The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data

Peter E. Hodges, Andrew H. Z. McKee, Brian P. Davis, William E. Payne and James I. Garrels*

Proteome Inc., 100 Cummings Center, Suite 435M, Beverly, MA 01915, USA

Received October 27, 1998; Accepted October 28, 1998

ABSTRACT

The Yeast Proteome Database (YPD) is a model for the organization and presentation of comprehensive protein information. Based on the detailed curation of the scientific literature for the yeast *Saccharomyces cerevisiae*, YPD contains more than 50 000 annotations lines derived from the review of 8500 research publications. The information concerning each of the ~6100 yeast proteins is structured around a convenient one-page format, the Yeast Protein Report, with additional information provided as pop-up windows. Protein classification schema have been revised this year, defining each protein's cellular role, function and pathway, and adding a Functional Abstract to the Yeast Protein Report. These changes provide the user with a succinct summary of the protein's function and its place in the biology of the cell, and they enhance the power of YPD Search functions. Precalculated sequence alignments have been added, to provide a crossover point for comparative genomics. The first transcript profiling data has been integrated into the YPD Protein Reports, providing the framework for the presentation of genome-wide functional data. The Yeast Proteome Database can be accessed on the Web at http://www.proteome.com/YPDhome.html

INTRODUCTION

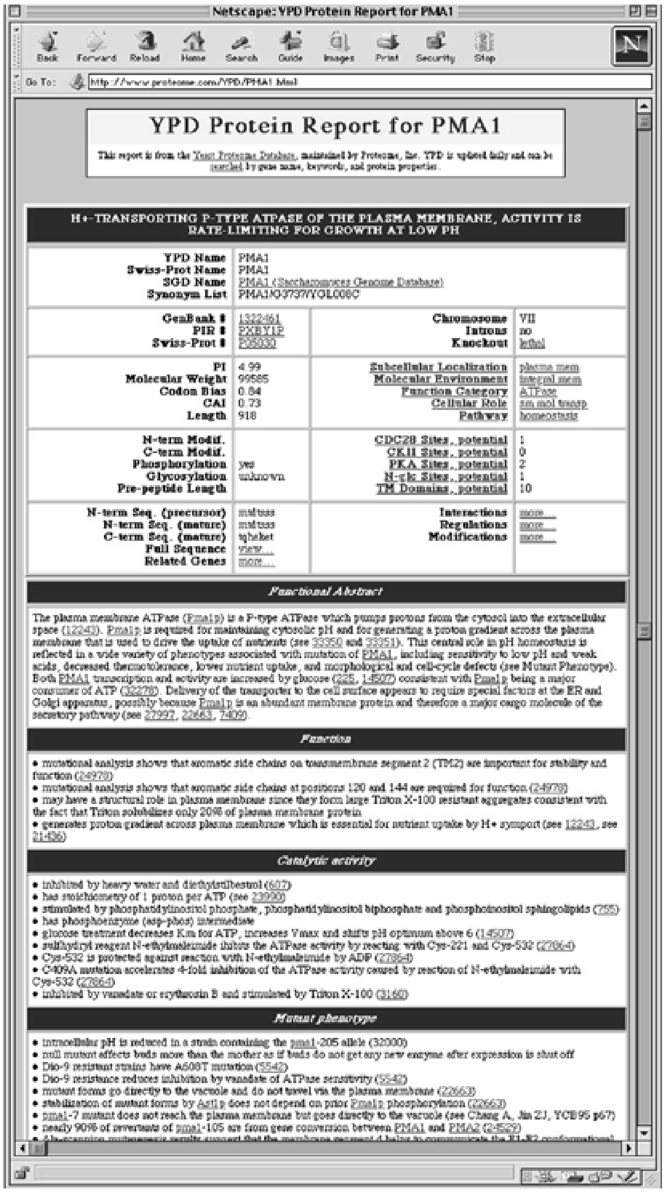
The Yeast Proteome Database (YPD™) is the first annotated proteome database for any organism (1). YPD is annotated by in-depth curation of the research literature, and it is a proteome database because it contains entries for each known or predicted protein of *Saccharomyces cerevisiae*. YPD tabulates a variety of properties, functions and interactions of the proteins. It contains more than 50 000 annotation lines derived from a review of the scientific findings contained in 8500 articles. The information for each of the approximately 6100 yeast proteins is presented in a convenient one-page format, the Yeast Protein Report (Fig. 1). From the Report page, users can display pop-up windows with more detailed information or descriptions, such as the full protein sequence, protein–protein interactions, regulation of gene expression, protein modifications and sequence alignments with proteins from humans and model organisms. YPD is fully searchable, by gene name or synonym, by any keyword that appears in the annotation lines, or by any curated or calculated protein property. YPD is still growing rapidly, and will soon be a repository of the curated information from the entire research literature on yeast proteins. YPD is a heavily used resource on the World-Wide Web (http://www.proteome.com/YPDhome.html), where it is accessed by more than 2500 different academic users each week. This article summarizes the progress of YPD in curating the literature, highlights the features added to YPD over the last year and introduces areas that will be expanded in the next year.

THE GROWTH OF YPD CURATION

The annotations and properties contained in YPD are written by a staff of PhD level curators experienced in yeast research. The curatorial staff has read and annotated 8500 research articles, including nearly 3500 in the last year. Curation has accelerated to about 400 research papers every month, and with special emphasis on new articles, YPD curates more than 80% of the current literature on yeast proteins within two months of publication.

Since YPD tracks the full body of scientific publication on yeast proteins, we are in a unique position to measure the nature and progress of yeast science. The yeast scientific community continues to expand their scope. As an indicator, YPD tracks the numbers of yeast proteins that have an assigned function, as determined by genetic or biochemical experiments, as well as those proteins with a function predicted by sequence homology, and the remaining uncharacterized proteins (Table 1). In 1998 yeast researchers passed the milestone of having characterized half of the yeast proteome. Already, the phenotype is known for the disruption of 2692 genes, or 44% of the genome (tested at least as to viability). The number of proteins studied and reported on during the year was 1669 yeast proteins in 1995, 1787 proteins in 1996 and 2188 proteins in 1997. Clearly the scope of yeast protein research is expanding, and experimental data concerning more than a third of the yeast proteome is published each year.

*To whom correspondence should be addressed. Tel: +1 978 922 1643; Fax: +1 978 922 3971; Email: ypd@proteome.com


Netscape: YPD Protein Report for PMA1
Back
Forward
Reload
Home
Search
Guide
Images
Print
Security
Stop
Go To:
http://www.proteome.com/YPD/PMA1.html

YPD Protein Report for PMA1

This report is from the Yeast Proteome Database, maintained by Proteome, Inc. YPD is updated daily and can be searched by gene name, keywords, and protein properties.

H+-TRANSPORTING P-TYPE ATPASE OF THE PLASMA MEMBRANE, ACTIVITY IS RATE-LIMITING FOR GROWTH AT LOW PH

YPD Name	PMA1		
Swiss-Prot Name	PMA1		
SGD Name	PMA1 (Saccharomyces Genome Database)		
Synonym List	PMA1/G3737/YGL008C		
GenBank #	1322461	Chromosome	VII
PIR #	PXBY1P	Introns	no
Swiss-Prot #	P05030	Knockout	lethal
PI	4.99	Subcellular Localization	plasma mem
Molecular Weight	99585	Molecular Environment	integral mem
Codon Bias	0.84	Function Category	ATPase
CAI	0.73	Cellular Role	sm mol transp
Length	918	Pathway	homeostasis
N-term Modif.		CDC28 Sites, potential	1
C-term Modif.		CKII Sites, potential	0
Phosphorylation	yes	PKA Sites, potential	2
Glycosylation	unknown	N-glc Sites, potential	1
Pre-peptide Length		TM Domains, potential	10
N-term Seq. (precursor)	mdtsss	Interactions	more...
N-term Seq. (mature)	mdtsss	Regulations	more...
C-term Seq. (mature)	tqhekt	Modifications	more...
Full Sequence	view...		
Related Genes	more...		

Functional Abstract

The plasma membrane ATPase (Pma1p) is a P-type ATPase which pumps protons from the cytosol into the extracellular space (12243). Pma1p is required for maintaining cytosolic pH and for generating a proton gradient across the plasma membrane that is used to drive the uptake of nutrients (see 33350 and 33351). This central role in pH homeostasis is reflected in a wide variety of phenotypes associated with mutation of PMA1, including sensitivity to low pH and weak acids, decreased thermotolerance, lower nutrient uptake, and morphological and cell-cycle defects (see Mutant Phenotype). Both PMA1 transcription and activity are increased by glucose (225, 14507) consistent with Pma1p being a major consumer of ATP (32278). Delivery of the transporter to the cell surface appears to require special factors at the ER and Golgi apparatus, possibly because Pma1p is an abundant membrane protein and therefore a major cargo molecule of the secretory pathway (see 27997, 22663, 7409).

Function

- mutational analysis shows that aromatic side chains on transmembrane segment 2 (TM2) are important for stability and function (24978)
- mutational analysis shows that aromatic side chains at positions 120 and 144 are required for function (24978)
- may have a structural role in plasma membrane since they form large Triton X-100 resistant aggregates consistent with the fact that Triton solubilizes only 20% of plasma membrane protein
- generates proton gradient across plasma membrane which is essential for nutrient uptake by H+ symport (see 12243, see 21436)

Catalytic activity

- inhibited by heavy water and diethylstilbestrol (607)
- has stoichiometry of 1 proton per ATP (see 23930)
- stimulated by phosphatidylinositol phosphate, phosphatidylinositol biphosphate and phosphoinositol sphingolipids (755)
- has phosphoenzyme (asp-phos) intermediate
- glucose treatment decreases Km for ATP, increases Vmax and shifts pH optimum above 6 (14507)
- sulfhydryl reagent N-ethylmaleimide ihibits the ATPase activity by reacting with Cys-221 and Cys-532 (27864)
- Cys-532 is protected against reaction with N-ethylmaleimide by ADP (27864)
- C409A mutation accelerates 4-fold inhibition of the ATPase activity caused by reaction of N-ethylmaleimide with Cys-532 (27864)
- inhibited by vanadate or erythrosin B and stimulated by Triton X-100 (3160)

Mutant phenotype

- intracellular pH is reduced in a strain containing the pma1-205 allele (32000)
- null mutant affects buds more than the mother as if buds do not get any new enzyme after expression is shut off
- Dio-9 resistant strains have A608T mutation (5542)
- Dio-9 resistance reduces inhibition by vanadate of ATPase sensitivity (5542)
- mutant forms go directly to the vacuole and do not travel via the plasma membrane (22663)
- stabilization of mutant forms by Ast1p does not depend on prior Pma1p phosphorylation (22663)
- pma1-7 mutant does not reach the plasma membrane but goes directly to the vacuole (see Chang A., Jin ZJ, YCB96 p67)
- nearly 90% of revertants of pma1-105 are from gene conversion between PMA1 and PMA2 (24529)
- Ala-scanning mutagenesis results suggest that the membrane segment 4 helps to communicate the E1-E2 conformational

Figure 1. A representative YPD Yeast Protein Report page describing the plasma membrane ATPase encoded by the PMA1 gene. Users can access this page via the World Wide Web at http://www.proteome.com/YPD/PMA1.html . The page includes a YPD Title Line which is the best short description for the function of the protein (uppermost), a table of protein properties (immediately below the YPD Title Line), annotations from the curated literature sorted by topic headings (main body, not completely shown) and the list of references (below the annotations, not shown). Details of protein properties are provided by pop-up windows. The new properties Cellular Role and Pathway, as well as the new annotation topic Functional Abstract provides users with a brief synopsis of the nature and function of the protein.

NEW FEATURES TO YPD

YPD is now based in a relational (Oracle) format which affords major improvements in the structuring of search queries. Another major advantage will be the enforcement of a controlled vocabulary for discussing protein roles, functions, subcellular localizations, modifications, complexes, pathways, genetic regulators and phenotypes of associated mutations.

Table 1. Growth of YPD content by release number

YPD release	Date	Total proteins	Known function[a]	Similarity[b]	Unknown function[c]
1.2	Nov. 23, 1994	3020	1729	387	904
2.0	Dec. 8, 1994	3142	1750	450	942
3.0	Feb. 1, 1995	3512	1871	524	1117
4.0	Jun. 6, 1995	4046	1951	667	1428
4.1	Jul. 7, 1995	4305	2012	729	1564
5.0	Nov. 30, 1995	4559	2187	859	1913
6.0	Aug. 3, 1996	6021	2369	1231	2421
7.01	Jan. 21, 1997	6045	2507	1192	2346
8.01	Jan. 6, 1998	6096	2837	1067	2192
8.42	Oct. 21, 1998	6085	3094	961	2030

[a] Proteins characterized through genetic or biochemical experiments.
[b] Proteins that have not been characterized but have sequence similarity to characterized proteins.
[c] Proteins of unknown function.

YPD has expanded the classification scheme for proteins, to better define the proteins for the reader and to allow more powerful searches. These data are displayed together in an expanded Properties table, designed to provide a comprehensive but brief profile of each protein for the reader to orient themselves before proceeding to the more detailed annotation lines (Fig. 1). A Cellular Role assigns the protein to one or more essential cellular processes. The Functional Categories have been expanded to better define protein activities. Where appropriate, proteins are assigned to a Pathway defined by biochemistry, genetics or cell biology. In addition, a new annotation topic, the ‘Functional Abstract’ has been introduced. The Functional Abstract is a distillation of the most relevant knowledge which describes each protein’s function. With these new features, the reader will have a more comprehensive understanding of the protein’s function and its role in cell biology.

GUIDING COMPARATIVE GENOMICS THROUGH RICH PROTEIN CONNECTIONS

Yeast is likely to be the first cell where the function of every protein is understood. Already more than half of yeast proteins have been characterized (Table 1). The true power of a comprehensive, model proteome database is to fill in the gaps in the fragmentary data from other organisms. YPD serves that role, allowing researchers studying other organisms to better understand proteins of interest by investigating their yeast homologs. Tracing the links from one yeast protein to all those proteins with which it interacts allows scientists to predict the homolog’s function and interaction in other organisms. As an entry point for comparative genomic analysis, YPD now provides sequence alignments on the Related Genes pop-up window. Alignments are based on the BLAST program (2) with refinement by a Smith–Waterman algorithm (3). Alignments are presented between each yeast protein and all other yeast proteins, human proteins and proteins from other model organisms. The output is precalculated and formatted to speed its presentation, but updated regularly to remain current.

Most importantly, YPD provides links that are not based simply on sequence similarity. YPD connects proteins with common physical properties or common gene regulation. YPD connects enzymes to their substrates, kinases to their targets, and transcription factors to the proteins they regulate. For example, following links from the PMA1 page for the yeast plasma membrane ATPase (http://www.proteome.com/YPD/PMA1.html , shown in part in Fig. 1), YPD connects to six other databases with relevant information, links to 27 different yeast proteins with some relationship to Pma1p, and provides access to further details from 122 research papers via their PubMed abstracts. Via precalculated BLAST reports, Pma1p is connected to the 10 most similar yeast proteins and to homologs in human and model organisms. Via the YPD Search page, an immeasurable number of links connect PMA1 to other yeast proteins that share a common protein property or that share a common keyword from the annotation of their experimental data.

YPD IS MAKING SENSE OF FUNCTIONAL GENOMICS

With the completion of the yeast genome in 1996 (4), it was found that about one-third of the yeast proteins were uncharacterized and had no similarity to characterized proteins of other species. The rate at which new proteins are characterized by the yeast research community, using conventional genetic and biochemical experiments, has held steady at about 30 proteins per month (Table 1). While this rate is impressive, it will still be some years before functions of the unknown proteins are discovered. One of the benefits of the complete genome sequence has been a shift in research emphasis toward genome-wide experiments of gene function, including hybridization to high-density DNA microarrays (DNA chips, 5–10), serial analysis of gene expression (SAGE, 11,12), systematic gene disruptions (13–17), systematic studies of protein subcellular localization (15–17), and systematic two-hybrid analysis of protein–protein interactions (18). These ‘functional genomic’ experiments are rapidly adding new information regarding the functions of all the yeast genes. Such experiments are helping to fill the gaps in the knowledge of yeast biology, and they are helping industrial researchers to find new targets for drug discovery.

Functional genomics experiments typically generate tens of thousands of data points per experiment. The data cannot be reduced to just a few experimental results, necessitating new forms of presentation. Most experiments thus far have been presented as a printed publication, in which only the highlights are discussed, and as a Web site where the full data table can be accessed (see ref. 7 and http://cmgm.stanford.edu/pbrown/explore/index.html , for example). As yet there is no standardized means for presentation of this data, and scientists are left with the task of collecting data from various Web sites with no unified presentation platform for the data. Furthermore, these experiments cannot be analyzed in the absence of broader knowledge of yeast genes. Usually only minimal descriptions of each gene are given, leaving each researcher on his own to track the significance of each gene’s change in expression. With 6000 genes and hundreds of data sets soon to be available, there is little chance that any investigator can obtain all the results from these experiments that may be relevant to his/her work. Presentation of functional genomic datasets in the context of YPD would overcome these problems.

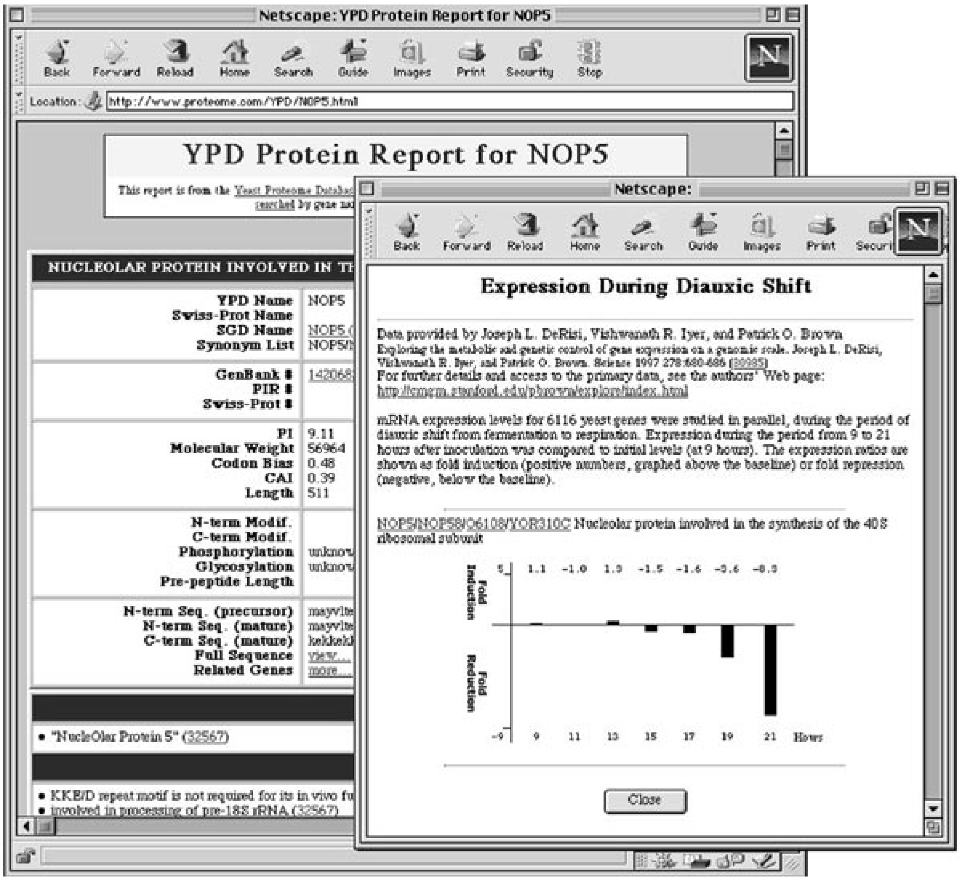
This past year YPD introduced the first presentation of functional genomic data integrated into the proteome database. The data, kindly provided by Joseph DeRisi, Vishwanath Iyer and Patrick Brown, describe the effect of diauxic shift on transcript abundance, measured simultaneously for every gene in the genome (7, http://cmgm.stanford.edu/pbrown/explore/index.html). These expression data are displayed for nearly every YPD Protein Report, located with other genetic regulation experiments on the Regulation pop-up window (Fig. 2). Here the users can view the data in the context of experimental results obtained by other techniques, and in the context of all of the information known about the protein. Furthermore, powerful search capabilities through YPD allow the user to identify specific subsets of proteins with shared properties, and then refer to the diauxic shift data for each of these proteins. We plan to expand the presentation of transcript profiling datasets, and we actively seek more functional genomic data for inclusion in YPD.

YPD TITLE LINES™ PROVIDE MEANING TO HIT LISTS

One of the most frequently encountered difficulties in interpretation of functional genomic data is making sense of long lists of similarity hits. Whether reading a BLAST output or analyzing a genome-wide transcript profile, the problem is the same. The output ‘hit list’ does not contain enough information to scan the list for meaningful scientific leads. Researchers have warned of the dangers of inaccurate or outdated annotation of database entries (19–22). YPD can help by providing access to YPD Title Lines. Every yeast protein is described by a single line, distilling the essence of the protein’s function. As each newly published research paper is read by a YPD curator, the YPD Title Lines are reevaluated, and rewritten as necessary. As a result, the description of each yeast protein is always current. The YPD Title Lines can be accessed by downloading the YPD Spreadsheet (see http://www.proteome.com/YPDspreadsheet.html), and can be used by the academic researcher to interpret functional genomic data, or, with permission, can be used to annotate Web site data or tables for publication. By regularly updating their copy of the YPD Title Lines, a researcher can keep old data fresh, and even find new interpretations of existing data. As more is discovered about the function of uncharacterized proteins, new meaning may be found in old datasets. For example, DeRisi *et al*. (7, http://cmgm.stanford.edu/pbrown/explore/index.html) observed the global repression of the proteins involved in ribosomal biogenesis during diauxic shift. However, since that publication much has been learned about the proteins involved in ribosomal RNA processing. A reanalysis of the transcript profiles for the newly characterized small nucleolar ribonucleoprotein (snoRNP) components Nop5p, Nop56p, Cbf5p and Nhp2p shows that each of these proteins is repressed in a fashion seen for the previously characterized snoRNP proteins Gar1p and Nop1p (e.g., Fig. 2).

FUTURE DEVELOPMENT OF YPD

YPD curates all newly published articles concerning yeast proteins and is making a major effort to complete curation of the older literature. In the near future, we will complete the assignment of protein roles, functions and pathways based on experimental evidence in the curated literature, and the proteins will be summarized in Functional Abstracts. Consistent with the principle that yeast is the best eukaryotic model organism, we are dedicated to providing comparative genomic data, as illustrated by the new ‘Related Genes’ protein property field, with additional comparative genomic features to follow. Finally, with the shift in



Location: http://www.proteome.com/YPD/NOP5.html

YPD Protein Report for NOP5

This report is from the Yeast Proteome Databa
searched by gene na

NUCLEOLAR PROTEIN INVOLVED IN TH

YPD Name	NOP5
Swiss-Prot Name	
SGD Name	NOP5 (
Synonym List	NOP5/N
GenBank #	142068
PIR #	
Swiss-Prot #	
PI	9.11
Molecular Weight	56964
Codon Bias	0.48
CAI	0.39
Length	511
N-term Modif.	
C-term Modif.	
Phosphorylation	unknow
Glycosylation	unknow
Pre-peptide Length	
N-term Seq. (precursor)	mayvlte
N-term Seq. (mature)	mayvlte
C-term Seq. (mature)	kekkekk
Full Sequence	view...
Related Genes	more...

- "NucleOlar Protein 5" (32567)

- KKE/D repeat motif is not required for its in vivo fu
- involved in processing of pre-18S rRNA (32567)

Expression During Diauxic Shift

Data provided by Joseph L. DeRisi, Vishwanath R. Iyer, and Patrick O. Brown
Exploring the metabolic and genetic control of gene expression on a genomic scale. Joseph L. DeRisi, Vishwanath R. Iyer, and Patrick O. Brown. Science 1997 278:680-686 (30985)
For further details and access to the primary data, see the authors' Web page:
http://cmgm.stanford.edu/pbrown/explore/index.html

mRNA expression levels for 6116 yeast genes were studied in parallel, during the period of diauxic shift from fermentation to respiration. Expression during the period from 9 to 21 hours after inoculation was compared to initial levels (at 9 hours). The expression ratios are shown as fold induction (positive numbers, graphed above the baseline) or fold repression (negative, below the baseline).

NOP5/NOP58/O6108/YOR310C Nucleolar protein involved in the synthesis of the 40S ribosomal subunit

Figure 2. Representative pop-up window for the transcript profile during diauxic shift from the protein report for the NOP5 gene. This pop-up window is accessed through the Regulations protein property line and provides a convenient, easily understood graphical representation of the functional genomic data provided by DeRisi, Iyer and Brown (7, http://cmgm.stanford.edu/pbrown/explore/index.html). A brief description of experimental detail is followed by the gene name and synonyms, the YPD Title Line as the best short functional description and a graph of the relative transcript abundance over the time of the diauxic shift. Expression of NOP5 is strongly repressed late in diauxic shift, as is the expression of other genes encoding snoRNP proteins.

research emphasis toward genome-wide experiments of gene function, YPD aims to be a repository for functional genomic data from the scientific community. As already shown in the pop-up window for regulations, YPD can present transcription profiles of the yeast genome in a meaningful context. YPD will continue to include additional transcript profiles and other functional genomic data, including the results of systematic gene disruptions, genome-wide two-hybrid screening, and serial analysis of gene expression (SAGE).

HOW TO SUBMIT PROTEIN DATA TO YPD

We appreciate the feedback from our users, concerning new data submission, additions, clarifications and corrections. Personal communications will be cited as such, and functional genomic datasets are especially welcomed. Any correspondence should be directed to ypd@proteome.com or by mail to the address of the authors.

CITING YPD

Authors wishing to make use of the information provided by YPD should cite this article as a general reference for the access and content of YPD.

ACKNOWLEDGEMENTS

We appreciate the contributions of a number of scientists for reviews of the direction, scope and content of YPD: Bruno André, Charles Cole, Les Grivell, Sepp Kohlwein and Jon Warner. We thank Joseph DeRisi, Vishwanath Iyer and Patrick Brown for allowing us to present their data as the prototype functional genomic dataset, and all the researchers who have contributed datasets since then. We acknowledge the help and cooperation of the staffs at the Saccharomyces Genome Database (http://genome-www.stanford.edu/Saccharomyces/) and at the Munich Information Centre for Protein Sequences (http://www.mips.biochem.mpg.de/). Most wholeheartedly, we thank the members of the yeast scientific community. It is their dedication, innovation and hard work that provide the scientific content of YPD, and their comments, corrections and suggestions that help keep it accurate. The development of YPD has been partially funded by a Phase II SBIR grant from the National Institute of General and Medical Sciences (R44 GM54110-02).

REFERENCES

1 Hodges,P.E., Payne,W.E. and Garrels,J.I. (1998) *Nucleic Acids Res*., **26**, 68–72.
2 Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1995) *Nucleic Acids Res*., **25**, 3389–3402.

3 Waterman,M.S. (1995) *Introduction to Computational Biology: Maps, Sequences and Genomes*. Chapman & Hall, London.

4 Goffeau,A., Barrell,B.G., Bussey,H., Davis,R.W., Dujon,B., Feldmann,H., Galibert,F., Hoheisel,J.D., Jacq,C., Johnston,M. *et al.* (1996) *Science*, **274**, 546–567.

5 Cho,R.J., Cambell,M.J., Winzeler,E.A., Steinmetz,L., Conway,A., Wodicka,L., Wolfsberg,T.G., Gabrielian,A.E., Landsman,D., Lockhart,D.J. and Davis,R.W (1998) *Mol. Cell*, **2**, 65–73.

6 DeRisi,J., Penland,L., Brown,P.O., Bittner,M.L., Meltzer,P.S., Ray,M., Chen,Y., Su,Y.A. and Trent,J.M. (1996) *Nature Genet*., **14**, 457–460.

7 DeRisi,J.L., Iyer,V.R. and Brown,P.O (1997) *Science*, **278**, 680–686.

8 Lashkari,D.A., DeRisi,J.L., McCusker,J.H., Namath,A.F., Gentile,C., Hwang,S.Y., Brown,P.O., Davis,R.W. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 13057–13062.

9 Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H. and Brown,E.L. (1996) *Nature Biotechnol*., **14**, 1675–1680.

10 Wodicka,L., Dong,H., Mittmann,M., Ho,M.-H. and Lockhart,D.J (1997) *Nature Biotechnol*., **15**, 1359–1367.

11 Velculescu,V.E., Zhang,L., Vogelstein,B., Kinzler,K.W (1995) *Science*, **270**, 484–487.

12 Velculescu,V.E., Zhang,L., Zhou,W., Vogelstein,J., Basrai,M.A., Bassett,D.E.,Jr, Hieter,P., Vogelstein,B. and Kinzler,K.W. (1997) *Cell*, **88**, 243–251.

13 Shoemaker,D.D., Lashkari,D.A., Morris,D., Mittmann,M. and Davis,R.W. (1996) *Nature Genet*., **14**, 450–456.

14 Smith,V., Chou,K.N., Lashkari,D., Botstein,D. and Brown,P.O. (1996) *Science*, **274**, 2069–2074.

15 Burns,N., Grimwade,B., Ross-Macdonald,P.B., Choi,E.Y., Finberg,K., Roeder,G.S. and Snyder,M. (1994) *Genes Dev*., **8**, 1087–1105.

16 Ross-Macdonald,P., Sheehan,A., Roeder,G.S. and Snyder,M. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 190–195.

17 Erdman,S., Lin,L., Malczynski,M. and Snyder,M. (1998) *J. Cell. Biol*., **140**, 461–483.

18 Fromont-Racine,M., Rain,J.-C. and Legrain,P. (1997) *Nature Genet*., **16**, 277–282.

19 Botstein,D. and Cherry,J.M. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 5506–5507.

20 Ermolaeva,O., Rastogi,M., Pruitt,K.D., Schuler,G.D., Bittner,M.L., Chen,Y., Simon,R., Meltzer,P., Trent,J.M. and Boguski,M. (1998) *Nature Genet*., **20**, 19–23.

21 Hieter,P. and Boguski,M. (1997) *Science*, **278**, 601–602.

22 Smith,T.F. (1998) *Trends Genet*., **14**, 291–293.