# **Proteomics Databases and Websites**

# Mohieddin Jafari<sup>\*1</sup>, Ali Masoudi-Nejad<sup>2</sup>

<sup>1</sup>Student's Research Committee, Faculty of Paramedical Science, Shahid Beheshti University of Medical Sciences, Tehran, Iran. <sup>2</sup>Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics and Center of Excellence in Systems Biology, University of Tehran, Tehran, Iran.

\*Corresponding Author: Email: mo.jafari@sbmu.ac.ir (M.Jafari)

## ABSTRACT

Information avalanche (overload or expansion) in various scientific fields is a novel issue turned out by a number of factors considered necessary to facilitate their record and registration. Though, the biological science and its diverse fields like proteomics are not immune of this event and even may be as the event's herald. On the other hand, time as the most valued anxiety of human has encountered a huge mass of information. Therefore, in order to maintain access and ease the understanding of information in several fields some emprises have been prepared. Bioinformatics is an upshot of this anxiety and emprise. Interestingly, proteomics through studying proteins collection in alive things has covered a great portion of bioinformatics. Consequently, a noteworthy outlook on proteomics related databases (DBs) and websites not only can help investigators to face the upcoming archive of databases but also estimate the volume of the needed facilitates. Furthermore, enrichment of the DBs or related websites must be the priority of researchers. Herein, by covering the major proteomics related databases and websites, we have presented a comprehensive classification to simplify and clarify their understanding and applications.

Keywords: Proteomics; database and websites; proteome; mass spectrometry; classification

#### **INTRODUCTION**

The rising information growth rate confirms the formation of a problematic future in which new and more useful methods are needed to face and manage the delivered data [2]. Nowadays in biology, mainly in the modern fields like genomics and proteomics, the information management is truly complicated. As we know human genome  $(3 \times 109 \text{ bp})$  carries the information of about 30,000 genes [3]. Additionally, taking the world's seven-billion population and also the various informatics fields like proteomics and metabolomics into account, and even disregarding other organisms which sometimes have bigger genomes than human being, the information web appears more complicated.

Currently, the knowledge of the informatics backed by a-half--a-century experience has established strong bases to handle the information. Finally in the 1990, this field was integrated with biological information and named bioinformatics afterwards [4]. Generally and simply the main concerns of bioinformatics in its early years are conducting clever, fast and approved analyses of the information along with scientific studies by the experts in various fields of computer, statistics, mathematics, chemistry and especially biology to heal the difficulties and guide the investigations.

In a general definition, informatics is a process in which information detection rate or the conception of sent signals including patterns, structures, functions and furthermore their exchange into the knowledge occurs in a more trouble-free state [35].

Proteome, the protein collection of an organism in distinct conditions in lifetime, is one of the most important issues of bioinformatics [5]. According to the definition and by assigning special conditions, we know that in each new condition, different amount-kind collection of the proteins will be encountered that is named subproteome. However, generally, we can divide studies on proteomics in four major groups [1]:

Expression proteomics:

It is the approach that involves the creation of quantitative maps of expressed proteins from any living things in a list, similar to cDNA arrays for transcription profiling. Cell map or interaction proteomics:

It determinates subcellular location of proteins and protein-protein interactions by the purification of organelles or protein complexes followed by mass spectrometric identification and bioinformatics tools.

Functional proteomics:

It addresses the potential role of the target protein in a cellular process by systematic perturbation or functional inactivation of proteins within a given physiological environment.

Structural proteomics:

It is another term of structural biology for proteins or the large-scale determination of 3D structures of proteins.

This creates a more complicated archive than genomic information. For instance, considering 30,000 genes in human genome, around 400,000 proteins are produced [6, 7]. Considering the more annotations of proteome data than genome, now, if we put these data together with structural, functional and evolutional data, we will encounter a great amount of data. Bioinformatics delivers these data and then investigates the knowledge in a more sensible way. Obviously, the World Wide Web besides archiving these data facilitates this goal.

As a consequence of robust growth rate in information quantity and also accumulation of data in the web-based proteome DBs, attentive reflection must be performed. The aim of this paper is to introduce the bulk number of these DBs and related websites as well as their general classification in order to illustrate the resource categories with the structural, topical and applicable priorities in data retrieving and management.

## **DB** and web site exploration

In order to find the data bases and related web sites we picked up four methods; first, general search via general explorer motors Google and Yahoo. Second, after entering the related DBs we looked for the links existing in each home page and found the other data bases. In third method, we used the search result of protein cross references. Finally we used the scientific texts like papers and books in this direction [10-11]

Classification of the DB and web sites

We used four general features to classify DBs as follows:

1) The subject of each DB

2) The relationship between different DBs

3) Primary DBs used in the secondary DBs

4) User-Interface of DBs.

Concerning the first three features, the home page and about section of each DB were the main criteria of judgment. After that, using criterions and citations of Google reference priority was calculated. In order to find the databases and related websites, a four-phased search protocol was conducted. First, general search via general search engines such as Google, Bing and Yahoo was conducted. Then, after entering the related DBs the links existing in each home page were searched and other databases were found. In the third phase, the search result of protein cross references was used. Finally scientific papers and books related with the issue were used [8, 9].

## **PROTEOMICS WEBSITES**

The classified groups were arranged and shown in the figure 1 which has columns including recommended reference priority, name, descriptions/URL, citation links, number of databases and references. Herein, the proteomics related DBs and websites have been divided into five main groups. (For more description, some parts of supplementary data presented following)

## 1) Reference group

The first groups, reference group, include five websites that have the most citation link in other four groups. For example, three of them are; 1) ExPASy proteomics server; 2) human proteome organization (HUPO); 3) Matrix science.

Generality of data is a prominent special issue in these web sites which have not been seen in the other related sites. Moreover, the greatest citations within each group, is nominated in the reference group.

Among these web sites ExPASy is known as the most important protein search and identification and other services site which includes various educational and scientific services in addition to high cross references, full and general annotations of proteins. It is known as one of the most useful and valuable protein primary DB. Altogether, many of the secondary DBs are formed based on this DB. In the HUPO, as an international scientific organization and collaborations, general information about conferences and congresses of the world's proteome society and also the associate centers are available for the members.

Matrix science, the last website, is the most famous and user friendly biological mass spectrometry database which is have a Mascot search as a powerful search engine with three main categories: 1) Peptide Mass Fingerprint (The experimental data are a list of peptide mass values from an enzymatic digest of a protein) 2) Sequence Query (One or more peptide mass value associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query), 3) MS/MS Ion Search (Identification based on raw MS/MS data from one or more peptides).

## 2) Educational and service group

In the second group, the most important Educational services related to proteomics including lessons and portals of worldwide universities, or specific institutions, congresses and lab works are gathered (11 websites).

Albeit number of citation links is not a good criterion for ranking especially in educational websites but to observe similar format in total study did such that.

## 3) Association with Universities & Institutes

The third group includes the web sites that are related to a special university or institution and also data obtained from the completed or ongoing research projects in their interested topics and methods, something that seems necessary for all proficient centers (9 websites), besides the global databases. Thereto their useful information about their projects for different researcher, these web sites show the procedure and interests of these universities and institutions and also a prospect about their future plans. These website are classified in a single group to emphasis that if somebody want to search some databases, they sometimes should be known university or institute which that DBs are belong to it.

## 4) Protein integrity identification

The fourth group includes the advanced (expert) bioinformatics analysis tools sites containing the secondary DBs which are applicable in mass spectrometry in order to define the integrity of the proteins (14 websites). Whereas one of the main areas in proteomics that bioinformatics science with its algorithms and tools flaunt, is the analysis of mass spectrometry data to identification proteins.

Other area which we could see bioinformatics' traces is related to the advanced statistical methods such as kinds of multivariate statistical analysis, data mining, meta analysis, clustering, discrimination, etc. In this domain by integrate different data obtained from different sources or databases, be attempted to close to the answers of our questions or problems and get new questions and clues for following research.

In addition to Matrix science" that is described in" Reference group", The PRIDE or PRoteomics IDEntifications database is a centralized, standards compliant, public data repository for proteomics data. It has been developed to provide the proteomics community with a public repository for protein and peptide identifications together with the evidence supporting these identifications. It has been developed through a collaboration of the EMBL-EBI and Ghent University in Belgium. In addition to identifications, PRIDE is able to capture details post-translational of modifications coordinated relative to the peptides in which they have been found.

PSORT WWW Server is one of the website that derived in proteomics database by new bioinformatics algorithm and new statistical approach. PSORT is a computer program for the prediction of protein localization sites in cells. It receives the information of an amino acid sequence and its source origin, e.g., Gramnegative bacteria, as inputs. Then, it analyzes the input sequence by applying the stored rules for various sequence features of known protein sorting signals. Finally, it reports the possibility for the input protein to be localized at each candidate site with additional information. In the other word this website is a useful tool for cell-map proteomics.



Figure 1: Summarized of classified list of DBs and a pie plot associated to different organisms in proteomics DBs (details in supplementary data)

#### 5) Specific organisms' DBs.

The last group is based on research topics with a focus on special organelles, organs or organisms. Therefore, this group has four proteomics sub-groups involving various species (4 websites), single cells (17 websites), plants & animals (20 websites) and human (22 websites). A variety of distinguished issues is the advantage of this group useful for researchers dealing with centralized studies. Some topics of these websites are similar to the topic or purpose of special institutes or centers that website is related to it, but some institutes or centers have some websites with variable approaches in different organisms.

What is important is that such databases which are available on important topics, can play a vital role in the process of knowledge discovery by leading researchers to the unanswered deserve questions.

As important criteria for web visiting, in these kinds of website usually you can find all of things are related to subject of database. In following for example, it is described some website with something it could be found on them.

GenProtEC, which is about E.coli genome and proteome database, is dedicated to the functions encoded by the Escherichia coli K-12 (strain MG1655) genome defined in the GenBank Accession No. NC\_000913.2 deposit.

Annotation work includes multiple types of information:

- I. Sequence similarity to orthologues as defined by Darwin (start and end of aligned region, identity, and PAM distance).
- II. Resolution of fused proteins into modular units with independent functions.
- III. Identification of sequence similar protein groups within E. coli that is clustered by transitive relationships. The sequence similarity is limited to PAM 200 and an alignment of at least 83 amino acids.

IV. Updated literature references.

- V. Classification of gene products by their gene type and by their cellular role(s). The MultiFun classification system for cellular roles is used to assign gene products to one or more roles. MultiFun has been converted to Gene Ontology terms.
- VI. Families of proteins related by structure and biochemical reaction mechanisms (work in progress).

VII. SCOP superfamily identification and location (e.g. binding site domains) for E. coli proteins.

Another example for plant's proteomics website is Soybean Proteome Database, which is based on data of soybean proteins separated by twodimensional polyacrylamide gel electrophoresis (2D-PAGE). The goal of the database is to provide a repository for 2D-PAGE and proteomics information for functional analysis. The majority of the data is focused on soybean, which is an important crop to supply vegetable oil and protein. Especially, it is focused on the proteins involved in the soybean response to flooding. The database integrates multiple omics. The whole database is coordinated based on a scheme of differential omics to identify time-variant proteins under flooding stress.

Human related DBs haves been placed in the last group as a specific organism because of its importance and multiplicity of websites. One of them is Genome Medicine Database of Japan Proteomics that is also called GeMDBJ Proteomics is an integrated proteome database useful in cancer research. The following information is available in the database:

- I. Expression levels of proteins identified in surgical specimens and cell lines by 2D-DIGE.
- II. Protein identified by liquid chromatography tandem mass spectrometry (LC-MS/MS), with corresponding peptide information.
- III. Clinical and biological information pertaining to the samples.

GeMDBJ Proteomics has been developed as part of the Proteome Bioinformatics Project of the National Cancer Center Research Institute, Tokyo. This database aims to enhance proteome-based research for the benefit of human health and welfare, and will keep expanding as new proteome information becomes available.

## CONCLUSION

After the study, around 96 other proteomics related web sites were indentified from August 1<sup>st</sup> (2009) to March 26<sup>th</sup> (2010). This collection actually includes the most important and famous proteomics related web sites holding the highest cross-links amongst other sites and also a number of links cited in some other web sites in Google. First of all, the reference priority and the way of using them are of great importance. Therefore, considering thematic and applicable criteria, we introduce and recommend the five possible groups. Figure 1 shows classified list of five groups. However, in this classification the recommended reference priorities in each group based on the relation between the topic of the web sites and the applicability of the primary DBs in greater volubility are included in each chart. Additionally, in order to compare the scientific volubility of the mentioned web sites using various web analyzing tools, it is applicable to employ the references to each of the web sites or being linked in which we believe that the latter one can show this aim in more accurately.

It may be inquiring that why we don't have all proteomics related information in a general reference DB. The answer could be that the majority of the information is located in the ExPASy server [34] but the advantage of the other mentioned DBs is to focus on various issues. Additionally each has specialized tools and interfaces based on user-styles to facilitate the search and retrieve data in a fast and more confident system. Additionally, creating a general outlook of proteomics resources, future applications for each of them will be facilitated via generated information.

Some other DBs (except ExPASy which have generality of tools) such as Proteome Database, PRIDE PRoteomics IDEntifications database, ProteinProspector, PRIDE PRoteomics IDEntifications database, SALAD database and Microbial Research for Infection Biology Server contain different tools for various useful and applicable analyses.

Conclusion: Meanwhile using the reference DBs, it's recommended for researchers to take advantage of specialized DBs close to their research program. This could make DBs richer and stronger for directing deeper future plans and ideas. Also, it's recommended that researchers in other fields of proteomics and omics, research institutions and universities try to establish such DBs with the aim of ordering their proteomics investigations. Besides

## REFERENCES

1. Ann Finney Batiza, Bioinformatics, Genomics, and Proteomics: Getting the Big Picture (Biotechnology in the 21st Century), Chelsea House Publications, 2009.

2. Lyman P, Varian HR. How much information? Technical report, School of

scientific applications and worldwide awareness about proteomics, these DBs will make this field progress smoother. In this regard, plenty of software have been developed for these DBs such as "Make 2D-DBII", a tool of ExPASy website, and GDPE (Griss Proteomics database Engine).

## Key issues

- In the biology section, especially modern technology related sections like genomics overload of information is ever-growing.
- Informatics knowledge has been established during the past fifty years to decrease the information overload anxiety. In 1970 this knowledge was integrated with biology data and created a new branch named bioinformatics.
- According to proteomics definition we are aware that in each new condition we are exposed to plenty of new protein combination which is beyond of genomic data.
- Proteomics related web sites are classified to five groups based on their topic including; 1) reference group; 2) education and service group; 3) association with Universities and institutions; 4) Protein integrity identification; and finally 5) specific organisms DBs.
- The most important protein search, identification and other applicable services tools are employed in ExPASy protein server which includes various educational and scientific services. High cross references and inclusive annotation of proteins are the valuable advantages of this web site as a primary protein DB nominating it as the most popular and appreciated Web site.

## ACKNOWLEDGEMENT

This work was derived from PhD dissertation results. Authors would thank Sadegh Azimzadeh Jamalkandi, Dr. Armin Madadkar Sobhani , Dr. Hamid Moghaddasi and Dr. Mostafa Rezaei Tavirani for their consultations during this project.

Information Management (ed. Berkeley, U), 2000.

3. Dolezel J, Bartoš J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. Cytometry 2003; 51 (2), 127-128 4. Hogeweg P, Interactive instruction on population interactions. Comput Biol Med 1978; 8:319-27.

5. Wilkins MR, Pasquali C, Appel RD et al. From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. Nature Biotechnology 1996; 14(1), 61-65.

6. McDermott J, Samudrala R. Bioverse: functional, structural and contextual annotation of proteins and proteomes. Nucleic Acids Res. 2003; 31, 3736-3737.

7. Tavirani MR, marashi A, galambar F, Mostafavi M. Proteomics, Andishe Zohoor Publication, Tehran 2006.

8. Batiza AF. Bioinformatics, Genomics, And Proteomics: Getting the Big Picture, Chelsea House 2005.

9. Gilany K, Moens L, Dewilde S. Mass spectrometry-based proteomics in the life science: a review. Journal of Paramedical Siences (JPS) 2009; 1(1).

10. Appel RD, Bairoch A, Hochstrasser DF. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends Biochem Sci 1994; 19(6), 258-260.

11. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999; 20(18), 3551-3567.

12. Hanash S, Celis JE. The Human Proteome Organization. Molecular & Cellular Proteomics 2002; 1(6), 413-414.

13. Bini L, Sanchez-Campillo M, Santucci A et al. Mapping of Chlamydia trachomatis proteins by Immobiline-polyacrylamide two-dimensional electrophoresis: Spot identification by N-terminal sequencing and immunoblotting Electrophoresis. Electerophoresis 1996; 17(1), 185-190.

14. Lemkin PF, Thornwall G, Evans J. Comparing 2-D Electrophoretic Gels Across Internet Databases, Totowa, 2005.

15. Lemkin PF, Thornwall G. Flicker image comparison of 2-D gel images over the Interne, Totowa t, 2002.

16. Lemkin P. NCI 2DWG Image Meta-Database, 1999.

17. Rasmussen RK, Ji H, Eddes JS et al. Two-dimensional electrophoretic analysis of human breast carcinoma proteins: Mapping of proteins that bind to the SH3 domain of mixed lineage kinase MLK2. Electrophoresis 1997; 18, 588-598.

18. Jungblut PR, Seifert R. Analysis by high-resolution two-dimensional electrophoresis of differentiation-dependent alterations in cytosolic protein pattern of HL-60 leukemic cells. J Biochem Biophys Meth 1990; 21, 47-58.

19. Serres MH, Goswami S, Riley M. GenProtEC: an updated and improved analysis of functions of Escherichia coli K-12 proteins. Nucleic Acids Res 2004; 32(1), D300-302.

20. Steinmetz lm, Scharfe C, Deutschbauer AM et al. Systematic screen for human disease genes in yeast. Nat Genet 2002; 31(4), 400-404.

21. Kubis S, Baldwin A, Patel R et al. The Arabidopsis ppi1 mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. Plant Cell 2003; 15(8), 1859-187.

22. Tunica DG, Yin X, Sidibe A et al. Proteomic analysis of the secretome of human umbilical vein endothelial cells using a combination of free-flow electrophoresis and nanoflow LC-MS/MS. Proteomics 2009; 9, 4991.

23. Sorin C, Negroni L, Balliau T et al. Proteomic analysis of different mutant genotypes of Arabidopsis thaliana led to the identification of thirteen proteins correlating with adventitious root development. Plant Physiol 2005; 140, 349-364.

24. Ferry-Dumazet H, Houel G, Montalent P et al. PROTICdb: a web-based application to store, track, query, and compare plant proteome data. Proteomics 2005; 5(8), 2069-2081.

25. Riano-Pachon DM, Nagel A, Neigenfind J et al. GabiPD: The GABI Primary Database - a plant integrative 'omics' database. Nucleic Acids Research 2009; 37(Database issue), D954-959.

26. Jungblut P, Otto A, Zeindl-Eberhardt E et al. Protein composition of the human heart: The construction of a myocardial two-dimensional electrophoresis database. Electrophoresis 1994; 15, 685-707.

27. Hubbard MJ, Faught MJ, Carlisle BH, Stockwell PA. ToothPrint, a proteomic database for dental tissues. Proteomics 2001; 1, 132-135.

28. Evans G, Wheeler CH, Corbett JM, Dunn MJ. Construction of HSC-2DPAGE: A two-diemnsional gel electrophoresis database of heart proteins. Electrophoresis 1997;18, 471-479.

29. Elstner M, Andreoli C, Klopstock T, Meitinger T, Prokisch H. The Mitochondrial Proteome Database MitoP2. Methods Enzymol 2009; 457, 3-20.

30. Zhang Y, Zhang Y, Adachi J et al. MAPU: Max-Planck Unified database of organellar, cellular, tissue and body fluid proteomes. Nucleic Acids Res 2007; 35, D771-779.

31. Pleissner KP, Sander S, Oswald H, Regitz-Zagrosek V, Fleck E. The construction of the World Wide Webaccessible myocardial two-dimensional gel electrophoresis protein database "HEART-2DPAGE": A practical approach. Electrophoresis 1996; 17, 1386-1392.

32. Noel-Georis I, Bernard A, Falmagne P, Wattiez R. Database of bronchoalveolar lavage fluid proteins. J Chromatogr B Analyt Technol Biomed Life Sci 2002; 771(1-2), 221-236.

33. Ohlmeier A, Kastaniotis AJ, Hiltunen JK, Bergmann U. The yeast mitochondrial proteome, a study of fermentative and respiratory growth. J Biol Chem 2004; 279(6), 3956-3979.

34. Minuchehr Z, Goliaei B. Protein database. IJPT 2004, 3, 1-11.

35.http://bioinformatics.ut.ac.ir/armin/cours es/index.php. (Ed.^(Eds) Madad-kar

#### Supplementary data: Classification of proteomics related websites including their citation links, included databases and citation based priority.

Priority	name	Website Self-Description/URL	Citation Link	Number of DBs	Ref			
1) Reference web sites								
1	ExPASy Proteomics Server (Expert Protein Analysis System)	Analyzes protein sequences and structures as well as 2-D PAGE. (http://expasy.org)	2180	24	[10]			
2	Master's degree in proteomics and bioinformatics	The Master's in proteomics and bioinformatics purposes theoretical courses, wet-lab and dry-lab hands-on experience to study bioinformatics and proteomes. The key topics investigated in the research laboratories are reflected in the teaching program of the Master's degree in Proteomics and Bioinformatics. (http://www.mpb.unige.ch/)	466	-	-			
3	Matrix science	This site features Mascot, a powerful search engine that uses mass spectrometry data to identify proteins from primary sequence databases. ( <u>http://www.matrixscience.com</u> )	191	15	[11]			
4	HUPO website (Human Proteome Organization)	An international scientific organization representing and promoting proteomics through international cooperation and collaborations by fostering the development of new technologies, techniques and training. (http://www.hupo.org)	104	-	[12]			
5	SIENA-2DPAGE	A Two-dimensional polyacrylamide gel electrophoresis database. (http://www.bio-mol.unisi.it/2d/2d.html)	19	3	[13]			
2) Education	al and service group							
1	Master's degree in proteomics and bioinformatics	Described earlier.	466	-				
2	HUPO website (Human Proteome Organization)	Described earlier.	104	-				
3	The Open2Dprot Project	The Open2Dprot project is a community effort to create an open source n-dimensional (n-D) protein expression data pipeline- analysis system. It is downloadable and could be used for exploratory data analysis of protein expression data across sets of n-dimensional (n-D) data from research experiments. Interchangeable subproject modules are being developed for 2-dimensional data including 2D-PAGE. (http://open2dprot.sourceforge.net/index.html)	23	-	[14] [15]			
4	NCI 2DWG Image Meta-Database	A meta-database contains information about data in other databases. The 2DWG database is an organized as a spreadsheet table with each gel image represented by a row. Data for each gel entry includes tissue, species, cell-line, image URL	10	-	[16]			

		database URL, gel protocol, organization URL, image properties, map URL if it exists, etc. It uses a simple database search engine dbEngine to access selected rows of gels in the full 2DWG table (http://www-lecb.nciferf.gov/2dwgDB/)			
5	Proteomics Courses Database	The objectives are to hold and arrange courses and meetings on issues connected to Proteomics to facilitate the exchange of scientific information between biochemists, molecular biologists, biologists, chemists and related scientists generally and especially in Swizerland by holding meetings and discussions and by other appropriate means. (http://www.swissproteomicsociety.org/courses/list.pl?location)	8	-	-
6	NCI Flicker Web Server	Flicker is a method to compare images from different Internet sources on your Web browser. Scientists around the world often work on similar image data. A great deal of of this kind of data is being published on the Internet each year. In the case of 2D protein electrophoretic gel images, maps identifying proteins in these gels are becoming increasingly available. Visually comparing 2D sample gels against these 2D gel database maps may suggest putative protein spot identification in many cases. Flicker was originally developed for comparing 2D protein gels across the Internet. (http://www-lecb.ncifcrf.gov/flicker/)	8	-	-
7	Protein Identification Tutorial	Contents of this site includes: Introduction, Protein Identification Philosophy, Protein Identification Strategies, Peptide-Mass Fingerprinting, Peptide Sequence Tag, MS/MS Protein ID, Searching With Complex Proteomics Data, Databases, Tutorial References and Additional Suggested Readings. (http://www.ionsource.com/tutorial/protID/idtoc.htm)	3	-	-
8	Proteomics Database Portal	Creation of PRIME, a centralized, web-based resource for tracking and managing the data generated by proteomics research projects are its tasks. In the process of protein analysis, from running gels to digesting selected proteins and running mass-spectrometer analyses for protein identification, large amounts of data are generated. (http://www.ctaalliance.org/MCBI/Proteomics.html)	0	-	-
9	Thermo scientific	Technologies and integrated solutions help tackle the analytical challenges of the emerging field of proteomics. They create innovative laboratory solutions in protein separation science, automation, mass spectrometry, informatics, and computer science. (http://reg.accelacomm.com/servlet/Frs.frs?Context=LOGENT RY&Source=proteomics&Source_BC=35&Script=LP/8019830 6/reg&&gclid=CPjZxuiW150CFcmEzAodQFfv2g)	0	-	-
10	Isoelfocus 3.0	A multimedia computer-aided learning program on Isoelectric Focusing and its derived techniques as tools in biomedicine and agro food industries. (http://www.snv.jussieu.fr/Bio/logiciels/isoelfocus/isoelfocus.ht ml)	0	-	-
11	Seattle Proteome Center (SPC)	The goal of the Seattle Proteome Center (SPC) is to evolve as a local, highly interactive, multi-disciplinary group, bringing together scientific expertise from diverse areas such as biology, chemistry, proteomics, physics, bioinformatics, and statistics, to enhance and develop innovative proteomic technologies and apply them to biological questions relevant to systems biology and, ultimately, human health and wellbeing.			

		The focus of the Center is the development of an array of new, systematic tools and techniques for the analysis of proteins and to create assays to comprehensively study the dynamics of cells in health and disease. Validated technologies with generic applicability are broadly disseminated here from both tools, techniques and comprehensive data sets in an open source manner. The SPC is supported in part by the Duchy of Luxembourg, the National Institute of General Medical Sciences (NIGMS), the National Science Foundation (NSF), the National Human Genome Research Institute (NHGRI), and in the past by the National Heart, Lung, and Blood Institute (NHLBI).			
		( <u>http://www.proteomecenter.org/)</u>			
3) Associated	l with Universities & Inst	itute			
1	SIENA-2DPAGE	Described earlier.	19	3	-
2	University of Antwerp - Proteome Database	Public 2D gel electrophoresis profiles and protein identifications repository. (http://www.pdata.ua.ac.be/)	10	3	-
3	Joint Proteomics Laboratory	This web site focuses on analytical biochemistry, technical developments in protein separation and characterization, and Proteomics. (http://www.ludwig.edu.au/jpsl/jpslhome.html)	8	2	[17]
4	Washington University Inner Ear Protein Database	An all-out effort has been initiated towards establishment of an Inner Ear Proteomics Project, i.e. towards the study of the 'total protein complement of the genome'. Unlike the genome of an organism, which is essentially fixed, the proteome varies with tissue, cell organelles, as well as developmental stage, disease state, or environmental conditions. This endeavor, while gargantuan, is absolutely worthwhile and quite feasible. (http://oto2.wustl.edu/thc/innerear2d.htm)	4	1	-
5	Fachgebiet Proteomik	Focuses on the methodical development of 2D electrophoresis with immobilized pH gradients (IPG-Dalt) for proteome analysis, and the application of proteomics to various fields of research by studying differential protein expression using modern separation and detection technologies as well as mass spectrometry. (http://www.weihenstephan.de/blm/deg/)	2	1	-
6	JHU Proteomic Database	It is used to store raw, processed and final proteomic data from a large number of methods, technologies, clinical and physiological studies from cardiac studies through the National Heart, Lung and Blood Institute (NHLBI). This database will be used to mine information about proteins, protein patterns and post translational modifications (PTMs) and the associated methods and technologies. The data will be used to help create "protein profiles" to determine not only if disease is present, but the severity and progression of disease in the patient. (http://proteomics.jhu.edu/dl/pathidb.php)	2	1	_
7	Proteomics Database of the Medical University of Vienna	Currently the database contains: 22 cell types	1	3	-

		<ul> <li>4.548 proteins</li> <li>200000 peptides</li> <li>110 experiments conducted by 1 project group</li> <li>published in 9 papers</li> </ul>			
		(http://www.meduniwien.ac.at/proteomics/database/#simple_se arch/)			
8	University of Tennessee Proteome and Transcriptome Database	This site features 2D-PAGE reference databases of several mammalian organs, cells, and biological fluids. These maps are being developed to study global protein expression in health and disease. Each reference database contains a digitized 2D gel image, and data on proteins identified in the 2D gel map.	0	3	_
9	2de Database	(http://www.utmem.edu/proteomics/)	0	2	-
			Ŭ	-	
4) Associated	d to identifying protein (N	Iass spectrometry & Bioinformatics analyzing method)			
1	Matrix science	Described earlier.	191	15	-
2	PSORT WWW Server	PSORT is a computer program for the prediction of protein localization sites in cells. It receives the information of an amino acid sequence and its source origin, e.g., Gram-negative bacteria, as inputs. Then, it analyzes the input sequence by applying the stored rules for various sequence features of known protein sorting signals. Finally, it reports the possibility for the input protein to be localized at each candidate site with additional information. (http://psort.ims.u-tokyo.ac.jp/)	115	-	-
3	PRIDE PRoteomics IDEntifications database	The PRIDE PRoteomics IDEntifications database is a centralized, standards compliant, public data repository for proteomics data. It has been developed to provide the proteomics community with a public repository for protein and peptide identifications together with the evidence supporting these identifications. It has been developed through a collaboration of the EMBL-EBI and Ghent University in Belgium. In addition to identifications, PRIDE is able to capture details of post-translational modifications coordinated relative to the peptides in which they have been found. (http://www.ebi.ac.uk/pride)	114	-	-
4	Laboratory of Mass Spectrometry and Gaseous Ion Chemistry	Includes multiple tools for searching protein sequence, retrival and analysis of protein sequences, finding modifications on proteins and identifying proteins matching mass spectra. (http://prowl.rockefeller.edu/)	62	_	-
5	The Global Proteome Machine Organization	The Global Proteome Machine Organization was set up so that scientists involved in proteomics using tandem mass spectrometry could use that data to analyze proteomes. The projects supported by the GPMO have been selected to improve the quality of analysis, make the results portable and to provide a common platform for testing and validating proteomics results.	29	8	-
6	PeptideAtlas	PeptideAtlas is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. Mass spectrometer output files are collected for human, mouse, yeast, and several other organisms, and searched using the latest search engines and protein sequences. All results of sequence and spectral library searching are subsequently processed through the Trans	29	-	-

		Proteomic Pipeline to derive a probability of correct identification for all results in a uniform manner to insure a high quality database, along with false discovery rates at the whole atlas level. Results may be queried and browsed at the PeptideAtlas web site. The raw data, search results, and full builds can also be downloaded for other uses.			
		(http://www.peptideatlas.org/)			
7	RESID	The RESID Database of Protein Modifications is a comprehensive collection of annotations and structures for protein modifications including amino-terminal, carboxyl-terminal and peptide chain cross-link post-translational modifications. (http://www.ebi.ac.uk/RESID/index.html)	18	-	-
8	Open Proteomics Database	OPD is a public database for storing and disseminating mass spectrometry based proteomics data. The database currently contains roughly 3,000,000 spectra representing experiments from 5 different organisms. (http://bioinformatics.icmb.utexas.edu/OPD)	9	-	-
9	Database of Interacting Proteins	The DIPTM database catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein- protein interactions. The data stored within the DIP database were curated, both, manually by expert curators and also automatically using computational approaches that utilize the knowledge about the protein-protein interaction networks extracted from the most reliable, core subset of the DIP data. (http://dip.doe-mbi.ucla.edu/dip/Main.cgi)	5	-	-
10	Integr8: Access to complete genomes and proteomes	Provides easy access to integrated information about deciphered genomes and their corresponding proteomes. Available data includes DNA sequences; protein sequences; statistical genome and proteome analysis and information about orthology, paralogy, and synteny. (http://www.ebi.ac.uk/integr8/EBI-Integr8-HomePage.do)	3		-
11	PEDANT 3 database	The pedant genome database provides exhaustive automatic analysis of genomic sequences by a large variety of bioinformatics tools. For example the following pre-computed analyses are available to analyze protein function: Blast similarity searches against the non-redundant protein sequence database, motif searches against the Pfam, BLOCKS and PROSITE databases. Prediction of cellular roles and functions are made based on high stringency BLAST searches against protein sequences with manually assigned functional categories according to the Functional Catalogue (FunCat) developed by MIPS and Biomax AG. Protein structure analysis includes the similarity-based identification of known 3D structures and structural domains, by searching against the PDB and SCOP databases, and predictions of transmembrane regions, low similarity regions and non-globular domains. (http://pedant.gsf.de/)	0	-	-
12	SALAD Database	SALAD is a tool of systematic comparison of proteome data among the species and survey conserved motifs Alignment diagram and the Associating Dendrogram. (http://salad.dna.affrc.go.jp/salad/en/)	0	-	_
13	Protein Prospector	Proteomics tools for mining sequence databases in conjunction with Mass Spectrometry experiments. (http://prospector.ucsf.edu/prospector/mshome.html)	0	-	-

14	pepMapper	Mapper, the UMIST protein search tool which uses Mass Spec. data produced by the digestion of a protein to identify a match to a protein from a database. (http://www.nwsr.manchester.ac.uk/mapper)		_	-
5) proficient	DB about organisms				[
a) W	eb sites related to variou	s species			
1	ExPASy Proteomics Server	Described earlier.	2180	24	
2	Proteome 2D-PAGE Database	Recovers expressive information about the characterized proteins on 2D-PAGE maps as proteomes of the various organisms such as; <i>Bacillus amyloliquefaciens, Bacillus anthracis, and etc.</i> (http://web.mpiib-berlin.mpg.de/cgi-bin/pdbs/2d-page/extern/index.cgi)	18	3	[18]
3	Proteome Database System for Microbial Research for Infection Biology	(http://www.mpiib-berlin.mpg.de/2D-PAGE/)	18	17	[18]
4	Negative Proteome Database	( <u>http://superfly.ucsd.edu/negprot/</u> )	-	10	-
b) w	eb sites related to single c	ell organisms			
1	E.coli genome and proteome database	<ul> <li>GenProtEC is dedicated to the functions encoded by the Escherichia coli K-12 (strain MG1655) genome defined in the GenBank Accession No. NC_000913.2 deposit. Annotation work includes multiple types of information:</li> <li>Sequence similarity to orthologues as defined by Darwin (start and end of aligned region, identity, and PAM distance).</li> <li>Resolution of fused proteins into modular units with independent functions.</li> <li>Identification of sequence similar protein groups within E. coli that are clustered by transitive relationships. The sequence similarity is limited to PAM 200 and an alignment of at least 83 amino acids.</li> <li>Updated literature references.</li> <li>Classification of gene products by their gene type and by their cellular role(s). The MultiFun classification system for cellular roles is used to assign gene products to one or more roles. MultiFun has been converted to Gene Ontology terms.</li> <li>Families of proteins related by structure and biochemical reaction mechanisms (work in progress).</li> <li>SCOP superfamily identification and location (e.g. binding site domains) for E. coli proteins.</li> </ul>	55	1	[19]
2	Yeast Deletion Project and Proteomics of	The YDPM database surveys to support the Yeast Deletion and the Mitochondrial Proteomics Project. The database is hyperlinked with other public databases. The project aims to increase the understanding of mitochondrial function and	20	1	[20]

	Mitochondria Database	biogenesis in the context of the cell.			
		(http://www-deletion.stanford.edu/YDPM/YDPM_index.html)			
3	DBase: 2D-PAGE Database of Escherichia coli	It's a 2D-PAGE Database of <i>Escherichia coli</i> (in various growing conditions) at University of Bielefeld - Fermentation Engineering Group (AGFT) in Germany. (http://2dbase.techfak.uni-bielefeld.de/cgi-bin/2d/2d.cgi)	8	1	-
4	2D PAGE Aberdeen ANALYSIS of MICROBIAL PROTEOMES Department of Medical Microbiology, University of Aberdeen	It is directed towards the analysis of virus and bacterial protein synthesis using (2D-PAGE). The proteomes of a number of medically important human bacterial pathogens are actively under investigation including <i>Haemophilus influenzae</i> , <i>Streptococcus pneumoniae</i> and <i>Neisseria meningitidis</i> . Similar studies of Helicobacter pylori are also underway. In addition, 2D-PAGE is being used to investigate protein synthesis in picornavirus infected cells with a specific emphasis on determining the mechanism of persistent Coxsackie B virus infection. This includes the analysis of virus evolution during persistent infection in laboratory grown cell cultures. (http://www.abdn.ac.uk/~mmb023/2dhome.htm)	8	5	-
5	PHCI-2DPAGE Parasite Host Cell Interaction 2D-PAGE database	Parasite-host cell interaction 2D-PAGE database, at Department of Medical Microbioloy and Immunology, University of Aarhus, Denmark. ( <u>http://www.gram.au.dk/</u> )	7	1	-
6	Yeast Protein Map	YPM reports our last knowledge on protein identification of the Yeast Protein Map. To date, YPM shows 716 identified spots corresponding to 485 different genes. (http://www.ibgc.u-bordeaux2.fr/YPM/)	5	1	-
7	Proteome Database of Lactococcus lactis	Proteome Database of <i>Lactococcus lactis</i> 2D Reference Maps & Summarized Protein Information. (http://www.wzw.tum.de/proteomik/lactis/)	4	1	-
8	Methanococcus jannaschii 2D database	Methanococcus jannaschii 2D database at Argonne protein mapping group server, Chicago, USA. (http://proteomeweb.anl.gov/)	4	1	-
9	Pseudomonas syringae 2DE Maps	Pseudomonas syringae 2DE Maps from Lee Research Group at Cornell University, USA. (http://www.leelab.org/resources/syringae/index.html)	3	1	_
10	The Proteome Map of C. glutamicum	Partially Proteome map of <i>C. glutamicum</i> (soluble, secreted and membranes proteins) at Institute of Biotechnology, Jülich, Germany. ( <u>http://www.fz-juelich.de/ibt/cg-proteomic/</u> )	2	1	-
11	ANALYSIS of BACTERIAL PROTEOMES	Current proteomic researches in the Food Quality and Security team of the Microbiology Research Unit are focused on (i) the analysis of the proteome of the food borne pathogen <i>Listeria</i> <i>monocytogenes</i> (2-D PAGE database), (ii) the study of the variations of <i>L. monocytogenes</i> and <i>Staphylococcus xylosus</i> (starter for fermented milk and meat products) subproteomes during growth in biofilm, from the adhesion on an abiotic surface to the biofilm formation and ageing, (iii) the comparative subproteome analysis of <i>L. monocytogenes</i> strains from different serovars and origins and (iv) the study of adaptative response to environmental and technological stresses of food borne pathogen bacteria or spoilage bacteria of meat products.	2	2	-

		(http://www1.clermont.inra.fr/proteome/index_angl.htm)			
12	Compluyeast-2D-DB	Compluyeast-2D-DB is a 2DPAGE federated database built with the Make2DDB package and integrated at the WORLD- 2DPAGE of the ExPASy proteomics web server. (http://compluyeast2dpage.dacya.ucm.es/)	1	1	-
13	Plasmodium 2D Database	Plasmodium falciparum 2-DE database at Indian Institute of Science, Bangalore, India. (http://utlab3.biochem.iisc.ernet.in/cgibin/Plasmo2Dbase/Plasm o2Dbase.cgi)	0	1	-
14	SWICZ:Swiss-Czech Proteomics Server	Proteomics Database of <i>Streptomyces coelicolor</i> (liquid cultivation). (http://proteom.biomed.cas.cz/strepto/cc1_strep.php)	0	1	-
15	Helicobacter pylori 26695 pH-Dependent Proteins	Helicobacter pylori Proteome 2-D Gel Database, at Kenyon College, Ohio, USA. (http://biology.kenyon.edu/slonc/helico/)	0	1	-
16	Yeast Mitochondrial Proteome	Comprises to 2-D gel electrophoresis, image analysis and protein identification by mass spectrometry. The 2-D gel electrophoresis and image analysis facility includes IEF and SDS-PAGE. Gels are scanned by an ImageScanner for Coomassie stained and silver stained gels or a Molecular Imager FX Pro for autoradiography and fluorescence. The protein pattern is analyzed by specific software. For mass spectrometry the core facility operates two different instruments. A VOYAGER DE-STR MALDI-TOF, and a Q- TOF electrospray mass spectrometer. The MALDI-TOF has its main application in the identification of proteins from 2-D gels, while the Q-TOF can provide sequence information of peptides and is thus used to characterize proteins (post-translational modifications) in more detail. (http://www.biochem.oulu.fi/proteomics/ymp.html)	0	1	-
17	Saccharomyces cerevisiae IPG 6-12	The Nottingham Arabidopsis Stock Centre (NASC) provides seed and information resources to the International Arabidopsis Genome Programme and the wider research community. (http://www.weihenstephan.de/blm/deg/2ddb/)	0	1	-
c) w	eb sites related to plants a	and animals			
1	NASC Proteomics database for Arabidopsis data	(http://proteomics.arabidopsis.info/)	82	1	[21]
2	The Plant Proteome Database	PPDB is a Plant Proteome DataBase for <i>Arabidopsis thaliana</i> and maize ( <i>Zea mays</i> ). Initially PPDB was dedicated to plant plastids, but has now expanded to the whole plant proteome – hence it was renamed from Plastid PDB to Plant PDB in November 2007. (http://ppdb.tc.cornell.edu/)	16	2	-
3	Rice Proteome Database	Rice Proteome Database contains information on proteins identified from several tissues and organelles on 2D-PAGE reference maps. (http://cdna01.dna.affrc.go.jp/RPD/main_en.html)	8	1	-

4	vascular-proteomics	(http://www.vascular-proteomics.com/)	7	1	[22]
5	The Arabidopsis mitochondrial proteome project	<ul> <li>Arabidopsis thaliana mitochondrial proteome project was started for a comprehensive investigation of mitochondrial functions in plants. As a first step the proteome of mitochondria isolated from suspension cell cultures was characterized by high resolution 2D gel electrophoresis and mass spectrometry. Currently the proteomes of selected Arabidopsis knock-out mutants concerning nuclear genes of mitochondrial proteins are under investigation.</li> <li>(http://www.gartenbau.uni-hannover.de/arabidopsis.html)</li> </ul>	5	1	_
6	Arabidopsis 2D gel	(http://www.isv.cnrs-gif.fr/gel2d/)	5	1	-
7	Seed Proteome	(http://www.seed-proteome.com/)	4	2	-
8	plastid protein database	plprot - a plastid protein database which integrates data from large scale proteome analyses of different plastid types including etioplasts, chloroplasts, chromoplasts and the undifferentiated proplastid-like organelles of tobacco BY2 cells. This comparison allows establishing a core proteome that is common to all plastid types and provides furthermore information about plastid type-specific functions.	4	1	-
		( <u>http://www.piprot.etnz.ch/</u> )			
9	Soybean Proteome Database	Soybean Proteome Database is based on data of soybean proteins separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). The goal of the database is to provide a repository for 2D-PAGE and proteomics information for functional analysis. The majority of the data is focused on soybean, which is an important crop to supply vegetable oil and protein. Especially, it is focused on the proteins involved in the soybean response to flooding. The database integrates multiple Omics. The whole database is coordinated based on a scheme of differential Omics to identify time-variant proteins under flooding stress.	3	1	-
10	Two dimensional gel electrophoresis	2-D gels for maritime pine at INRA, Bordeaux, France. (http://www.pierroton.inra.fr/genetics/2D/)	3	1	-
11	Proteomics of oilseeds	The following website was created to disseminate data from NSF-Young Investigator Award Plant Genome Research program-funded research project DBI-0332418 "Proteomics of Seed-filling in Oilseeds" awarded October 1, 2003. (http://www.oilseedproteomics.missouri.edu/)	2	4	-
12	Mt PROTEOMICS	This web site covers the symbiotic proteomics of Medicago truncatula roots at INRA, Epoisses, France.	2	1	-
13	PROTICdb	PROTICdb is a web-based application mainly designed to store and analyse plant proteome data obtained by 2-D PAGE and mass spectrometry (MS). (http://cms.moulon.inra.fr/proticdb/web_view/index.php)	2	2	[23] [24]
14	GABI primary database (GabiPD)	2-DE database for 8 different <i>Arabidospis thaliana</i> tissues (inflorescence, stem, primary leaf, leaf, seed, seedling, silique, root) of the Genomanalyse im biologischen System Pflanze, Germany.	1	1	[25]

		(http://gabi.rzpd.de/projects/Arabidopsis_Proteomics/)			
15	Rice Proteome Database	Rice Proteome Database contains information on proteins identified from several tissues and organelles on two- dimensional polyacrylamide gel electrophoresis (2D-PAGE) reference maps. (http://gene64.dna.affrc.go.jp/RPD/database_en.html)	1	1	-
16	Proteomics and Protein Structure Study Group	Proteomics and Protein Structure Study Group, at Milano University, Italy Co Sera (during dioestrus), positive acute- phase reactants detected in serum from a cow with udder edema (inset C) Rat Serum (control Sprague-Dawley (CD) female and male rats, turpentine-treated rats, diabetic rats, stroke-prone rats), urine (control female and male rats, stroke-prone rats), cerebrospinal fluid (control and stroke-prone rats) Mouse Serum. (http://linux.farma.unimi.it/)	1	3	-
17	Drosophila Proteomics Database	The fruit fly, <i>Drosophila melanogaster</i> , has been utilized as a model of genetic organism to understand the fundamental biological systems in human. In this database, a proteomic database of Drosophila using a combination of two-dimensional gel electrophoresis and mass spectrometry has been constructed. This database contains a list of identified proteins, a reference protein map of two-dimensional gel, and detailed information of MS analyses. (http://www.ou.edu/journals/dis/ProteomicsDatabase/Proteomic <u>s_Home.htm</u> )	1	1	_
18	2DPAGE DATABASE	The goal of this database is to provide a repository for 2D- PAGE and proteomics information for plant research. The majority of the data is focused on the model legume, <i>Medicago</i> <i>truncatula</i> ; however, data from other species are included and welcomed. (http://www.noble.org/2dpage/)	0	8	-
19	FishProm at the University of Aberdeen Fish Proteomics	This website contains information on proteins identified on 2DE gels from Salmonid fish tissue extracts. You can locate these proteins on the 2DE maps or retrieve in a table of all the protein entries. (http://www.abdn.ac.uk/fishprom/index.shtml)	0	1	-
20	SOLstIS	The tomato proteome database of INRA developes a proteomic approach on the genetic analysis of tomato organoleptic and nutritional quality. Most of experiments combine two- dimensional polyacrylamide gel electrophoresis (2-D PAGE) for protein separation and mass spectrometry (MS) for protein identification. SOLstIS distributes the experimental results and give access to our database (http://w3.avignon.inra.fr/solstis/)	-	1	-
d) Web sites	related to human				
1	Human 2D-page database for proteome analysis in health and disease	The 2-D PAGE databases contain data on proteins identified on various reference gel images. You can display protein names and information on specific protein spots by clicking on the image in which you are interested. Also, you can search by protein name, keywords, Mr and pI or organelle or cellular component. (http://proteomics.cancer.dk/index.html)	22	1	-

2	HEART-2DPAGE	This is the 2-D PAGE database for human heart proteins. It contains data on proteins identified on the 2-D PAGE maps of ventricle and atrium of human heart and includes proteinchemical data obtained at the Wittmann Institute of Technology and Analysis of Biomolecules and at the Free University of Berlin.	14	1	[26]
3	Sys-BodyFluid Database	This body fluid protein database,Sys-BodyFluid, contains 11 body fluid proteomes, including plasma/serum, urine, cerebrospinal fluid, saliva, bronchoalveolar lavage fluid, synovial fluid, nipple aspirate fluid, tear fluid, seminal fluid, human milk, and amniotic fluid. Over 10,000 proteins are included in the Sys-BodyFluid. These body fluid proteome data come from 50 peer-review publications of different laboratories all over the world. (http://www.biosino.org/bodyfluid)	8	1	-
4	Proteomics database for dental tissue proteins	(http://toothprint.mdhs.unimelb.edu.au/fmi/xsl/toothprint/home. xsl)	6	3	[27]
5	HSC-2DPAGE	The Protein Biochemistry Group carries out research into alterations in protein expression in heart disease and in response to cardiac transplantation. (http://www.doc.ic.ac.uk/vip/hsc-2dpage/)	6	4	[28]
6	Mitochondrial Proteome	The aim of this database is to provide a comprehensive list of mitochondrial proteins of yeast, mouse, <i>Arabidopsis thaliana</i> , neurospora and human. Datasets relevant to the study of the mitochondrial proteome are integrated and accessible via search tools and links. They include computational predictions of signaling sequences and summarized results from proteome mapping, mutant screening, expression profiling, protein-protein interaction and cellular sublocalization studies. Predictive scores (score based on support vector machine prediction) are given and are calculated using the integration of the included datasets and annotated reference sets of mitochondrial proteins. The reference sets are regularly updated.	5	5	
7	GeMDBJ Proteomics	<ul> <li>Genome Medicine Database of Japan Proteomics (GeMDBJ Proteomics) is an integrated proteome database useful in cancer research. The following information is available in the database:</li> <li>Expression levels of proteins identified in surgical specimens and cell lines by 2D-DIGE.</li> <li>Protein spots identified by liquid chromatography tandem mass spectrometry (LC-MS/MS), with corresponding peptide information.</li> <li>Clinical and biological information pertaining to the samples.</li> <li>GeMDBJ Proteomics has been developed as part of the Proteome Bioinformatics Project of the National Cancer Center Research Institute, Tokyo. This database aims to enhance proteome-based research for the benefit of human health and welfare, and will keep expanding as new proteome information becomes available.</li> <li>(https://gemdbj.nibio.go.jp/dgdb/DigeTop.do)</li> </ul>	5	1	
8	MitoProteome Database	MitoProteome is an object-relational database of human mitochondrial protein sequences generated from information obtained from a comprehensive curation of public databases and direct experimental evidence as well. MitoProteome contains both mitochondrial- and nuclear-encoded protein sequences.	1	1	[29]

		The initial release (2004) contained 847 human mitochondrial proteins, 615 of which had been identified by LC/MS studies of human heart mitochondria.			
9	Adipocyte Proteome Database	Adipocyte Proteome is a Database of the Department for Proteomics and Signal Transduction at the Max-Planck-Institut for Biochemistry. (http://141.61.102.16/)	1	1	[30]
10	Heart High- Performance 2-DE Database	Heart High-Performance 2-DE Database contains about 3300 proteins of ventricle resolved by large gels. The database includes the carefully examined data published in 1994, new entries of proteins identified in the lab, and proteins identified by comparison of other 2-DE heart databases. The identification of proteins was improved by combination of N-terminal and internal sequencing, amino acid analysis, immunostaining, MALDI-MS peptide mass fingerprinting, post source decay MALDI-MS, and ladder sequencing by carboxypeptidase P. (http://web.mpiib-berlin.mpg.de/hp-2dpage/)	1	1	[31]
11	OGP-WWW	(http://proteomewww.glycob.ox.ac.uk/2d/2d.html)	1	1	-
12	Proteomic unit	This proteomic resource contains information regarding 2-DE images and clickable maps and experimental results emphasizing the application of two-dimensional electrophoresis and protein mapping (2-DE) to different proteomic projects.	1	3	[32]
13	KAIKO 2DDB	Two-dimensional polyacrylamide gel electrophoresis federated database for the analysis of silkworm protein from National Institute of Agrobiological Sciences, Japan. (http://kaiko2ddb.dna.affrc.go.jp/cgi-bin/search_2DDB.cgi)	0	1	-
14	Nucleolar Proteome Database	This site contains all of the data from Lamond and Mann laboratories' ongoing proteomic analysis of human nucleoli. Using high sensitivity mass spectrometry and stringent criteria, so far ~700 human nucleolar proteins have been identified. More recently, they used a quantitative proteomic approach for the temporal characterization of protein flux through the nucleolus, determining the kinetics of up to 489 nucleolar proteins after various drug treatments and demonstrating that there is no unique, complete proteome for the nucleolus, but rather an overlapping set of proteomes that are relevant to different cell states or conditions. The data can be searched in a variety of ways (see below). (http://www.lamondlab.com/NOPdb/)	0	1	-
15	TMIG-2DPAGE database for researches on Gerontology and Geriatrics at TMIG Proteomics Collaboration Center	The concept of this proteome database is the research on cellular aging was reported in the INABIS '98 Symosium. (http://proteomeback.tmig.or.jp/2D/index.html)	0	1	-
16	MGH-PGA Proteomic Tools	The MGH-PGA Proteomic Database is based on NCBI's protein database, with the additional information about coding DNA sequences. This database was constructed by parsing the human and mouse GenPept release from NCBI. Very short peptide sequences (< 23 amino acids) were excluded from this database. Most immunoglobulin and T cell receptor variable region genes were also excluded to produce a better non-redundant database. The corresponding DNA coding regions were then retrieved from GenBank via the CDS feature and were included in the database.	0	2	-

		(http://pga.mgh.harvard.edu/servlets/)			
17	Cornea 2D-PAGE	Human Cornea 2D-PAGE was constructed and maintained by Sivert Bertelsen, using the Make2D-DB II package (ver. 1.00.a) from the WORLD-2DPAGE of the ExPASy web server. (http://www.cornea-proteomics.com/)	0	1	-
18	DOSAC-COBS 2D Page	University of Palermo, Italy Human contains various breast cancer samples (cells, serum, etc.), dermal fibroblasts, mammary cells, lymphocytes, U937 and HL60 cell lines. (http://www.dosac.unipa.it/cgi-bin/2d/2d.cgi)	0	1	-
19	REPRODUCTION- 2DPAGE	(http://reprod.njmu.edu.cn/cgi-bin/2d/2d.cgi)	0	2	-
20	HUVEC.COM	Proteomics of Human Umbilical Vein Endothelial Cell at Hopital Saint-Antoine, Paris, France) Human Umbilical vein endothelial cell. (http://www.huvec.com/)	0	1	-
21	Tooth Pulp 2-D	(http://www.biochem.oulu.fi/proteomics/teeth.html)	0	2	[33]
22	The LECB 2-D PAGE Gel Images Data Sets	Includes Human leukemias (AML, ALL, CLL, HCL and other), HL-60 cell line, Molt-4 cell line, serum (Fetal Alchohol Syndrome, Rett Syndrome), Urine (cadmium-exposure, mercury-exposure) - 500 gel images. (http://www.lecb.ncifcrf.gov/2DgelDataSets/)	0	1	-