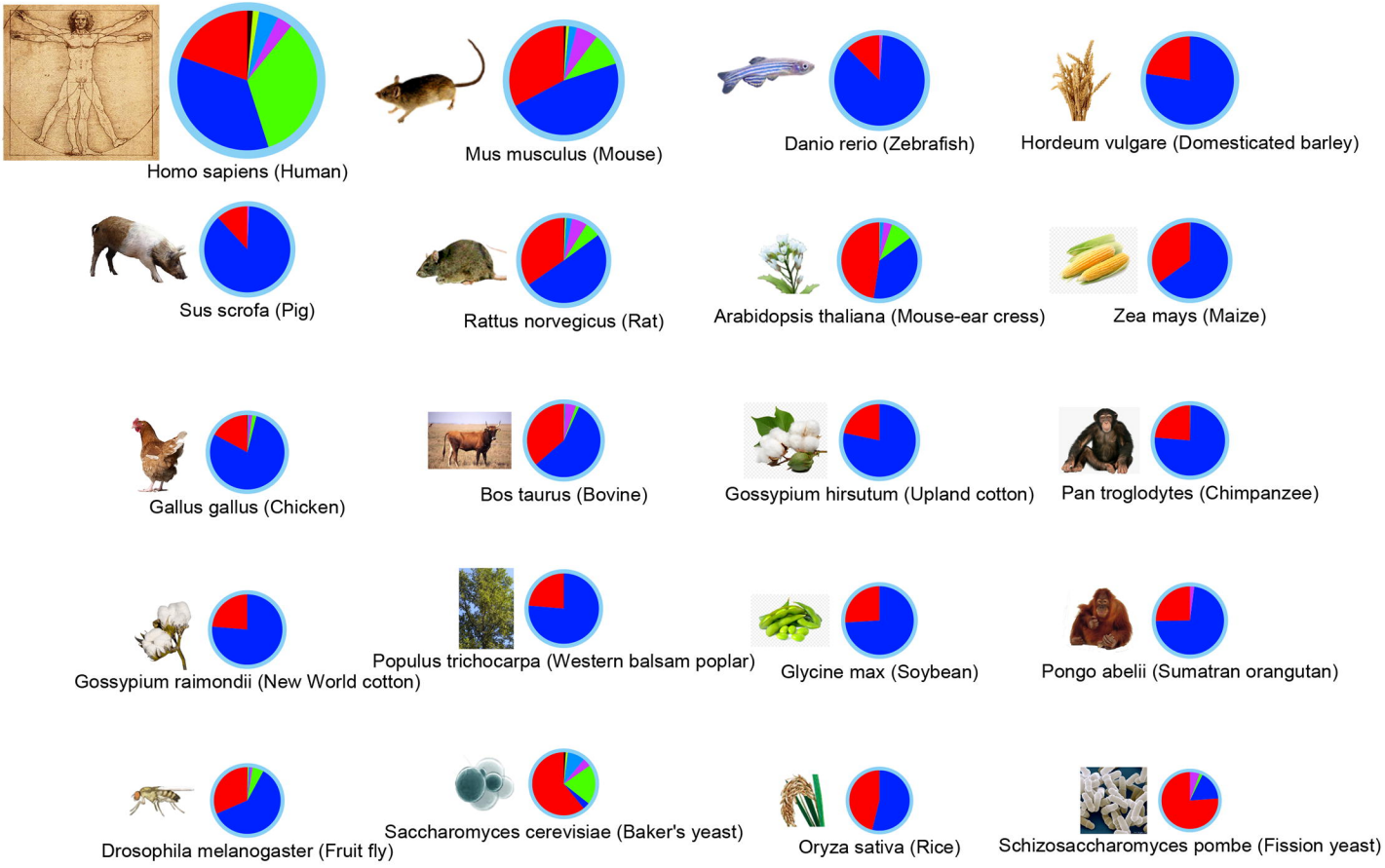
Protein Names:	Amyloid beta A4 precursor protein-binding family B member 1-interacting protein (APBB1-interacting protein 1) (Proline-rich EVH1 ligand 1) (PREL-1) (Proline-rich protein 73) (Rap1-GTP-interacting adapter molecule) (RIAM) (Retinoic acid-responsive proline-rich protein 1) (RARP-1)
Gene name:	APBB1IP
Organism:	*Homo sapiens (Human)*
Taxonomy (NCBI):	9606
Sequence Length:	666
Status:	reviewed
Source Type:	UniProtKB MBPpred Show More Transmembrane Interactor -- Transmembrane Protein(s): Q8IUH5
UniProt AC:	Q7Z5R6
Present on Lipid Raft (score):	Yes (0)
Additional Information:	Show

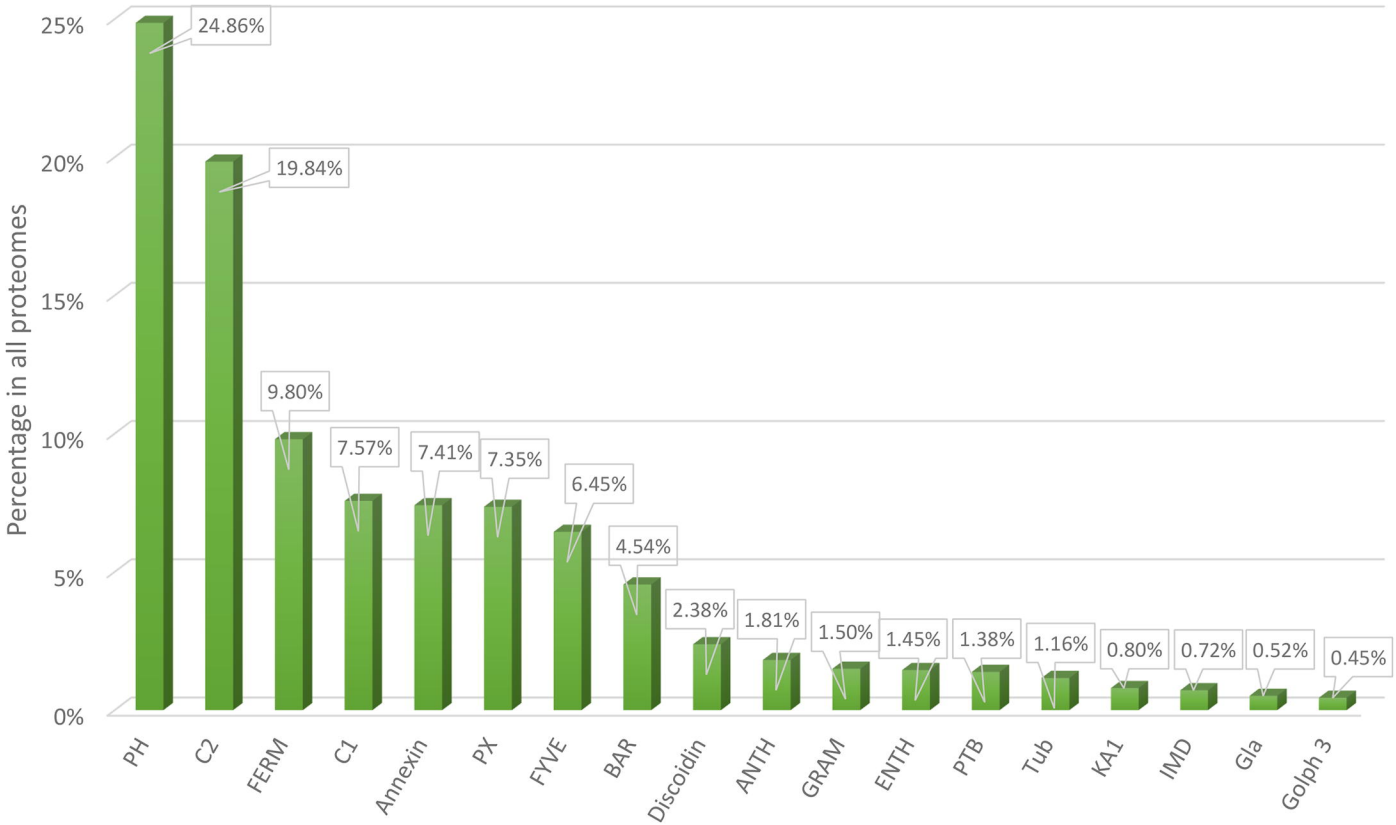
Interaction Network

Right click on Protein Nodes to go to UniProt.

Cross-References

BioGrid:	120012
DrugBank:	No cross references
EMBL:	AB085852 AY152730 BC035636 BC054516 CH471072
HGNC:	17379
IntAct:	Q7Z5R6
KEGG:	hsa:54518
OMIM:	609036
Orphanet:	No cross references
PDB:	2MWN 3ZDL
Pfam:	PF00169 PF00788
PROSITE:	PS50003 PS50200
Reactome:	R-HSA-354192 R-HSA-354194 R-HSA-372708 R-HSA-5674135 R-HSA-6802946 R-HSA-6802948 R-HSA-6802949 R-HSA-6802952 R-HSA-6802955


Homo sapiens (Human)
Mus musculus (Mouse)
Danio rerio (Zebrafish)
Hordeum vulgare (Domesticated barley)
Sus scrofa (Pig)
Rattus norvegicus (Rat)
Arabidopsis thaliana (Mouse-ear cress)
Zea mays (Maize)
Gallus gallus (Chicken)
Bos taurus (Bovine)
Gossypium hirsutum (Upland cotton)
Pan troglodytes (Chimpanzee)
Gossypium raimondii (New World cotton)
Populus trichocarpa (Western balsam poplar)
Glycine max (Soybean)
Pongo abelii (Sumatran orangutan)
Drosophila melanogaster (Fruit fly)
Saccharomyces cerevisiae (Baker's yeast)
Oryza sativa (Rice)
Schizosaccharomyces pombe (Fission yeast)


Percentage in all proteomes
25%
20%
15%
10%
5%
0%
24.86%
19.84%
9.80%
7.57%
7.41%
7.35%
6.45%
4.54%
2.38%
1.81%
1.50%
1.45%
1.38%
1.16%
0.80%
0.72%
0.52%
0.45%
PH
C2
FERM
C1
Annexin
PX
FYVE
BAR
Discoidin
ANTH
GRAM
ENTH
PTB
Tub
KA1
IMD
Gla
Golph 3