Published in final edited form as:

J Proteome Res. 2007 October ; 6(10): . doi:10.1021/pr070325f.

YPED: A Web-Accessible Database System for Protein Expression Analysis

Mark A. Shifman^{*,†,‡}, **Yuli Li**^{†,‡}, **Christopher M. Colangelo**^{§,||}, **Kathryn L. Stone**^{§,||}, **Terence L. Wu**^{§,||}, **Kei-Hoi Cheung**^{†,‡,⊥,#}, **Perry L. Miller**^{†,‡,¥}, and **Kenneth R. Williams**^{§,||}

Center for Medical Informatics, Department of Anesthesiology, Keck Biotechnology Resource Laboratory, Department of Genetics, Department of Molecular, Cellular and Developmental Biology, Department of Molecular Biophysics and Biochemistry, and Department of Computer Science, Yale University, New Haven, Connecticut

Abstract

We have developed an integrated web-accessible software system called the Yale Protein Expression Database (YPED) to address the need for storage, retrieval, and integrated analysis of large amounts of data from high throughput proteomic technologies. YPED is an open source system which integrates gel analysis results with protein identifications from DIGE experiments. The system associates the DIGE gel spots and image, analyzed with DeCyder, with mass spectrometric protein identifications from selected gel spots. Following in gel trypsin digestion, proteins in spots of interest are analyzed using MALDI-TOF/TOF on an AB 4700 or, more recently, on an AB 4800 with protein identifications performed by Mascot in conjunction with the AB GPS Explorer system. In addition to DIGE, YPED currently handles protein identifications from MudPIT, iTRAQ, and ICAT experiments. Sample descriptions are compatible with the evolving MIAPE standards. Tandem MS/MS results from MudPIT, and ICAT analyses are validated with the Trans-Proteomic Pipeline and then stored in the database for viewing and linking to the identified proteins. Researchers can view, subset, and download their data through a secure Web interface that includes a table containing proteins identified, a sample summary, the sample description, and a clickable gel image for DIGE samples. Tools are available to facilitate sample comparison and the viewing of phosphoproteins. A summary report with PANTHER Classification System annotations is also available to aid in biological interpretation of the results. The source code is open-source and is available from http://yped.med.yale.edu/yped_dist.

Keywords

protein expression database; mass spectrometry; MudPIT; DIGE; ICAT; iTRAQ; proteomics

© 2007 American Chemical Society

*To whom correspondence should be addressed. Mark Shifman, Yale Center for Medical Informatics, 300 George Street, Suite 501, New Haven, CT 06520; mark.shifman@yale.edu.

†Center for Medical Informatics.

‡Department of Anesthesiology.

§Keck Biotechnology Resource Laboratory.

||Department of Molecular Biophysics and Biochemistry.

⊥Department of Genetics.

#Department of Computer Science.

¥Department of Molecular, Cellular and Developmental Biology.

Introduction

Gene profiling methods (e.g., DNA microarrays¹) facilitate genome-scale analysis of transcriptional alterations occurring under different experimental or disease conditions, while protein profiling (an essential component of proteomics) characterizes on a proteome scale post-transcriptional and post-translational changes.^{2,3} In recent years, we have witnessed significant advances in realizing the great potential of biological and clinical applications of proteomics⁴ of which protein profiling plays a major role. Protein profiling can be achieved using several different experimental techniques typically involving the use of mass spectrometry. Some of these techniques are gel-based such as, DIGE⁵ (difference fluorescence 2D gel electrophoresis), while others such as MudPIT⁶ (Multidimensional Protein Identification Technology), ICAT⁷ (Isotope Coded Affinity Tag profiling), SILAC⁸ (Stable isotope labeling with amino acids in cell culture) and iTRAQ⁹ (Multiplexed Isobaric Tagging Technology) are not.

The need for proteomic data management has been recognized, as high-throughput proteomic technologies, such as mass spectrometry, are increasingly used for a wide variety of biomedical experiments.¹⁰ As the amount of proteomic data continues to grow, analysis and storage of the results have become a critical challenge. In the past several years, a number of proteomic databases have been developed. While some databases such as SBEAMS,¹¹ PEDro,¹² Proteios,¹³ CPAS,¹⁴ PRIDE¹⁵ and DBParser¹⁶ are more comprehensive with emphases on storing both raw and processed data, other databases such as WORLD-2DPAGE¹⁷ are designed for a specific proteomic technology (e.g., 2-D gel). As proteomic technologies are evolving, the data management requirements and needs will also change. The database design therefore should be flexible enough to handle such technological changes. For example, databases that are designed for 2-D gel electrophoresis will not be immediately applicable to capturing data generated using the more recently developed difference (fluorescence) 2-D gel electrophoresis (DIGE) technology. As other new protein profiling techniques arise, a database is needed that can handle data/protein identification results generated using such techniques. YPED is designed to capture DIGE data in addition to data generated by other protein profiling techniques including MudPIT, ICAT, and iTRAQ. YPED is unique in its ability to combine both DIGE gel analysis with mass spectrometry database results in a web accessible format.

Methods and Results

YPED is a web-accessible database for managing high throughput proteomic analyses. The system currently handles analysis requisition, result reporting and sample comparison for MudPIT, DIGE, iTRAQ, and ICAT samples. YPED will handle SILAC in the near future. Scientists can order analyses and describe samples via the web interface. Results are viewed via a secure web interface and can be downloaded as an Excel spreadsheet for local analysis. A demonstration login is available at <http://yped.med.yale.edu>.

YPED consists of three components: an Oracle database server, a web interface and an ftp file server. The database server, Oracle9i Enterprise Edition, stores the user information, sample descriptions, and results, (i.e., protein/peptide identifications and expression ratios). The web interface was developed using the Tomcat web server (current version 5.5) running on a RedHat Linux server. The web application was developed using Java 1.4.2 and the Struts Framework (current version 1.3.5). The ftp file server (whose current storage capacity is 4.5 TeraByte) houses the gel images and is an archive for the result files in XML format. The web interface is encrypted and transmitted using the Secure Sockets Layer (SSL) since some of the information provided by the user may be sensitive.

YPED's analysis requisition module interface facilitates a detailed description of samples by the users for each of the different types of requested proteomic experiments. Sample descriptions include information on experiment type, species, tissue type, etc. Quantitative information such as protein concentration and sample volume and amount are also recorded along with an optional text comment. For a DIGE experiment, as shown in Figure 1, specific information for running the gel such as pI range or percent polyacrylamide is included as is the fluorescent label for each of the individual samples that is to be pooled. Similarly, for ICAT and iTRAQ samples YPED stores the specific label used to modify each of the sample components. The data items captured are compatible with the MIAME¹⁸ and MIAPE¹⁹ guidelines. The vocabulary for describing organisms is a subset of the NCBI Taxonomy²⁰ and that for describing tissues is taken from the NCI Thesaurus.²¹

YPED can handle MudPIT, iTRAQ, and ICAT data generated with any type of LC-MS/MS instrument, including the Q-ToF type mass spectrometers in use in our laboratory. The tandem MS spectra are analyzed to identify the corresponding peptides and proteins using either the commercial database search programs SEQUEST²² and Mascot,²³ or the open source search program X!Tandem²⁴ in conjunction with pluggable scoring.²⁵ The Trans-Proteomic Pipeline²⁶ (TPP) developed by the Institute for Systems Biology is a suite of programs used in YPED to validate the database search results. This suite includes PeptideProphet²⁷ and ProteinProphet²⁸ which together compute probabilities of the indicated protein actually being present in the analyzed sample. These probability scores are incorporated into YPED. ICAT experiments can be quantified with either XPRESS²⁹ or ASAPRatio.³⁰ The output of the suite consists of the XML files, pepXML, and protXML, which are filtered and entered into YPED. For iTRAQ samples, quantitation is done using ProteinPilot (ABI) and the tab delimited result files are inserted into YPED.

Two-dimensional gel electrophoresis is performed using the DIGE platform with CyDye DIGE fluor labeling reagents, GE Healthcare Immobiline DryStrip Gels, the Ettan IPGphor IEF System, and pre-poured polyacrylamide gels from Jule Inc. The gel images are captured and analyzed using a GE Healthcare Typhoon 9410 Fluorescence Imager and DeCyder software. The differentially expressed spots that meet a predefined criterion are then picked using the GE Healthcare Ettan Spot Picker and analyzed using an AB 4700 or, more recently, an AB 4800 MALDI TOF/TOF mass spectrometer after in gel trypsin digestion on the Ettan TA Digester robot. For MALDI-TOF/TOF analysis, samples are named/identified in the AB 4000 Series Explorer Spot Set Manager using the following nomenclature: principal investigators name_ gel #_spot #_96 well plate position. The spot # is obtained from the DeCyder pick list which is outputted in a text file. The AB GPS Explorer software (v.3.6) is used to submit the combined MS and MS/MS data from the MALDI-TOF/TOF to a Mascot database search.

After analysis, the results from the image analysis of the DIGE gel (exported as an XML file) are merged with the proteins identified via the Mascot analysis of the MS/MS mass spectra (Figure 2). For a given gel, first the MALDI spots from the mass spectrometer are merged with the corresponding protein and peptide identifications — while retaining the spot number as an identifier. These identifications are then merged with the expression ratios contained in the DeCyder result XML file. The merged results are stored in an XML file which is then inserted into YPED.

The results can be viewed via the Web interface. The result page for each analysis type consists of a statistical overview summarizing the number of proteins, peptides and quantitation and a table of results containing the proteins identified. The result table can also be downloaded as Excel spreadsheets for local analysis. For MudPIT, ICAT, and iTRAQ results, a protein probability cutoff can be selected to restrict the number of proteins viewed.

The result table for these analyses contains the ProteinProphet probability, hyperlinks for obtaining more information on the identified proteins and for obtaining peptide sequence and quantitation details (ICAT or iTRAQ). Links are also available for viewing sample information and an image of the HPLC cation exchange chromatogram (fraction) trace, which is used as a first separation step in MudPIT, ICAT and iTRAQ analyses of complex cell and tissue extracts.

As shown by arrow (a) in Figure 3, the DIGE results can be queried based on Cy5/Cy3, ratios. In addition to the protein identifiers, the DIGE result table contains the spot number, Cy5/Cy3 or when Cy2 is utilized, the Cy5/Cy2 and Cy3/Cy2 ratios and specific information pertaining to the Mascot search. This includes the GPS Explorer scores derived from the Mascot scores for both the combined peptide mass fingerprint and MS/MS analysis search (db search score), as well as the score that resulted from only the analysis of the MS/MS data (total ion score). The peptide sequences identified are also listed by clicking on the "peptide #" link. The "Mascot detail" link lists information on the protein such as molecular weight and pI. The "spot number" link contains important information from the DeCyder analysis such as the individual Cy labeled spot volumes, peak heights, slopes, ratios and the x and y coordinates. The "Protein ID" link pulls up the indicated accession number from the database searched. As shown by arrow (b) in Figure 3, a link is available to view the DIGE gel image for the selected sample (Cy5/Cy3, Cy3/Cy2 or Cy5/Cy2). The image is annotated with colored circles indicating which spots were picked for analysis, with blue spots being up-regulated and red down regulated. Dark blue or red spots indicate spots from which proteins have been identified. Clicking on one of these dark blue/red circles returns specific information on the proteins identified in that spot as well as the spot number (see arrow c in Figure 3).

The interpretation of the proteomic results can be a daunting task, given the potentially large number of proteins identified as being differentially regulated in a given experiment. To aid in this task, the result pages offer the option of obtaining a summary of the results annotated by the PANTHER Classification System.³¹ We have developed a client to the PANTHER Classification System for viewing biological process, molecular function, or pathway summaries on a given result set of identified proteins. Protein identifiers for NCBItr, IPI databases and SwissProt are translated to their EntrezGene IDs and submitted to PANTHER where summary pie charts are generated. Protein identifiers from the Celera databases can also be submitted. We have implemented this seamless link between YPED and PANTHER using the Jakarta Commons HttpClient library.³² As indicated by arrow (d) in Figure 3, after selecting one of the options, i.e., Biological Process, Molecular Function, or Pathway, the PANTHER web site is queried and a summary report containing a pie chart and list of classifications is returned. One can also download a list of the protein identifications which can be manually submitted to PANTHER for more detailed analysis.

Another challenge is the comparison and summarization of proteomic results from different experiments. In the most simple instance, YPED provides tools for displaying distinct, unambiguously identified proteins, pairwise intersections between samples, and identified proteins common to all samples (for comparisons of three or more samples). In less straightforward instances, YPED also provides tools to help deal with the often large collections of MS/MS spectra that potentially each may have arisen from more than one unique protein. In these instances simply finding the intersection or set difference of all the indistinguishable proteins in two experiments will result in a large and potentially overwhelming collection of indistinguishable proteins. Rather, YPED filters out the indistinguishable proteins and uses one protein as an indicator for each indistinguishable group.

Since discovery and comparative profiling of phosphoproteomes is another area of very active research - which follows from the important role this post-translational modification often plays in modulating protein activity, we have employed another strategy to facilitate the viewing and interpretation of protein phosphorylation. A checkbox is provided on the result menu page for filtering phosphorylated peptides and proteins. Only phosphorylated proteins are displayed when this option is selected and the phosphorylated peptides are highlighted in red. A web link to our phosphopeptide probability calculation based on fractional mass³³ is also provided to help validate the identification. An example screen showing a phosphopeptide and the results of the fractional mass probability calculation is shown in Figure 4.

Discussion, Status, and Further Plans

The ability to effectively handle the management and dissemination of proteomic data is critically important to be able to take advantage of continuing advancements in the comparative analysis of proteomes. Additionally, investigators and their collaborators need more than just a list of proteins identified in a given sample. They need tools that enable them to obtain concise, informative comparisons between different sets of experiments.

A variety of proteomic database systems have been described including SBEAMS,¹¹ PEDro,¹² Proteios,¹³ CPAS,¹⁴ PRIDE¹⁵ and DBParser,¹⁶ each addressing various aspects and issues, e.g., PEDro with an eye toward standards, CPAS incorporating experimental workflow in concert with standard file formats and the Trans-Proteomic Pipeline, and PRIDE capturing proteomics data for dissemination in a publicly accessible data warehouse. YPED has similar functionality to Proteios with the additional feature of integration of the TPP validation and specific emphasis on the seamless incorporation of DIGE results.

The driving force for YPED was to store the experimental results from our often highly complex proteomic analyses in a database that enables *non-mass spectrometrists* a concise and understandable overview of their data and the ability to extract significant biological findings for publication and downstream validation. Although YPED includes database search scores, “prophet” probability values, and estimated false positive error tables, the users must still make a judgment concerning level of error they are willing to accept. Currently, YPED has the ability for collecting the necessary sample information to be MIAPE compliant while allowing researchers to submit samples for a variety of proteomic analyses. This data should facilitate data mining in the future.

To effectively manage the results of the hundreds of research investigators with whom we interact, we needed to develop a database which would provide users with data results in an interactive format. The Yale University W. M. Keck Biotechnology Resource handles (fee for service) protein profiling analyses for investigators across the globe. The Keck MS and Proteomics Resource also supports several closely associated NIH Centers. These include the Yale/NHLBI Proteomics Center, Yale/NIDA Neuroproteomics Center, Yale Cancer Center Proteomics Shared Resource and the Proteomics Core for the Northeast Center of Excellence in Biodefense. The Keck MS & Proteomics Resource annually carries out more than 300 DIGE experiments, with approximately 27,000 protein spots of interest being subjected to MALDI-Tof/Tof-based protein identifications. To date, more than 800 DIGE experiments have been completed. YPED has made a significantly positive impact on our ability to effectively and quickly disseminate this large volume of information in a user-friendly format that has greatly decreased the analysis turn around time. MudPIT, iTRAQ and ICAT experiments total approximately 100 samples analyzed to date. YPED is heavily relied on for the final data reports and currently has 128 research investigators and 504 result sets archived.

Currently we are using only part of the Trans-Proteomic Pipeline (TPP) for performing peptide/protein validation, we will explore the use of other programs that are or will be offered by TPP. For example, Libra³⁴ is a recently developed TPP module that can potentially be useful for iTRAQ quantification. For large-scale protein identification searches that allow for possible post-translational modifications, we need to move to high-performance-computing (HPC) solutions. We have started this effort by developing an optimized parallel version of X!Tandem to take advantage of multiple cluster nodes with dual CPUs to substantially speed up the massive database searches.³⁵ Such HPC solutions are key to the successful utilization of high-throughput protein profiling technologies.

Conclusions

We have presented an overview of a versatile and user-friendly database system for managing, querying, viewing, interpreting, and archiving data derived from multiple, state-of-the-art protein profiling technologies. Building such a mission critical system requires multidisciplinary collaboration (e.g., biostatistics, computer science, laboratory researchers and bioinformatics) to address complex issues and needs including data analysis, performance, and flexibility/extensibility of database design. In addition, users have played an active and important role in providing input and feedback on a system and user interface design which is able to meet their needs. Finally, our system has demonstrated the importance of interoperability with a variety of computation and annotation resources such as the Trans-Proteomic Pipeline and PANTHER. For the system to be usable and successful, it is important to provide an integrated environment that incorporates the natural workflow of data acquisition, retrieval, and interpretation. YPED is an evolving system with new components added as the need requires.

YPED is freely and publicly available from the authors at http://yped.med.yale.edu/yped_dist under the terms of the GNU General Public License and a demonstration login is available at <http://yped.med.yale.edu>.

Acknowledgments

This project has been funded in whole or in part with Federal funds from NIH/NHLBI contract N01-HV-28186 and NIH/NIDA grant P30 DA018343.

References

1. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet.* 1999; 21(1 Suppl):33–37. [PubMed: 9915498]
2. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 2003; 422(6928):198–207. [PubMed: 12634793]
3. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science.* 2006; 312(5771):212–217. [PubMed: 16614208]
4. Wilkins, MR.; Williams, KL.; Appel, RD.; Hochstrasser, DF., editors. *Proteome research: new frontiers in functional genomics.* Springer; Berlin; New York: 1997. p. xviii-243.
5. Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, Pognan F, Hawkins E, Currie I, Davison M. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics.* 2001; 1(3):377–396. [PubMed: 11680884]
6. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol.* 2001; 19(3):242–247. [PubMed: 11231557]
7. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol.* 1999; 17(10):994–999. [PubMed: 10504701]

8. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002; 1(5):376–386. [PubMed: 12118079]
9. Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*. 2007; 7(3):340–350. [PubMed: 17177251]
10. Boguski MS, McIntosh MW. Biomedical informatics for proteomics. *Nature*. 2003; 422(6928): 233–237. [PubMed: 12634797]
11. SBEAMS web page. <http://www.sbeams.org>
12. Garwood K, McLaughlin T, Garwood C, Joens S, Morrison N, Taylor CF, Carroll K, Evans C, Whetton AD, Hart S, Stead D, Yin Z, Brown AJ, Hesketh A, Chater K, Hansson L, Mewissen M, Ghazal P, Howard J, Lilley KS, Gaskell SJ, Brass A, Hubbard SJ, Oliver SG, Paton NW. PEDRo: a database for storing, searching and disseminating experimental proteomics data. *BMC Genomics*. 2004; 5(1):68. [PubMed: 15377392]
13. Levander F, Krogh M, Warell K, Garden P, James P, Hakkinen J. Automated reporting from gel-based proteomics experiments using the open source Proteios database application. *Proteomics*. 2007; 7(5):668–674. [PubMed: 17295359]
14. Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A, Fang R, Faca V, Gafken P, Zhang H, Whiteaker J, States D, Hanash S, Paulovich A, McIntosh MW. Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. *J Proteome Res*. 2006; 5(1):112–121. [PubMed: 16396501]
15. Jones P, Cote RG, Martens L, Quinn AF, Taylor CF, Derache W, Hermjakob H, Apweiler R. PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res*. 2006; 34(Database issue):D659–D663. [PubMed: 16381953]
16. Yang X, Dondeti V, Dezube R, Maynard DM, Geer LY, Epstein J, Chen X, Markey SP, Kowalak JA. DBParser: web-based software for shotgun proteomic data analyses. *J Proteome Res*. 2004; 3(5):1002–1008. [PubMed: 15473689]
17. Appel RD, Bairoch A, Sanchez JC, Vargas JR, Golaz O, Pasquali C, Hochstrasser DF. Federated two-dimensional electrophoresis database: a simple means of publishing two-dimensional electrophoresis data. *Electrophoresis*. 1996; 17(3):540–546. [PubMed: 8740178]
18. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*. 2001; 29(4):365–371. [PubMed: 11726920]
19. Taylor CF. Minimum Reporting Requirements for Proteomics: A MIAPE Primer. *Proteomics*. 2006; 6(2):39–44. [PubMed: 17031795]
20. NCBI Taxonomy web page. <http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html>
21. NCI Thesaurus web page. <http://ncit.nih.gov>
22. Eng JK, McCormack AL, Yates JR III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*. 1994; 5(11): 976–989. [PubMed: 24226387]
23. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999; 20(18):3551–3567. [PubMed: 10612281]
24. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004; 20(9):1466–1467. [PubMed: 14976030]
25. MacLean B, Eng JK, Beavis RC, McIntosh M. General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics*. 2006; 22(22): 2830–2832. [PubMed: 16877754]
26. Keller A, Eng J, Zhang N, Li XJ, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol*. 2005; 1:2005 0017. [PubMed: 16729052]

27. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem.* 2002; 74(20):5383–5392. [PubMed: 12403597]
28. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem.* 2003; 75(17):4646–4658. [PubMed: 14632076]
29. Han DK, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol.* 2001; 19(10):946–951. [PubMed: 11581660]
30. Li XJ, Zhang H, Ranish JA, Aebersold R. Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal Chem.* 2003; 75(23):6648–6657. [PubMed: 14640741]
31. Panther Classification System web page. <http://www.pantherdb.org/>
32. Commons HttpClient library web page. <http://jakarta.apache.org/commons/httpclient/>
33. Bruce C, Shifman MA, Miller P, Gulcicek EE. Probabilistic enrichment of phosphopeptides by their mass defect. *Anal Chem.* 2006; 78(13):4374–4382. [PubMed: 16808444]
34. Libra web page. <http://tools.proteomecenter.org/Libra.php>
35. Bjornson RD, Carriero NJ, Colangelo C, Shifman M, Cheung KH, Miller PL, Williams K. X!! Tandem, an improved method for running X!Tandem in parallel on collections of commodity computers. *J Proteome Res.* 2007 in press.

User name: Demo Account for YPED Analysis Type: DIGE Funding Source: NIDA

Experiment Type: normal vs diseased comparison

Sample Name: YPED DIGE DEMO1

DIGE Gel Number: d1 PI Range Desired: 3-10 %Polyacrylamide Desired: 12.5% Sample Buffer: std Internal Standard Desired: no

Sample Comment: This is an example DIGE requisition

Sample Name Part1

Sample Type: normal/control Label: Cy-3 Species: Homo sapiens Protein mg/ml: 1 Volume ul: 1 Total Sample to Load ug: 1

Tissue: Brain Cell Type: Cell Line: Developmental Stage: ☒ AAA requested

Sample Part1 Comment:

Sample Name Part2

Sample Type: normal/control Label: Cy-5 Species: Homo sapiens Protein mg/ml: 1 Volume ul: 1 Total Sample to Load ug: 1

Tissue: Spinal_Cord Cell Type: Cell Line: Developmental Stage: ☒ AAA requested

Sample Part2 Comment:

Figure 1.
DIGE sample requisition form.

DIGE

Merging Results

Figure 2.
DIGE data flow.

Figure 3. Query and display of DIGE results with links to an active gel image and to the PANTHER classification system. Arrows: (a) Result menu to result page, (b) Result page to DIGE gel image, (c) Specific spot to protein identification page, (d) Result page to PANTHER summary.

Mascot Peptides			
Protein ID	IPI00207370		
Protein Name	Tax_Id=10116 Neurabin-2		
Peptides Identified	17		
Percent Coverage	26		
Peptide Sequence	Search Score	Ion Mass Matched	Precursor Mass
ADSRTLHR Phospho (ST)[3] Fractional Mass P values	-1	1092.49462890625	0
AEAVPEEEASSSVATLER	8	1874.89770507812	1874.89526367188
AQLEQSVEENKER	47	1559.76599121094	1559.76098632813
DRGPGGQQPPQHR	-1	1429.70434570313	0

Phospho Fraction Mass Probabilities			
Protein ID	IPI00207370		
Protein Name	Tax_Id=10116 Neurabin-2		
Peptide	ADSRTLHR Phospho (ST)[3]		
Ion Mass	1092.49462890625		
Fractional Probability	PPM	Species	
0.87	1	human	
0.71	5	human	
0.66	10	human	
0.93	1	mouse	
0.73	5	mouse	
0.70	10	mouse	

Figure 4.
Mascot peptide display page with phosphorylated peptide indicated in red and link to phosphopeptide probability calculation based on fractional mass.