Chapter 2

Proteomic Data Storage and Sharing

Shivakumar Keerthikumar and Suresh Mathivanan

Abstract

With the advent of high-throughput genomic and proteomic techniques, there is a massive amount of multidimensional data being generated and has increased several orders of magnitude. But the amount of data that is cataloged in the central repositories and shared publicly with the scientific community does not correlate the same rate at which the data is generated. Here, in this chapter, we discuss various proteomics data repositories that are freely accessible to the researchers for further downstream meta-analysis.

Key words Proteins, Peptides, Databases, False discovery rate, Cancer, Mass spectrometry

1 Introduction

The applications of mass spectrometry in identification and quantification of proteins in complex biological samples is rapidly evolving [1–3]. Recent technical advances in mass spectrometer to measure the abundance of proteins have further increased the amount of multidimensional data being generated [4]. As a result, significant interests have been created to characterize the proteome of many cell types and subcellular organelles [5–9]. There are three different layers of proteomic data that is generated using mass spectrometry-based techniques: raw data, peptide/protein data (also known as "result" or "peak list") and metadata. Raw data is basically a binary format file which most of the proteomic tools like MSConvert (http://proteowizard.sourceforge.net/tools.shtml) converts further into human readable formats such as mgf, XML, pkl, and txt files. Metadata contains experimental details, type of instruments, modifications and search engines/tools used [10]. In order to disseminate these different types of data to the scientific community, researchers have constantly thrived to develop central repository to store and share these humongous data [11–13].

Here, we focus on publicly available free centralized resources that disseminate all kinds of proteomics data and tools which further aid in downstream analysis to gain new biological insights that benefit the scientific community.

2   Online Proteomics Community Resources

Currently, there are wide varieties of online resources (Table 1) that host different types of proteomics data at different level and software tools to further mine these data. The most commonly and widely used proteomic resources are discussed here.

2.1   PRoteomics IDEntifications (PRIDE) Database

The PRIDE database is most widely used centralized, publicly available proteomic repository which stores and manages all three different levels of proteomic data such as raw data, peak list file and metadata. The PRIDE database established at European Bioinformatics Institute, UK has a Web-based, user-friendly query and data submission system as well as documented application programming interface besides local installation [14]. Recently, the new PRIDE archival system (http://www.ebi.ac.uk/pride/archive/) has replaced the PRIDE database. The PRIDE archive system supports community recommended Proteomic Standard Initiative (PSI) data formats and is an active founding member of ProteomeXchange (PX) consortium (http://www.proteomicexchange.org/). The main concept behind such consortium is to standardize the mass spectrometry proteomics data and automate the sharing of these data between the repositories to benefit the end users [15].

The PRIDE archive system also stores many software tools such as PRIDE Inspector, PRIDE converter and PX submission tool to further streamline the data submission process and its visualization to aid scientific community. All these software tools including Web modules are developed in JAVA and are open source (https://code.google.com/archive/p/ebi-pride/). Besides funding agencies, many scientific journals such as *Nature Biotechnology, Proteomics, Molecular and Cellular Proteomics* and *Journal of Proteome Research* mandates submission of raw data and associated metadata to proteomics repository to support their publication which further elevated the public deposition of proteomics data. As a result, The PRIDE archive currently contains ~140 TBs size of data which constitutes 690 M spectra, 298 M and 66 M peptide and protein identification, respectively, spanning more than 500 different taxonomical identifiers.

2.2   PeptideAtlas

The PeptideAtlas (http://www.peptideatlas.org/) database is another freely available mass spectrometry derived proteomic data repository developed at Institute of Systems Biology, Seattle, USA.

Table 1
Overview of online proteomics resources

Database	Types of data stored	Link
PRIDE	Accepts Raw data, processed data and meta data	http://www.ebi.ac.uk/pride/archive/
Peptide Atlas	Accepts only Raw data and limited meta data	http://www.peptideatlas.org/
CPTAC	Allows download and dissemination of raw data, processed data and meta data relevant to cancer biospecimens collated through Proteomic Characterization centers (PCCs)	http://proteomics.cancer.gov/
Colorectal Cancer Atlas	Stores processed protein and peptide data after automatically analyzing the publicly available raw data from the proteomic repositories	http://www.colonatlas.org/
GPMDB	Stores processed protein and peptide data after automatically analyzing the publicly available raw data from the proteomic repositories. Supports data analysis	http://www.thegpm.org/
ProteomicsDB	Accepts Raw data, processed data and meta data. Allows download of raw data, processed protein and peptide data.	http://www.proteomicsdb.org/
Human Proteome Map	Allows download of processed protein and peptide data.	http://www.humanproteomemap.org/
Human Proteinpedia	Accepts processed and meta data.	http://www.humanproteinpedia.org/
Human Protein Atlas	Allows download of protein and RNA expression in normal and tumor tissues and cell types	http://www.proteinatlas.org/

Represents list of publicly available online proteomics resources and repositories discussed in this chapter

The PeptideAtlas accepts only spectra files either in the form of RAW, mzML or mzXML format and limited metadata. Once submitted, the raw spectra files are processed using standardized data processing pipeline known as Trans Proteomics Pipeline (TPP) [16] and stored in the SBEAMS (Systems Biology Experiment Analysis Management System)-Proteomics module. Further, peptides identified with high score are mapped to their respective genome sequence representing species/sample specific build [17, 18]. Currently, the PeptideAtlas has 19 organism specific build which includes many model organisms such as human, yeast, worms,

mouse, fly, rat, horse, and zebrafish, for important sample groups such as plasma, brain, liver, lung, colon cancer, heart, kidney, and urine.

The PeptideAtlas, similar to the PRIDE archive system, is one of the founding members of PX consortium that implemented standardization of the mass spectrometry-based proteomics data and automate the sharing of proteomic data across different repositories. Another important feature of the PeptideAtlas is investigation of proteotypic peptides which are defined as peptides that can uniquely and unambiguously identify specific protein. Currently, users can search proteotypic peptides from three different organisms such as human, mouse, and yeast. Identification of such high scoring peptides would further serve as most possible targets for Selected Reaction Monitoring (SRM) approach [19]. The PeptideAtlas SRM Experiment Library (PASSEL) is a component of the PeptideAtlas project that is designed to enable submission, dissemination, and reuse of SRM experimental results from analysis of biological samples. The raw data submitted via PASSEL are automatically processed and stored into the database which can be further downloaded or accessed via web interface [20].

Further, the distinct peptides and its associated proteins identified from the user submitted raw data files using TPP tool can be further depicted graphically in Cytoscape [21] plugin implemented in the PeptideAtlas. Overall, the PeptideAtlas depicts the normalized outlook of the user submitted data which further aid in genome annotation of different organisms using mass spectrometry derived proteomic data.

2.3   CPTAC (Clinical Proteomic Tumor Analysis Consortium) Portal

The CPTAC data portal (http://proteomics.cancer.gov/) launched in August 2011 by National Cancer Institute (NCI) is a freely available, centralized public proteomic data repository collected by proteomic characterization centers for the CPTAC framework. The proteomic characterization center constitutes of five teams namely Broad Institute of MIT and Harvard/Fred Hutchinson Cancer Research Center, Johns Hopkins University, Pacific Northwest National Laboratory, Vanderbilt University, and Washington University/University of North Carolina. The proteome characterization center implements proteomics candidate developmental pipeline for further protein identification and its verification to serve as high value targets for clinically useful diagnostics. In addition, proteomic data from The Cancer Genome Atlas (TCGA) data portal (http://cancergenome.nih.gov/), xenograft models and other tissue datasets of well-characterized genome using standardized Common Data Analysis Pipeline are analyzed to increase the significance of the results. The CPTAC data portal hosts mass spectrometry data of cancer biospecimens such as breast, colorectal, and ovarian cancer as well as global profiling of post-translational modifications of tumor tissues and

cancer cell lines which accounts to more than 6 TB data. The CPTAC data portal also hosts data from the Clinical Proteomic Technologies for Cancer Initiative from 2006 to 2011, which was mainly developed to address the pre-analytical and analytical variability issues that are major barriers in the field of proteomics. The major outcome of this program was the launch of the CPTAC data portal to understand the molecular basis of cancer using proteomic technology [22, 23].

2.4 Colorectal Cancer Atlas

Colorectal Cancer Atlas (http://www.colonatlas.org/) is web-based resource developed by integrating genomic and proteomic annotations identified precisely in colorectal cancer tissues and cell lines. It integrates heterogeneous data freely available in the public repositories, published articles [24] and in-house experimental data pertaining to quantitative and qualitative protein expression data obtained from variety of techniques such as mass spectrometry, western blotting, immunohistochemistry, confocal microscopy, immunoelectron microscopy, and fluorescence-activated cell sorting. Colorectal Cancer Atlas collates raw proteomic mass spectrometry and other proteomic experimental data specifically from colorectal cancer tissues and cell lines is processed using in-house pipeline. The proteins/peptides identified after <5 % FDR cutoff is stored in the backend database. Besides, mutation data largely obtained by large and small sequencing methods are also incorporated into the Colorectal Cancer Atlas database [13].

Currently, Colorectal Cancer Atlas hosts >62,000 protein identifications, >8.3 million MS/MS spectra, >13,000 colorectal cancer tissues and >209 cell lines. Further, Colorectal Cancer Atlas facilitate users to visualize these proteins identified in context of signaling pathways, protein–protein interactions, gene ontology terms, protein domains, and posttranslational modifications. Users can download the entire colorectal cancer data in tab-delimited format using the download page at http://colonatlas.org/download/.

2.5 Global Proteome Machine Database (GPMDB)

The Global Proteome Machine Database (http://www.thegpm.org/) is another open source mass spectrometry based proteomic repository, publicly available for the scientific community. The GPMDB periodically checks all the public proteomic repositories, downloads and reanalyzes the proteomic data using X! Tandem search engine. Besides, the users can also use spectral search engine called X! Hunter (http://xhunter.thegpm.org/) and proteotypic profiler called X! P3 (http://p3.thegpm.org/) [25] to analyze their data. The resultant peptide and protein lists after passing through the stringent automated quality test are stored into the backend database along with relevant metadata. Further, the results can be either viewed in the GPM website or downloaded through ftp or other interfaces. Besides, the users can

also submit their spectra files in different formats such as mgf, mzXML, pkl, mzData, dta, and common (for only big and compressed files) to GPM via 'Search Data' option available in the website. The most frequently checked public repositories for the suitable new proteomic data for reanalysis includes Proteome Xchange/PRIDE, PeptideAtlas/PASSEL, MassIVE (http://www.massive.ucsd.edu/), Proteomics DB, The Chorus Project (http://chorusproject.org/), and iProX (http://www.iprox.org/).

Recently, at the time of writing this chapter, the GPMDB released an updated version of the GPM Personal Edition-Fury to replace the old venerable Cyclone version and upgraded to the latest version of X! Tandem (Version 2015.12.15, Vengeance) which features speedy assignment of PTMs. In addition, the human and mouse protein identification information in GPMDB has been summarized into a collection of spreadsheets known as GPMDB Guide to Human Proteome (GHP) and GPMDB Guide to Mouse Proteome (GMP), respectively. This guide contains information organized into separate spreadsheets for each chromosome as well as mitochondrial DNA and made available for download at ftp://ftp.thegpm.org/projects/annotation/human_protein_guide/ and ftp://ftp.thegpm.org/projects/annotation/proteome_protein_guide/.

2.6 ProteomicsDB

ProteomicsDB (http://www.proteomicsdb.org/) is a human centric proteomic data repository developed jointly by Technical University Munich (TUM) and company SAP SE (Walldorf, Germany) and SAP Innovation Center and Cellzome GmbH (GSK Company). ProteomicsDB, an in-memory database, configured with 2 TB of random access memory (RAM) and 160 central processor units (CPU), designed for real-time analysis of big proteomic data. ProteomicsDB assembles raw proteomic data files from public repositories such as PRIDE, PeptideAtlas, MassIVE, ProteomeXchange, and many other individual laboratories as well as from in-house experiments and reprocess the files using MaxQuant [26] and MASCOT [27] software packages. The proteins and peptides identified after passing through quality control steps including FDR filters are deposited into ProteomicsDB.

ProteomicsDB came into the limelight in 2014 with the release of draft human proteome map assembled using mass spectrometry experiments on human tissues, cell lines, body fluids as well as data from PTM studies and affinity purifications [3]. Currently, at the time of writing, ProteomicsDB contains protein evidence for 15,721 of the 19,629 protein coding genes which constitutes 80 % coverage of human proteome. ProteomicsDB has a Web-based user-friendly interface through which users can search and download details of particular protein and peptide sequence via 'browse by proteins' and 'browse by chromosomes' options. Besides, users

can submit their raw mass spectrometry data files, peak list files and metadata associated with it only after creating a user account in the ProteomicsDB. The secure URL link generated. At the time of writing, there were more than 569 registered users, 76 projects and more than 400 experiments accounting to 7 TB of data in ProteomicsDB.

2.7 Human Proteome Map (HPM)

The Human Proteome Map (HPM) (http://www.humanproteomemap.org/) was developed to represent the draft study of human proteome map. The HPM database hosts high-resolution mass spectrometry proteomic data representing 17 adult tissues, six primary hematopoietic cells, and seven fetal tissues resulting in >84 % human proteome coverage. The mass spectrometry data was searched against Human RefSeq database (version 50 with common contaminants) using SEQUEST (http://fields.scripps.edu/sequest/) and MASCOT [27] search engines through Proteome Discoverer 1.3 platform (Thermo Scientific, Bremen, Germany). The peptides and proteins identified were represented as normalized spectral counts and for each peptide the high resolution MS/MS spectrum for the best scoring peptides can be visualized using Lorikeet JQuery plugin (http://uwpr.github.io/Lorikeet/). The results of the proteins and peptides can be queried and downloaded in the standard formats, but the databases currently do not support the submission of any new proteomic data [2].

2.8 Human Proteinpedia

Human Proteinpedia (http://humanproteinpedia.org/) [28, 29] was developed in 2008 [2] to facilitate the sharing and integration of human proteomic data. Besides, it allows scientific community to contribute and maintain protein annotations using protein distributed annotation system also known as PDAS. Further, protein annotations submitted by the users are mapped to individual proteins and made available using Human Protein Reference database (HPRD: http://www.hprd.org/) [30]. This allows the user to visualize experimentally validated protein–protein interaction networks, protein expressions in cell lines/tissues, post-translational modifications and subcellular localizations besides mass spectrometry derived peptides/proteins and spectral details.

Human Proteinpedia enables users to query at gene/protein level, by types of tissue expressions, posttranslational modifications, subcellular localizations, different mass spectrometer types, and experimental platforms. Using PDAS, the users are allowed to upload only processed data (peak list files) and meta-data containing experimental details into the back-end database either using normal or batch (for high-throughput data) upload system. The entire Human Proteinpedia data can be further downloaded freely by the scientific community at http://www.humanproteinpedia.org/download/ [31].

Currently, more than 240 different laboratories around the world has contributed proteomic data into Human Proteinpedia database which resulted in >4.8 M MS/MS spectra, >1.9 M peptide identifications, >150,000 protein expressions, >17,000 posttranslational modifications, >34,000 protein–protein interactions, and >2900 subcellular localizations from >2700 proteomic experiments.

2.9 Human Protein Atlas

The Human Protein Atlas (HPA: http://www.proteinatlas.org/) hosts expression and localization of majority of human protein coding genes based on both RNA and protein data. It was developed in 2005 as a large scale effort to quest where the proteins encoded by the human protein coding genes are expressed in the different tissues and cell types. Unlike other proteomic resources mainly depends on mass spectrometry based protein identifications, the HPA largely uses antibody based proteomics and transcriptomics profiling methods to locate and identify proteins in tissues and cell types. The transcriptomic data quantifies gene expression levels on different tissues and cell types while antibody based protein profiling methods characterize spatial cellular distribution for the corresponding proteins at different substructures and cell types of the tissues [32].

At the time of writing this chapter, the Human Protein Atlas version 14 known to contain RNA data for 99 % and protein data for 86 % of the predictive human genes and includes >11 million images with primary data from immunohistochemistry and immunofluorescence. The HPA contains >37,000 validate antibodies corresponding to 17,000 human protein coding genes collated from 46 human cell lines and tissue samples from 360 people (44 normal tissue types from 144 people and the 20 most common types of cancer from 216 people) [33].

Recently, tissue-based map of the human proteome data analyzed from 32 tissues and 47 cell lines using integrated OMICS approach is included in the Human Protein Atlas to further explore the expression pattern across the human body. In addition, global analysis of secreted and membrane proteins (secretome and membrane proteome), as well as an analysis of expression profiles for all proteins targeted by pharmaceutical drugs (druggable proteome) and protein implicated in cancer (cancer proteome) is integrated into the Human Protein Atlas [9].

3   Discussion

The amount of proteomics data being shared among the scientific community is still not well organized when compared to the humongous data that is being generated due to advancement in the proteomics field. The main reason for this can be attributed to the limited funding available for the maintenance of the database server, manpower, and other infrastructure. As a result, few of the

efficient repositories such as NCBI Peptidome [34, 35] and Tranche [10] are completely discontinued largely due to funding constraints. In order to sustain and serve the growing scientific community database like the CHORUS (https://chorusproject. org/), a cloud based platform for storage, analysis and sharing of mass spectrometry data is charging users with certain amount of fees based on type of services required. We urge the continuous usage of these proteomic resources and willingness to share the proteomics data to the scientific community will only keep these resources alive and stable. Further, these proteomics resources would aid as important discovery tools in the field of biomedical research.

References

1. Mathivanan S (2014) Integrated bioinformatics analysis of the publicly available protein data shows evidence for 96% of the human proteome. J Proteomics Bioinform 07:041–049. doi:10.4172/jpb.1000301

2. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A (2014) A draft map of the human proteome. Nature 509(7502):575–581. doi:10.1038/nature13302

3. Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese JH, Bantscheff M, Gerstmair A, Faerber F, Kuster B (2014) Mass-spectrometry-based draft of the human proteome. Nature 509 (7502):582–587. doi:10.1038/nature13319

4. Lesur A, Domon B (2015) Advances in high-resolution accurate mass spectrometry application to targeted proteomics. Proteomics 15 (5-6):880–890. doi:10.1002/pmic.201400450

5. Keerthikumar S, Gangoda L, Liem M, Fonseka P, Atukorala I, Ozcitti C, Mechler A, Adda CG, Ang CS, Mathivanan S (2015) Proteogenomic analysis reveals exosomes are more oncogenic than ectosomes. Oncotarget 6: 15375–15396

6. Onjiko RM, Moody SA, Nemes P (2015) Single-cell mass spectrometry reveals small molecules that affect cell fates in the 16-cell embryo. Proc Natl Acad Sci U S A 112(21): 6545–6550. doi:10.1073/pnas.1423682112

7. Lydic TA, Townsend S, Adda CG, Collins C, Mathivanan S, Reid GE (2015) Rapid and comprehensive 'shotgun' lipidome profiling of colorectal cancer cell derived exosomes. Methods 87:83–95. doi:10.1016/j.ymeth.2015.04.014

8. Habuka M, Fagerberg L, Hallstrom BM, Ponten F, Yamamoto T, Uhlen M (2015) The urinary bladder transcriptome and proteome defined by transcriptomics and antibody-based profiling. PLoS One 10(12):e0145301. doi:10.1371/journal.pone.0145301

9. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F (2015) Proteomics tissue-based map of the human proteome. Science 347(6220):1260419. doi:10.1126/science.1260419

10. No Authors Listed (2012) A home for raw proteomics data. Nat Methods 9(5):419

11. Keerthikumar S, Chisanga D, Ariyaratne D, Al Saffar H, Anand S, Zhao K, Samuel M, Pathan M, Jois M, Chilamkurti N, Gangoda L, Mathivanan S (2016) ExoCarta: a Web-based compendium of exosomal cargo. J Mol Biol 428(4):688–692. doi:10.1016/j.jmb.2015.09.019

12. Keerthikumar S, Raju R, Kandasamy K, Hijikata A, Ramabadran S, Balakrishnan L, Ahmed M, Rani S, Selvan LD, Somanathan DS, Ray S, Bhattacharjee M, Gollapudi S, Ramachandra YL, Bhadra S, Bhattacharyya C, Imai K, Nonoyama S, Kanegane H, Miyawaki T, Pandey A, Ohara O, Mohan S (2009) RAPID: resource of Asian primary immunodeficiency diseases. Nucleic Acids Res 37(Database issue):D863–D867. doi:10.1093/nar/gkn682

13. Chisanga D, Keerthikumar S, Pathan M, Ariyaratne D, Kalra H, Boukouris S, Mathew NA, Saffar HA, Gangoda L, Ang CS, Sieber OM, Mariadason JM, Dasgupta R, Chilamkurti N, Mathivanan S (2016) Colorectal cancer atlas: an integrative resource for genomic and proteomic annotations from colorectal cancer cell lines and tissues. Nucleic Acids Res 44(D1):D969–D974. doi:10.1093/nar/gkv1097

14. Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, O'Kelly G, Schoenegger A, Ovelleiro D, Perez-Riverol Y, Reisinger F, Rios D, Wang R, Hermjakob H (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 41(Database issue):D1063–D1069. doi:10.1093/nar/gks1262

15. Vizcaino JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H (2016) 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 44(D1):D447–D456. doi:10.1093/nar/gkv1145

16. Deutsch EW, Mendoza L, Shteynberg D, Slagel J, Sun Z, Moritz RL (2015) Trans-proteomic pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics Clin Appl 9(7-8):745–754. doi:10.1002/prca.201400164

17. Farrah T, Deutsch EW, Hoopmann MR, Hallows JL, Sun Z, Huang CY, Moritz RL (2013) The state of the human proteome in 2012 as viewed through PeptideAtlas.

J Proteome Res 12(1):162–171. doi:10.1021/pr301012j

18. Vizcaino JA, Foster JM, Martens L (2010) Proteomics data repositories: providing a safe haven for your data and acting as a springboard for further research. J Proteomics 73(11):2136–2146. doi:10.1016/j.jprot.2010.06.008

19. Pan S, Aebersold R, Chen R, Rush J, Goodlett DR, McIntosh MW, Zhang J, Brentnall TA (2009) Mass spectrometry based targeted protein quantification: methods and applications. J Proteome Res 8(2):787–797. doi:10.1021/pr800538n

20. Farrah T, Deutsch EW, Kreisberg R, Sun Z, Campbell DS, Mendoza L, Kusebauch U, Brusniak MY, Huttenhain R, Schiess R, Selevsek N, Aebersold R, Moritz RL (2012) PASSEL: the PeptideAtlas SRMexperiment library. Proteomics 12(8):1170–1175. doi:10.1002/pmic.201100515

21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504. doi:10.1101/gr.1239303

22. Ellis MJ, Gillette M, Carr SA, Paulovich AG, Smith RD, Rodland KK, Townsend RR, Kinsinger C, Mesri M, Rodriguez H, Liebler DC, Clinical Proteomic Tumor Analysis C (2013) Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. Cancer Discov 3(10):1108–1112. doi:10.1158/2159-8290.CD-13-0219

23. Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, Madhavan S, Ketchum KA (2015) The CPTAC data portal: a resource for cancer proteomics research. J Proteome Res 14(6):2707–2713. doi:10.1021/pr501254j

24. Mathivanan S, Ji H, Tauro BJ, Chen YS, Simpson RJ (2012) Identifying mutated proteins secreted by colon cancer cell lines using mass spectrometry. J Proteomics 76:141–149. doi:10.1016/j.jprot.2012.06.031

25. Craig R, Cortens JP, Beavis RC (2005) The use of proteotypic peptide libraries for protein identification. Rapid Commun Mass Spectrom 19(13):1844–1850. doi:10.1002/rcm.1992

26. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372. doi:10.1038/nbt.1511

27. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein

identification by searching sequence databases using mass spectrometry data. Electrophoresis 20(18):3551–3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

28. Mathivanan S, Ahmed M, Ahn NG, Alexandre H, Amanchy R, Andrews PC, Bader JS, Balgley BM, Bantscheff M, Bennett KL, Bjorling E, Blagoev B, Bose R, Brahmachari SK, Burlingame AS, Bustelo XR, Cagney G, Cantin GT, Cardasis HL, Celis JE, Chaerkady R, Chu F, Cole PA, Costello CE, Cotter RJ, Crockett D, DeLany JP, De Marzo AM, DeSouza LV, Deutsch EW, Dransfield E, Drewes G, Droit A, Dunn MJ, Elenitoba-Johnson K, Ewing RM, Van Eyk J, Faca V, Falkner J, Fang X, Fenselau C, Figeys D, Gagne P, Gelfi C, Gevaert K, Gimble JM, Gnad F, Goel R, Gromov P, Hanash SM, Hancock WS, Harsha HC, Hart G, Hays F, He F, Hebbar P, Helsens K, Hermeking H, Hide W, Hjerno K, Hochstrasser DF, Hofmann O, Horn DM, Hruban RH, Ibarrola N, James P, Jensen ON, Jensen PH, Jung P, Kandasamy K, Kheterpal I, Kikuno RF, Korf U, Korner R, Kuster B, Kwon MS, Lee HJ, Lee YJ, Lefevre M, Lehvaslaiho M, Lescuyer P, Levander F, Lim MS, Lobke C, Loo JA, Mann M, Martens L, Martinez-Heredia J, McComb M, McRedmond J, Mehrle A, Menon R, Miller CA, Mischak H, Mohan SS, Mohmood R, Molina H, Moran MF, Morgan JD, Moritz R, Morzel M, Muddiman DC, Nalli A, Navarro JD, Neubert TA, Ohara O, Oliva R, Omenn GS, Oyama M, Paik YK, Pennington K, Pepperkok R, Periaswamy B, Petricoin EF, Poirier GG, Prasad TS, Purvine SO, Rahiman BA, Ramachandran P, Ramachandra YL, Rice RH, Rick J, Ronnholm RH, Salonen J, Sanchez JC, Sayd T, Seshi B, Shankari K, Sheng SJ, Shetty V, Shivakumar K, Simpson RJ, Sirdeshmukh R, Siu KW, Smith JC, Smith RD, States DJ, Sugano S, Sullivan M, Superti-Furga G, Takatalo M, Thongboonkerd V, Trinidad JC, Uhlen M, Vandekerckhove J, Vasilescu J, Veenstra TD, Vidal-Taboada JM, Vihinen M, Wait R, Wang X, Wiemann S, Wu B, Xu T, Yates JR, Zhong J, Zhou M, Zhu Y, Zurbig P, Pandey A (2008) Human proteinpedia enables sharing of human protein data. Nat

Biotechnol 26(2):164–167. doi:10.1038/nbt0208-164

29. Kandasamy K, Keerthikumar S, Goel R, Mathivanan S, Patankar N, Shafreen B, Renuse S, Pawar H, Ramachandra YL, Acharya PK, Ranganathan P, Chaerkady R, Keshava Prasad TS, Pandey A (2009) Human proteinpedia: a unified discovery resource for proteomics research. Nucleic Acids Res 37 (Database issue):D773–D781. doi:10.1093/nar/gkn701

30. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A (2009) Human protein reference database—2009 update. Nucleic Acids Res 37 (Database issue):D767–D772. doi:10.1093/nar/gkn892

31. Muthusamy B, Thomas JK, Prasad TS, Pandey A (2013) Access guide to human proteinpedia. Curr Protoc Bioinformatics 1:121. doi:10.1002/0471250953.bi0121s41

32. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, Wernerus H, Bjorling L, Ponten F (2010) Towards a knowledge-based human protein Atlas. Nat Biotechnol 28(12):1248–1250. doi:10.1038/nbt1210-1248

33. Marx V (2014) Proteomics: an atlas of expression. Nature 509(7502):645–649. doi:10.1038/509645a

34. Slotta DJ, Barrett T, Edgar R (2009) NCBI peptidome: a new public repository for mass spectrometry peptide identifications. Nat Biotechnol 27(7):600–601. doi:10.1038/nbt0709-600

35. Csordas A, Wang R, Rios D, Reisinger F, Foster JM, Slotta DJ, Vizcaino JA, Hermjakob H (2013) From peptidome to PRIDE: public proteomics data migration at a large scale. Proteomics 13(10-11):1692–1695. doi:10.1002/pmic.201200514

![Springer logo](horse logo) **Springer**