Nucleic Acids Research, 2006, Vol. 34, Database issue ***D705–D711***
doi:10.1093/nar/gkj127

MitoP2: the mitochondrial proteome database—now including mouse data

H. Prokisch[1,2,*], C. Andreoli[1], U. Ahting[1], K. Heiss[1], A. Ruepp[3], C. Scharfe[4] and T. Meitinger[1,2]

[1]Institute of Human Genetics, Technical University of Munich, Munich, Germany, [2]Institute of Human Genetics and, [3]Institute for Bioinformatics, GSF National Research Center for Environment and Health, Neuherberg, Germany and [4]Department of Biochemistry and Stanford Genome Technology Center, 855 California Avenue, Palo Alto, CA 94304, USA

Received September 15, 2005; Revised and Accepted October 18, 2005

ABSTRACT

The MitoP2 database (http://www.mitop.de) integrates information on mitochondrial proteins, their molecular functions and associated diseases. The central database features are manually annotated reference proteins localized or functionally associated with mitochondria supplied for yeast, human and mouse. MitoP2 enables (i) the identification of putative orthologous proteins between these species to study evolutionarily conserved functions and pathways; (ii) the integration of data from systematic genome-wide studies such as proteomics and deletion phenotype screening; (iii) the prediction of novel mitochondrial proteins using data integration and the assignment of evidence scores; and (iv) systematic searches that aim to find the genes that underlie common and rare mitochondrial diseases. The data and analysis files are referenced to data sources in PubMed and other online databases and can be easily downloaded. MitoP2 users can explore the relationship between mitochondrial dysfunctions and disease and utilize this information to conduct systems biology approaches on mitochondria.

INTRODUCTION

The application of genomics to biology and medicine requires an understanding how specific gene variants contribute to phenotypes, in combination with a comprehensive knowledge of the 'parts list' of a cellular system and how these components are assembled into functional units (1). Mitochondria are ubiquitous and defined substructures of nucleated cells and lend themselves to systems biology approaches. However, in generic databases the annotation of mitochondrial proteins is often incomplete and does not always distinguish between proteins which have a confirmed mitochondrial subcellular localization and those which are only candidates according to preliminary experimental results or *in silico* predictions. For the human species, about half of the estimated 1500 proteins localized or functionally associated with mitochondria are known (2). Since the mitochondrial organelle is an evolutionarily conserved entity, systematic studies in model organisms are powerful to identify mitochondrial proteins in other organisms (3).

The MitoP2 database was created to consolidate and structure public information on mitochondrial proteins, their functions and associated human diseases (4,5). MitoP2 provides a wide variety of search functions to explore and download information and to access references in PubMed and other public databases. We have further expanded the manually annotated reference sets of mitochondrial proteins in yeast (522 proteins) and human (624 proteins), and have now added the section MitoP2-Mouse (615 proteins). For these three species, we integrated data from genome-wide approaches applied to the study of mitochondria, and assigned an evidence score of a candidate protein being mitochondrial (3). With the help of MitoP2, proteins involved in mitochondrial biogenesis and function have been identified and characterized (6,7). In addition, MitoP2 has enabled the identification of disease genes using positional candidate approaches (8–10).

MitoP2-YEAST

A wealth of information has been collected over the past several years from single gene and genome-wide studies of *Saccharomyces cerevisiae* (11). The list of yeast ORFs and protein annotations in MitoP2-Yeast are based on information in the *Saccharomyces* Genome Database (SGD; http://www.yeastgenome.org) (12). This MitoP2-Yeast update now

*To whom correspondence should be addressed. Tel: +49 89 3187 2890; Fax: +49 89 3187 3297; Email: prokisch@gsf.de

Table 1. Comparison of specificity and sensitivity for various approaches integrated in MitoP2 in determining the mitochondrial localization of proteins

Source	Total proteins	Specificity (%)	Sensitivity (%)
(A) MitoP2-Yeast datasets			
In silico predictions			
MitoProt II score > 0.8 (23)[a]	790	35	83
MITOPRED score > 80 (24)[a,b]	1045	34	68
PSORT II (25)[a]	981	27	51
Predotar (26)[a]	832	36	58
Bayesian prediction (37)[a]	500	42	40
Growth phenotype			
Deletion phenotype (15)[a]	381	50	37
Deletion phenotype (16)[a]	466	51	45
Mitochondrial-associated mRNA			
Mitopolysomes (38)[a]	303	23	13
Sublocalization of tagged proteins			
Ysubloc_01 (14)[a]	364	64	45
Ysubloc_02 (13)[a,b]	527	68	69
Protein–protein interaction			
High confidence interactions (22)[a]	188	62	22
Low confidence interactions (22)[a]	761	26	38
Proteomics of mitochondria			
Yprot_01 (19)[a,b]	177	79	27
Yprot_02 (3)[a,b]	546	50	52
Yprot_03 (20)[a]	749	51	73
Yprot_04 (39)[a]	252	61	29
Mitochondrial expression profiles			
Ytranscr_01 (3)[a]	1357	31	83
Ytranscr_02 (17)[a]	416	19	15
Ytranscr_03 (18)[a,b]	514	43	43
Potential orthologs/homologs			
Human mitochondrial ortholog[a,c]	565	60	65
Mouse mitochondrial ortholog[c]	425	68	55
Neurospora crassa mitochondrial ortholog[c]	337	84	55
MitoP2 calculations			
SVM score > 1	535	78	80
SVM score > 2	386	89	66
(B) MitoP2-Mouse datasets			
In silico predictions			
MITOPRED score > 80 (24)	2455	17	67
PSORT II (25)	4321	7	53
Proteomics of mitochondria			
Mprot_01 (33)[b]	132	77	17
Mprot_02 (34)[b]	359	72	42
Mitochondrial expression profile			
Mtranscr_01 (34)[b]	480	36	28
Sublocalization of tagged proteins			
MSubloc_01 (35)[b]	59	25	2
Potential orthologs/homologs			
Saccharomyces cerevisiae mitochondrial ortholog[c]	1561	16	41
N.crassa mitochondrial ortholog[c]	1030	26	43
Rickettsiae prowazeckii ortholog[c]	991	18	28
Human ortholog[c]	431	64	44
Human homolog with MitoP2 score > 70	421	71	48
MitoP2 calculation			
MitoP2 score > 70	996	47	76
(C) MitoP2-Human datasets			
In silico predictions			
MitoProt II score > 0.8 (23)	2559	12	43
MITOPRED score > 80 (24)	2892	15	61
PSORT II (25)	6125	5	45
Predotar (26)	2139	14	44
Proteomics of mitochondria			
Hprot_01 (32)[b]	736	37	38
Mprot_01 (33)[b]	156	83	10
Mprot_02 (34)[b]	478	60	31
Sublocalization of tagged proteins			
MSubloc_01 (35)[b]	62	26	80
Potential orthologs/homologs			
S.cerevisiae mitochondrial ortholog[c]	854	40	47
N.crassa mitochondrial ortholog[c]	523	48	35
R.prowazeckii ortholog[c]	1426	14	30
MitoP2 calculation			
MitoP2 score > 70	1002	52	73

[a]Datasets used for SVM training.
[b]Recently integrated datasets.
[c]Defined as bidirectional best BLAST hit or best BLAST hit $<1 \times 10^{-10}$.

provides annotated information for 522 mitochondrial reference proteins, which are based on experimental validation of each of these proteins. Recently, systematic cellular sublocalization studies estimated a total of 800 mitochondrial proteins presenting ~12% of the currently known yeast genes (13,14). Therefore, ~250–300 mitochondrial proteins are still missing. In order to identify these missing genes, we have validated and integrated genome-wide approaches applied to the study of mitochondria (3). MitoP2-Yeast datasets in Table 1 show 20 systematic approaches used for this purpose: phenotypes of single gene deletion mutant phenotypes (15,16); systematic subcellular localization studies (13,14); transcriptome datasets of differentially expressed genes including fermentable and non-fermentable growth conditions, the response to diauxic shift, and Hap4 transcription factor screening (3,17,18); proteome analyses of purified mitochondrial organelles (3,19,20); protein abundance measurements (21); and data from protein–protein interaction studies that include interactions to mitochondrial proteins (22). In addition to experimental datasets, mitochondrial proteins can be predicted *in silico* based on the presence of mitochondrial targeting sequences (23–26), and by sequence similarity to a known mitochondrial protein from other species (defined as bidirectional best BLAST hit or best BLAST hit with a score $<1 \times 10^{-10}$) (27). Data from each of these systematic studies can be searched and downloaded.

Using the MitoP2-Yeast reference proteins, it is possible to analyze the specificity and sensitivity of the data from genome-wide studies (Figure 1). Specificity is defined as the proportion of proteins of a dataset which are part of the reference set, while sensitivity is the proportion of reference set proteins which is covered by the dataset. In order to identify putative mitochondrial proteins, we calculated a MitoP2 score for each protein, reflecting the specificity of combined approaches which identified the particular protein (3). To further improve these predictions, we used a new approach utilizing a support vector machine (SVM, http://svmlight.joachims.org). The SVMs are learning machines based on statistical learning theory used for solving classification tasks. We trained the SVM using the MitoP2-Yeast reference set (522 proteins) and a set of 519 proteins with a known localization to other cellular compartments collected from SGD (http://www.yeastgenome.org/). For each of the 1041 proteins, we defined a 20-dimensional vector using the datasets of 20 systematic studies (see Table 1). This resulted in a

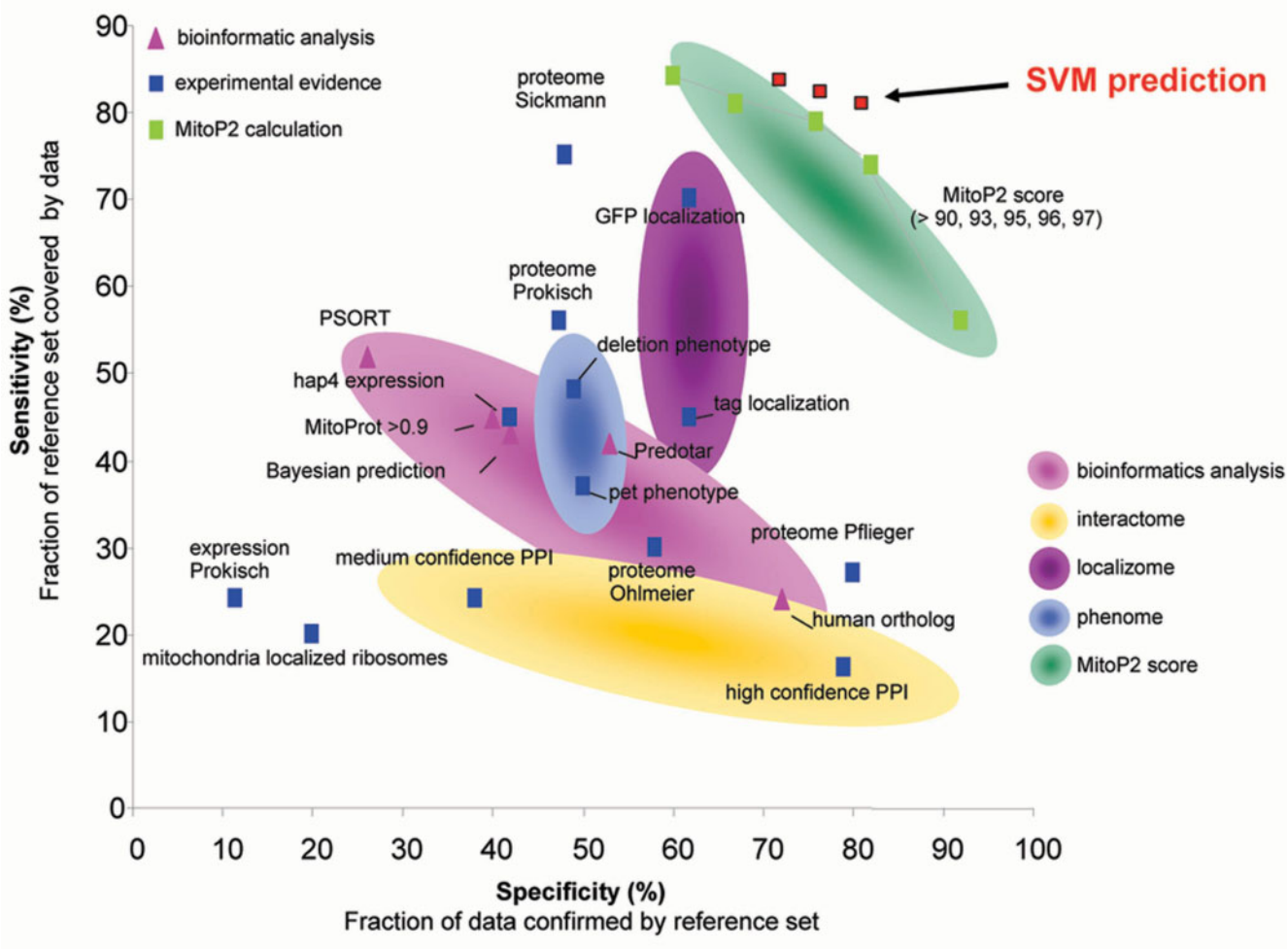

Figure 1. Systematic approaches to identify mitochondrial proteins. The yeast datasets were benchmarked against the mitochondrial reference set. Each point represents a dataset whose position is determined by benchmarking against the 522 reference proteins from MitoP2-Yeast. The different groups of approaches are highlighted using distinct colours: the bioinformatics datasets (purple) are PSORT (25), MitoProt >90 (23), Bayesian prediction (37), Predotar (26) and yeast proteins with human mitochondrial orthologs (MitoP2 database); the experimental datasets (blue) are as follows: hap4 expression (18), respiration induced expression (3), mitochondria localized ribosomes (38), deletion phenotype screen (16), tag localization (14), GFP localization (13), pet phenotypes (15), four mass spectrometry proteome studies (3,19,20,39) and high and medium confidence protein–protein interactions (PPI) (22) defined by interactions with known mitochondrial proteins (MitoP2 database). The predictive score for a mitochondrial protein (MitoP2 score; green) was based on the combination of the systematic datasets, calculated for different thresholds. The predictions using the SVM algorithm are shown in red for different thresholds.

20-dimensional input matrix, which was used to train the SVM (see also Supplementary Figure S1). After training, the SVM predicts mitochondrial proteins with a specificity of 78% and a sensitivity of 80% (SVM score >1). This analysis shows that a combination of datasets from genome-wide studies significantly increases the power of predicting mitochondrial proteins beyond the level achieved by any single study (Figure 1).

MitoP2-HUMAN AND MitoP2-MOUSE

We manually annotated mitochondrial reference proteins for human (624) and mouse (615) that now cover about half of the estimated mitochondrial proteins in these two species. These reference proteins present a subset of all the protein entries in the database: MitoP2-Human contains 36 504 proteins and MitoP2-Mouse contains 32 422 proteins. These datasets have been downloaded from the Swiss-Prot database (http://www.expasy.org/sprot/) (28). To identify putative orthologue proteins between human and mouse we calculated a bidirectional best BLAST hit or a best BLAST hit with score $<1 \times 10^{-10}$ between the two datasets. For each MitoP2 protein, we extracted descriptions, chromosomal positions, subcellular localization and literature references from Swiss-Prot. In addition, functional annotations such as biological processes and functional categories were extracted from the Gene Ontology database (GO; http://www.geneontology.org/). For MitoP2-Mouse, we annotated functional descriptions according to the MIPS functional catalogue (29), and provided access to DNA and protein sequence information. Each of these protein annotations is accompanied by its PubMed reference link. Phenotypic information on available mouse models are provided by the Mouse Genome Informatics database (MGI; http://www.informatics.jax.org/) (30). To date, more than 50 mouse models carrying mutations or deletions of mitochondrial genes have been investigated.

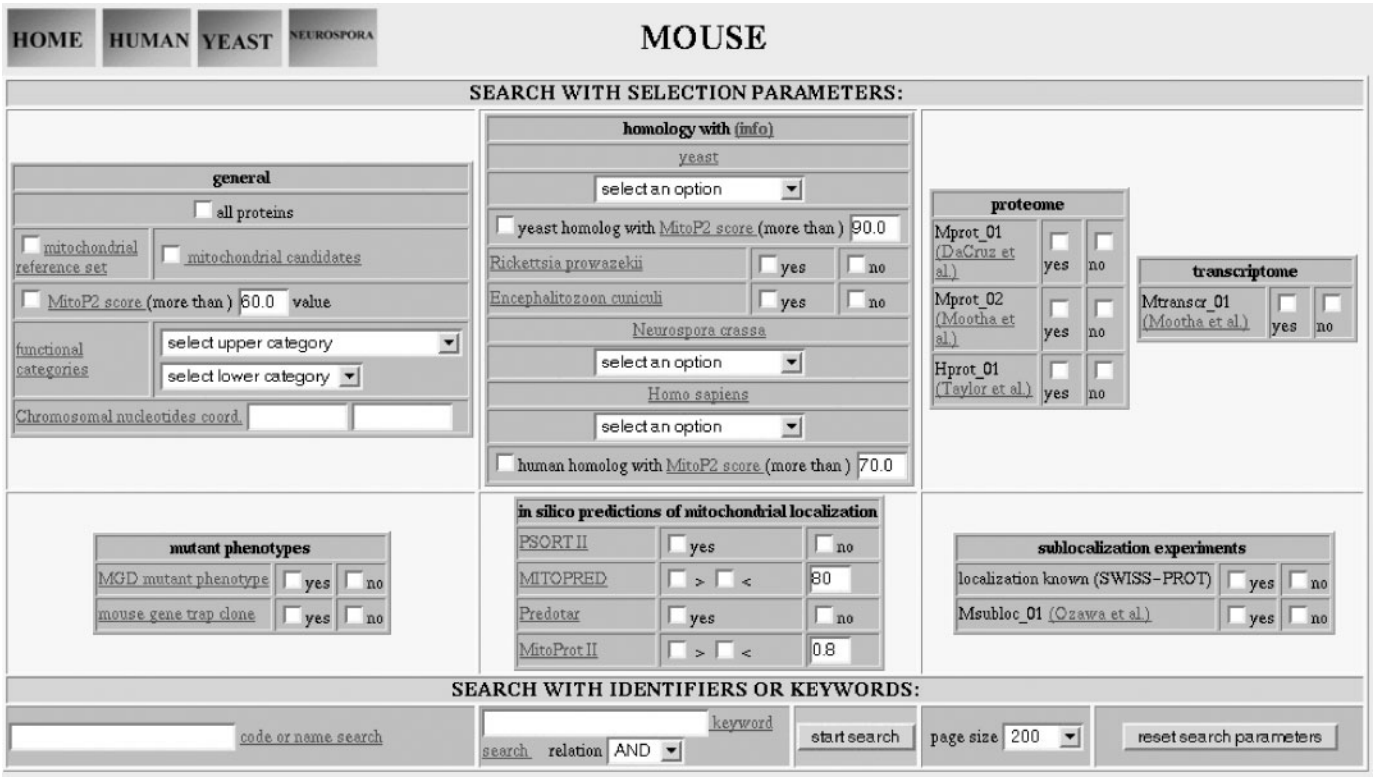

Figure 2. Screenshot of the MitoP2-Mouse query page. The MitoP2 query page is structured according to various groups of search parameters provided by the database. The search options are either linked to the online references or an explanation for this selection is provided.

For researchers interested in studying these models, MitoP2 provides links to the International Gene Trap Consortium (IGTC; http://www.genetrap.org/) to access the related mouse cell lines.

MitoP2-Mouse and MitoP2-Human provide similar search options that allow single or combined searches for individual database components (Figure 2). Database searches and downloads can be performed using keywords, genes names and the selection of datasets from systematic studies. For MitoP2-Human, two proteome studies on mitochondrial organelles purified from heart tissue are available, which have been integrated under the 'proteome' category (31,32). For Mitop2-Mouse, we integrated the datasets from three high-throughput studies that include two proteome experiments (33,34) and a subcellular localization study using split-enhanced green fluorescent protein (EGFP) (35). For mouse proteins identified using these approaches, we identified the respective putative orthologue proteins in human, and *vice versa*. The number of proteins in human and mouse datasets differ in part due to missing proteins in either one of the species in Swiss-Prot. The MitoP2 category 'transcriptome' predicts gene relationships based on similarities of their expression profiles (34). In addition to sequence similarity searches between human, yeast and mouse, MitoP2 provides *in silico* predictions for mitochondrial proteins utilizing established algorithms such as MitoProt II (23), PSORT II (25), Predotar (26) and MITOPRED (24). These programs allow the prediction of subcellular localizations of proteins based on their amino acid sequences. To illustrate the different search functions, users can select PSORT II under MitoP2-Mouse to extract 4321 proteins that include 323 entries from the mitochondrial reference set (7%). Alternatively, one can perform combined searches, for example, by selecting PSORT II and a human proteome dataset 'Hprot_01' (31,32), which then generates a list of 176 proteins that includes 56% of the mitochondrial reference set. This comparison demonstrates the trade-off between sensitivity and specificity: the combination of datasets reduces the total number of proteins (sensitivity), while it increases the specificity for mitochondrial proteins.

Each entry in MitoP2-Human and MitoP2-Mouse corresponds to a Swiss-Prot identifier with protein descriptions, annotated subcellular localization and sequence map positions according to UCSC genome browser (http://genome.ucsc.edu/). In addition, the single protein entry summarizes the information from *in silico* predictions, high-throughput experiments, the availability of mouse gene trap clones and the predictive MitoP2 score. An example for a single protein entry in MitoP2-Mouse, the adenine nucleotide (ADP/ATP) translocator 2, is shown in Figure 3. This figure shows in a matrix lane, the information available for this protein extracted from Swiss-Prot and the integrated genome-wide approaches, a list of functional annotations compiled from the MIPS catalogue that are linked to the Mouse Functional Genome Database (http://mips.gsf.de/genre/proj/mfungd/), PubMed links to the references, and a list of similar sequences from other species. The other parts of this entry, which are not shown in this figure, include a phenotype description of the associated mouse mutant, the Gene Ontology annotations for molecular protein functions and biological processes,

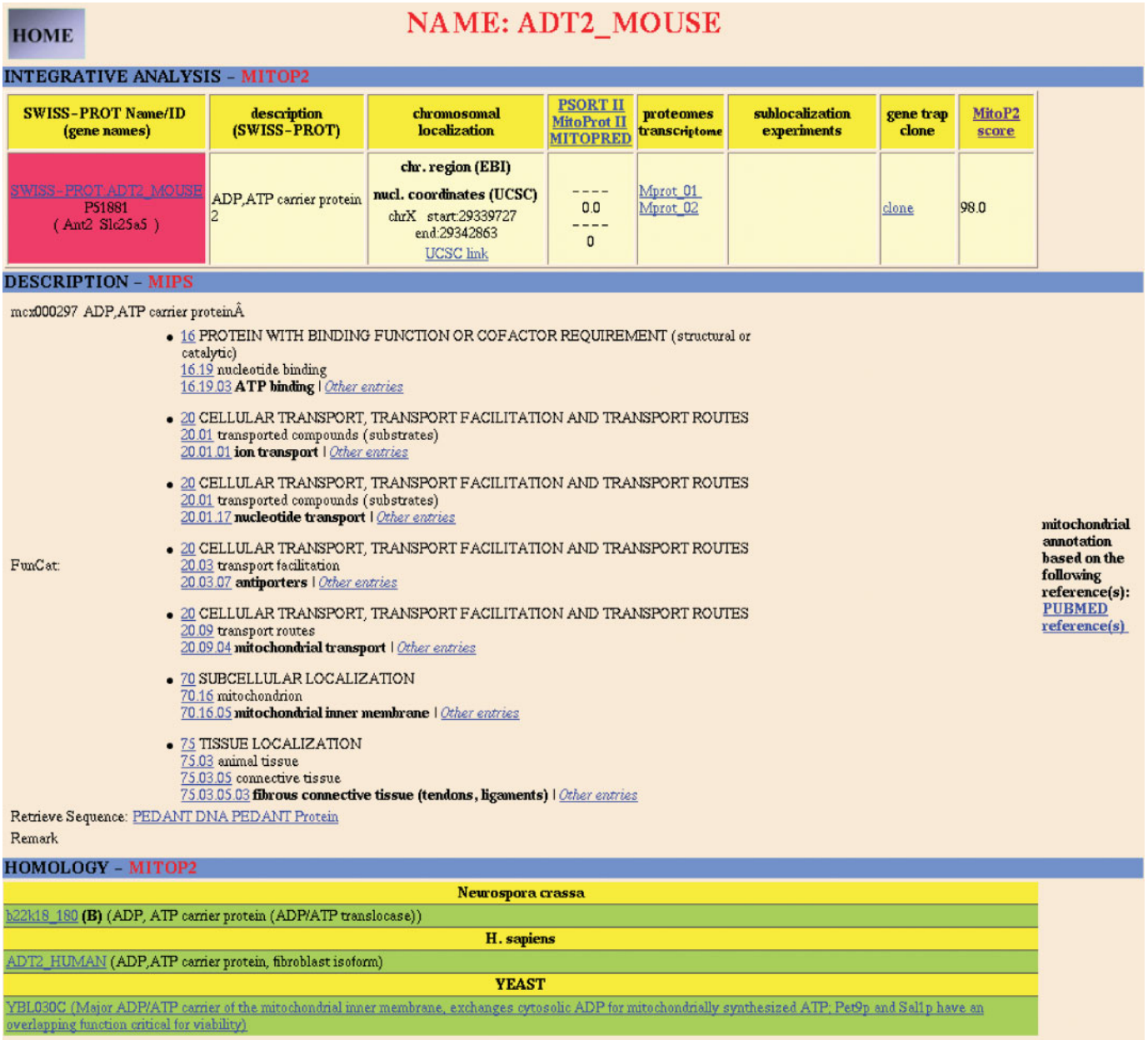

Figure 3. Example for protein entry in MitoP2-Mouse. As illustrated for the mitochondrial ADP/ATP carrier protein 2 (ADT2). MitoP2 provides for each protein entry the Swiss-Prot name and description, the chromosomal localization, results from mitochondrial prediction programs, data from proteome studies, available gene trap clones, functional annotations according to MIPS, PubMed reference links and homologous proteins in other species.

literature references on protein functions and variants that are listed by author names and title and a table of Swiss-Prot references.

For genes implicated in a hereditary disease, MitoP2 provides a link to the corresponding entry in the Online Mendelian Inheritance in Man database (OMIM; http://www.ncbi.nlm.nih.gov) (36). To date, more than 120 of the 624 human mitochondrial proteins are known to be involved in a hereditary disease. Mitochondrial disorders have a diversity of debilitating phenotypes and include a wide variety of neurodegenerative processes, cardiovascular disorders, diabetes mellitus and several cancer types. Many of these disease genes function in the metabolism of amino acids, nucleic acid, fatty acids and lipids, and energy production. The MitoP2 database enables the systematic identification of candidate genes to study mitochondrial diseases (5). Elpeleg *et al*. (8), for example, mapped a locus for hereditary mtDNA depletions associated with mitochondrial encephalomyopathy to a 21 Mb interval on chromosome 13. The mapping coordinates (i.e. 13:40878920 and 13:61359487) were used as a selection criteria to prioritize MitoP2 candidate genes among the 113 genes predicted in this region. In combination with a MitoP2 score >60, three proteins were identified as disease candidate genes. One of these genes (SUCLA2), a mitochondrial reference protein identified in two proteome experiments, was found to be mutated in affected members of the linkage

family. This study demonstrates that human disease genes can be identified using information provided by MitoP2.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The MitoP2 project is funded by the German National Genome Network (German Ministry for Education and Research, grant 01GR0411), BFAM (Bioinformatics for the Functional Analysis of Mammalian Genomes) and the MitEURO consortium. Funding to pay the Open Access publication charges for this article was provided by the GSF Research Centre.

Conflict of interest statement. None declared.

REFERENCES

1. Collins,F.S., Green,E.D., Guttmacher,A.E. and Guyer,M.S. (2003) A vision for the future of genomics research. *Nature*, **422**, 835–847.
2. Taylor,S.W., Fahy,E. and Ghosh,S.S. (2003) Global organellar proteomics. *Trends Biotechnol*., **21**, 82–88.
3. Prokisch,H., Scharfe,C., Camp,D.G.II, , Xiao,W., David,L., Andreoli,C., Monroe,M.E., Moore,R.J., Gritsenko,M.A., Kozany,C. *et al*. (2004) Integrative analysis of the mitochondrial proteome in yeast. *PLoS Biol*., **2**, e160.
4. Andreoli,C., Prokisch,H., Hortnagel,K., Mueller,J.C., Munsterkotter,M., Scharfe,C. and Meitinger,T. (2004) MitoP2, an integrated database on mitochondrial proteins in yeast and man. *Nucleic Acids Res*., **32**, D459–D462.
5. Scharfe,C., Zaccaria,P., Hoertnagel,K., Jaksch,M., Klopstock,T., Dembowski,M., Lill,R., Prokisch,H., Gerbitz,K.D., Neupert,W. *et al*. (2000) MITOP, the mitochondrial proteome database: 2000 update. *Nucleic Acids Res*., **28**, 155–158.
6. Szklarz,L., Guiard,B., Rissler,M., Wiedemann,N., Kozjak,V., van der Laan,M., Lohaus,C., Marcus,K., Meyer,H., Chacinska,A. *et al*. (2005) Inactivation of the mitochondrial heat shock protein Zim17 leads to aggregation of matrix Hsp70s followed by pleiotropic effects on morphology and protein biogenesis. *J. Mol. Biol*., **351**, 206–218.
7. van der Laan,M., Chacinska,A., Lind,M., Perschil,I., Sickmann,A., Meyer,H., Guiard,B., Meisinger,C., Pfanner,N. and Rehling,P. (2005) Pam17 is required for architecture and translocation activity of the mitochondrial protein import motor. *Mol. Cell. Biol*., **25**, 7449–7458.
8. Elpeleg,O., Miller,C., Hershkovitz,E., Bitner-Glindzicz,M., Bondi-Rubinstein,G., Rahman,S., Pagnamenta,A., Eshhar,S. and Saada,A. (2005) Deficiency of the ADP-forming succinyl-CoA synthase activity is associated with encephalomyopathy and mitochondrial DNA depletion. *Am. J. Hum. Genet*., **76**, 1081–1086.
9. Mootha,V.K., Lepage,P., Miller,K., Bunkenborg,J., Reich,M., Hjerrild,M., Delmonte,T., Villeneuve,A., Sladek,R., Xu,F. *et al*. (2003) Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc. Natl Acad. Sci. USA*, **100**, 605–610.
10. Tiranti,V., Hoertnagel,K., Carrozzo,R., Galimberti,C., Munaro,M., Granatiero,M., Zelante,L., Gasparini,P., Marzella,R., Rocchi,M. *et al*. (1998) Mutations of SURF-1 in Leigh disease associated with cytochrome *c* oxidase deficiency. *Am. J. Hum. Genet*., **63**, 1609–1621.
11. Reichert,A.S. and Neupert,W. (2004) Mitochondriomics or what makes us breathe. *Trends Genet*., **20**, 555–562.
12. Cherry,J.M., Ball,C., Weng,S., Juvik,G., Schmidt,R., Adler,C., Dunn,B., Dwight,S., Riles,L., Mortimer,R.K. *et al*. (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature*, **387**, 67–73.
13. Huh,W.K., Falvo,J.V., Gerke,L.C., Carroll,A.S., Howson,R.W., Weissman,J.S. and O'Shea,E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686–691.
14. Kumar,A., Cheung,K.H., Tosches,N., Masiar,P., Liu,Y., Miller,P. and Snyder,M. (2002) The TRIPLES database: a community resource for yeast molecular biology. *Nucleic Acids Res*., **30**, 73–75.
15. Dimmer,K.S., Fritz,S., Fuchs,F., Messerschmitt,M., Weinbach,N., Neupert,W. and Westermann,B. (2002) Genetic basis of mitochondrial function and morphology in *Saccharomyces cerevisiae*. *Mol. Biol. Cell*, **13**, 847–853.
16. Steinmetz,L.M., Scharfe,C., Deutschbauer,A.M., Mokranjac,D., Herman,Z.S., Jones,T., Chu,A.M., Giaever,G., Prokisch,H., Oefner,P.J. *et al*. (2002) Systematic screen for human disease genes in yeast. *Nature Genet*., **31**, 400–404.
17. DeRisi,J.L., Iyer,V.R. and Brown,P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.
18. Lascaris,R., Bussemaker,H.J., Boorsma,A., Piper,M., van der Spek,H., Grivell,L. and Blom,J. (2003) Hap4p overexpression in glucose-grown *Saccharomyces cerevisiae* induces cells to enter a novel metabolic state. *Genome Biol*., **4**, R3.
19. Pflieger,D., Le Caer,J.P., Lemaire,C., Bernard,B.A., Dujardin,G. and Rossier,J. (2002) Systematic identification of mitochondrial proteins by LC-MS/MS. *Anal. Chem*., **74**, 2400–2406.
20. Sickmann,A., Reinders,J., Wagner,Y., Joppich,C., Zahedi,R., Meyer,H.E., Schonfisch,B., Perschil,I., Chacinska,A., Guiard,B. *et al*. (2003) The proteome of Saccharomyces cerevisiae mitochondria. *Proc. Natl Acad. Sci. USA*, **100**, 13207–13212.
21. Ghaemmaghami,S., Huh,W.K., Bower,K., Howson,R.W., Belle,A., Dephoure,N., O'Shea,E.K. and Weissman,J.S. (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737–741.
22. von Mering,C., Krause,R., Snel,B., Cornell,M., Oliver,S.G., Fields,S. and Bork,P. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, **417**, 399–403.
23. Claros,M.G. (1995) MitoProt, a Macintosh application for studying mitochondrial proteins. *Comput. Appl. Biosci*., **11**, 441–447.
24. Guda,C., Fahy,E. and Subramaniam,S. (2004) MITOPRED: a genome-scale method for prediction of nucleus-encoded mitochondrial proteins. *Bioinformatics*, **20**, 1785–1794.
25. Nakai,K. and Horton,P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci*., **24**, 34–36.
26. Small,I., Peeters,N., Legeai,F. and Lurin,C. (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics*, **4**, 1581–1590.
27. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25**, 3389–3402.
28. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. *et al*. (2003) The Swiss-Prot protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*., **31**, 365–370.
29. Ruepp,A., Zollner,A., Maier,D., Albermann,K., Hani,J., Mokrejs,M., Tetko,I., Guldener,U., Mannhaupt,G., Munsterkotter,M. *et al*. (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res*., **32**, 5539–5545.
30. Bult,C.J., Blake,J.A., Richardson,J.E., Kadin,J.A., Eppig,J.T., Baldarelli,R.M., Barsanti,K., Baya,M., Beal,J.S., Boddy,W.J. *et al*. (2004) The Mouse Genome Database (MGD): integrating biology with the genome. *Nucleic Acids Res*., **32**, D476–D481.
31. Gaucher,S.P., Taylor,S.W., Fahy,E., Zhang,B., Warnock,D.E., Ghosh,S.S. and Gibson,B.W. (2004) Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid chromatography coupled with tandem mass spectrometry. *J. Proteome. Res*., **3**, 495–505.
32. Taylor,S.W., Fahy,E., Zhang,B., Glenn,G.M., Warnock,D.E., Wiley,S., Murphy,A.N., Gaucher,S.P., Capaldi,R.A., Gibson,B.W. *et al*. (2003) Characterization of the human heart mitochondrial proteome. *Nat. Biotechnol*., **21**, 281–286.
33. Da Cruz,S., Xenarios,I., Langridge,J., Vilbois,F., Parone,P.A. and Martinou,J.C. (2003) Proteomic analysis of the mouse liver mitochondrial inner membrane. *J. Biol. Chem*., **278**, 41566–41571.
34. Mootha,V.K., Bunkenborg,J., Olsen,J.V., Hjerrild,M., Wisniewski,J.R., Stahl,E., Bolouri,M.S., Ray,H.N., Sihag,S., Kamal,M. *et al*. (2003) Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell*, **115**, 629–640.

35. Ozawa,T., Sako,Y., Sato,M., Kitamura,T. and Umezawa,Y. (2003) A genetic approach to identifying mitochondrial proteins. *Nat. Biotechnol.*, **21**, 287–293.
36. Hamosh,A., Scott,A.F., Amberger,J., Bocchini,C., Valle,D. and McKusick,V.A. (2002) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **30**, 52–55.
37. Drawid,A. and Gerstein,M. (2000) A Bayesian system integrating expression data with sequence patterns for localizing proteins: comprehensive application to the yeast genome. *J. Mol. Biol.*, **301**, 1059–1075.
38. Marc,P., Margeot,A., Devaux,F., Blugeon,C., Corral-Debrinski,M. and Jacq,C. (2002) Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Rep.*, **3**, 159–164.
39. Ohlmeier,S., Kastaniotis,A.J., Hiltunen,J.K. and Bergmann,U. (2004) The yeast mitochondrial proteome, a study of fermentative and respiratory growth. *J. Biol. Chem.*, **279**, 3956–3979.