D964–D971 *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 16 November 2011*
doi:10.1093/nar/gkr936

dbDEPC 2.0: updated database of differentially expressed proteins in human cancers

Ying He[1,2], Menghuan Zhang[2,3], Yuanhu Ju[2], Zhonghao Yu[4], Daqing Lv[2], Han Sun[1], Weilan Yuan[5], Fei He[2], Jianshe Zhang[2], Hong Li[1,2], Jing Li[3], Rui Wang-Sattler[4], Yixue Li[1,2], Guoqing Zhang[2,*] and Lu Xie[2,*]

[1]Key Laboratory of Systems Biology, Chinese Academy of Sciences, Shanghai 200031, [2]Shanghai Center for Bioinformation Technology, Shanghai 200235, [3]Department of Bioinformatics and Biostatistics, Shanghai Jiaotong University, Shanghai 200240, P. R. China, [4]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg 85764, Germany and [5]Biomedical Engineering for School of Life Sciences and Technology, Tongji University, Shanghai 200092, P. R. of China

Received August 12, 2011; Revised October 10, 2011; Accepted October 11, 2011

ABSTRACT

A large amount of differentially expressed proteins (DEPs) have been identified in various cancer proteomics experiments, curation and annotation of these proteins are important in deciphering their roles in oncogenesis and tumor progression, and may further help to discover potential protein biomarkers for clinical applications. In 2009, we published the first database of DEPs in human cancers (dbDEPCs). In this updated version of 2011, dbDEPC 2.0 has more than doubly expanded to over 4000 protein entries, curated from 331 experiments across 20 types of human cancers. This resource allows researchers to search whether their interested proteins have been reported changing in certain cancers, to compare their own proteomic discovery with previous studies, to picture selected protein expression heatmap across multiple cancers and to relate protein expression changes with aberrance in other genetic level. New important developments include addition of experiment design information, advanced filter tools for customer-specified analysis and a network analysis tool. We expect dbDEPC 2.0 to be a much more powerful tool than it was in its first release and can serve as reference to both proteomics and cancer researchers. dbDEPC 2.0 is available at http://lifecenter.sgst.cn/dbdepc/index.do.

INTRODUCTION

Cancer is becoming one of the leading causes of death worldwide (1). In the past decade, advancements in high-throughput technologies, such as genome sequencing, expression microarrays and mass spectrometry (MS) have greatly improved our ability to understand the landscape of cancers (2) and helped to detect biomarkers in cancer diagnosis, prognosis and therapy which lead to better patient care (3,4).

Several distinguished cancer-related recourses such as the Cancer Genome Atlas (TCGA) (5), Cancer Genome Anatomy Project (CGAP) (6) and Oncomine (7) have been developed to explore human cancer genes at the level of the genome and mRNA. Many studies (8,9) have proved the efficiency and importance of investigating changes in cancer protein level using the technology of MS. A great amount of cancer proteomics data have been accumulating and require management of easy access and systematic analysis. However, protein landscape is more heterogeneous and complex than DNA and RNA, proteomics technology is still in pursuit of accuracy, coverage and repeatability (10). Few databases provide information on cancer proteome, let alone quantitative MS-based proteomic data. In 2001, a database in lung cancer (11) collected both genomic and proteomic data. Last year came out the Genome Medicine Database of Japan Proteomics (GeMDBJ) (12) collecting cancer proteome expression data produced by two-dimensional polyacrylamide gel electrophoresis (2D PAGE). Human Protein Atlas (13) database concentrates on protein

*To whom correspondence should be addressed. Tel: +86 21 6131 3672; Fax: +86 21 5406 5058; Email: xielu@scbit.org
Correspondence may also be addressed to Guoqing Zhang. Tel: +86 21 5406 5260; Fax: +86 21 5406 5058; Email: gqzhang@scbit.org

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

expression images in normal and cancerous tissues and cells generated by antibody-based immunofluorescence.

In 2009, we reported the first dbDEPC (14) that aimed to provide an overview of protein-level expression changes mainly detected by MS technology. This resource allows researchers to search their interested proteins and cancers; the retrieved protein entry provides the differential expression pattern seen in cancers, along with detailed annotations from various MS experiments. It also associated the differential proteins with genomic aberrance, such as single-nucleotide polymorphisms (SNPs), which might provide possible explanations from the genetic point of view. However, dbDEPC could have been more useful had there been more related data. With fast increment of cancer quantitative proteomic data in recent years, we have updated this database. In the new version of dbDEPC 2.0, the DEPs deposited have been doubled to over 4000, the cancer types increased to 20, and for the first time, 18 subtypes of some cancers are included. Altogether, the curated experiments have grown from 65 to 331. Other than largely expanded data volume, the search and browser functions are retained and enhanced, and quite a few new features have been added. Detailed experiment information is recorded, which allows for experiment to experiment comparison and filtering of results according to user-specific interests. This will give a more specific protein expression heatmap with a clearer biological indication, avoiding ambiguous comparison across heterogeneous experiments. Proteins usually do not function as independent events; rather they interact with each other to fulfill biological process. In dbDEPC 2.0, all DEPs can be viewed as networks in certain cancer types, showing the differential protein regulation network. Queried proteins can also be linked to DEP networks to see their possible involvement in cancer. We expect that both proteomics and cancer researchers can be benefited from the more resourceful updated version of dbDEPC. dbDEPC 2.0 is freely available to public domain at http://lifecenter.sgst.cn/dbdepc/index.do

DATA COLLECTION AND DATABASE CONSTRUCTION

Data collection goes through the following process:

(1) An automatically text mining was conducted on PubMed abstracts using names of cancer types (Table 1) in MeSH, MS-related words (MS, quantitative proteomic) and keywords describing expression changes (upregulated, downregulated and fold change). The current version of dbDEPC indexed papers published before March 2011.
(2) To control the data quality, each deposited data set went through the same rigorous manual review process as described in the previous version (14).
(3) Different identifiers (IDs) or names of DEPs were extracted from papers and uniformly mapped to UniprotKB (15) accession (3rd May 2011).
(4) To give additional annotation of each DEP, new external resources were integrated including the HUGO Gene Nomenclature Committee (HGNC) (16), Gene Ontology (17), Kyoto Encyclopedia of Genes and Genomes (KEGG) (18), STRING (19) and CanProVar (20).

Table 1. Human cancer types in dbDEPC 2.0

Cancer type	Subtype[a]
Lung adenocarcinoma	Non-small cell lung carcinoma
	Small cell lung carcinoma
HCC	Hepatitis C virus
	Hepatitis B virus
Breast cancer	Breast ductal carcinoma
Pancreatic carcinoma	Pancreatic ductal adenocarcinoma
Leukemia	Chronic myeloid leukemia
	Chronic lymphocytic leukemia
	Acute myeloid leukemia
	Acute lymphoblastic leukemia
Thyroid cancer	Papillary thyroid carcinoma
	Follicular thyroid carcinoma
	Follicular thyroid adenoma
Skin cancer[a]	Melanoma
	Non-melanoma
Brain tumor[a]	Neuroblastoma
Head and neck cancer[a]	Oral cancer
	Oral premalignant lesions
Gastric cancer	
Colorectal cancer	
Prostate cancer	
Esophageal cancer	
Cervical cancer	
Ovarian cancer	
Renal cell carcinoma[a]	
Lymphoma[a]	
Sarcoma[a]	
Testicular cancer[a]	
Gall bladder cancer[a]	

[a]Marked the new human cancer types and subtypes in dbDEPC 2.0.

In dbDEPC 2.0, data are managed by MySQL5 (http://www.mysql.com/) and a dynamic web interface is constructed with J2EE technology (http://www.oracle.com/technetwork/java/javaee) with a Tomcat server (http://tomcat.apache.org/). The tool modules including heatmap profile, network tool and venn-diagram-based experiment comparison tool are developed with R (http://www.r-project.org/).

DATABASE CONTENTS

In the current release, dbDEPC contains information on 4092DEPs in 20 different human cancer types and 18 subtypes (Table 1). The number of DEPs in breast cancer, hepatocellular carcinoma (HCC) and testicular cancer augmented dramatically during the past 2 years (Figure 1A). However, in all of the DEPs identified by the proteomics approach, only 5% were validated by the low-throughput assays such as western blot, immunohistochemistry, etc (Figure 1B). dbDEPC now has documented 331 MS experiments from 241 peer-reviewed publications. Data sets of proteomics studies on cervical cancer, gastric cancer, HCC and breast cancer have sharp increments by 14, 9.67, 8.5 and 5.69 times, respectively (Figure 1C).

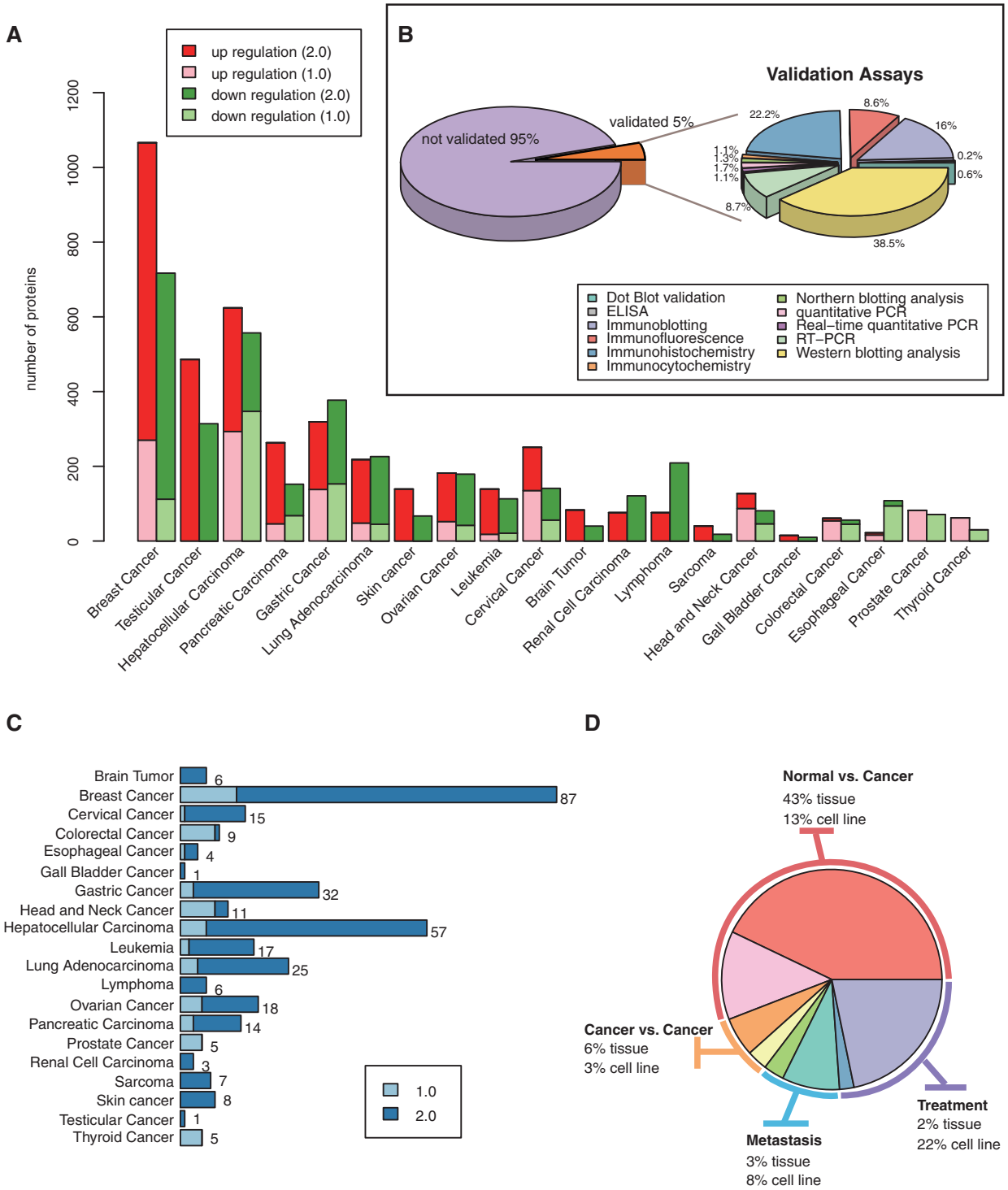

Figure 1. Data content additions in dbDEPC 2.0. (**A**) Number of DEPs in each cancer from version 1.0 to 2.0. (**B**) The percentage of DEPs validated by low-throughput assays. (**C**) Increasing number of experiment datasets in each cancer. (**D**) The percentage of experiment data sets in four types of experimental design.

NEW FEATURES

Better experimental descriptions

We have refined the experimental descriptions following the framework of PRIDE database (21), a public repository standard for proteomics research. The experiment page now delineates publication reference, experimental design, sample information, protocol for protein identification and quantification procedure, the total number of identified proteins (peptides and spectra) and the definition of threshold for differential expression. According to the experimental designs, all datasets could be

categorized into four types of studies, i.e. normal versus cancer comparison, cancer versus cancer comparison, cancer metastasis studies and treatment research. They account for 56%, 9%, 11% and 24% of total experiments, respectively (Figure 1D).

New search and filter tools

In addition to query by proteins or cancers in the last version, the search tool now allows users to browse all the experiments. Regardless of query from proteins, cancers or experiments, dbDEPC 2.0 first provides the related experiments briefly, showing the experiment ID, cancer name, experimental design, control sample, case sample and the total number of upregulated and downregulated DEPs in the experiment (Figure 2A), which give users a first sight of their interested experiments.

We now provide various optional filters in search function to help users focus on their most interested results. The filter tools cover experimental design, cancer name, sample type (tissue or cell type), sample size (<10, 10–50, 50–100 and >100), organism (*Homo sapiens*, *Mus musculus* and *Rattus norvegicus*), MS technology (LC-MS/MS, MALDI–TOF MS, etc.), proteomics search engine (Mascot, Sequest, etc.) and protein quantification approach (two-dimensional gel electrophoresis (2DE), 2-D Fluorescence Difference Gel Electrophoresis (2D-DIGE), Isotope-labeling and label-free). After filtering the experiment results, users can select them and click the button 'View Proteins' to navigate the DEPs list that gives the UniprotKB accession of each DEP and a quick overview of its differential expression ratio in each experiment (Figure 2B). The protein list is visualized by three tags. Under the 'All' tag it shows all the results; the 'Validated' tag displays only the DEPs validated by low-throughput assays and the 'Not validated' list the complementary candidates, which offer possible chance for further verification.

New outlook of heatmap profiles

dbDEPC profile page creating a heatmap facilitates users to investigate differential expression profiles among different cancers. In the last version, we found 63 proteins showing conflicting expression changes (up or down) in six studies on HCC, which might be due to different samples or heterogeneous experimental designs. To avoid the imbalance among different types of experiments, in this version we allow users to extract data sets from identical experimental design to create the heatmaps. Users input a list of proteins, select cancers of interest as before, and pick one type of experimental design (Figure 3A). The page will return a heatmap visualizing the expression change of query proteins across multiple cancers and a table listing the DEPs briefly (Figure 3B). Up and downregulations are represented by red and green color, respectively, for normal versus cancer comparison and the color grades correspond to the confidence scores. For each DEP, the score is calculated as the number of homogeneous experiments on certain cancer supporting its up or downexpression.

DEPs association network

Another new feature we would like to highlight is the DEPs association network, a tool in attempt to find the association of query proteins with reported DEPs in particular cancers. The association network background came from the STRING 9.0 (19). Users input query proteins, select one experimental design, one or several interested cancers and set an association confidence score threshold. The database will present a network of the first-layer-associated DEPs with the query proteins, and the associated protein list in particular cancers (Figure 3C). We also provide a downloadable zip file containing two files. One describes the links between the DEPs and query proteins, and the other represents the expression patterns (upregulated or downregulated) of the DEPs in the network. These two files can be imported by Cytoscape (http://www.cytoscape.org/) that allows users to modify the networks by themselves. This tool helps users find the associated candidate markers reported in other studies, which provide further information to the queried proteins and give possible hint for their next step research.

This network tool also demonstrates a cancer-specific protein differential regulation network, and helps detect important hub proteins from a systematic view. We calculated the degree of each node (namely, the number of connections or edges the node has to other nodes) in our database. As the basic topological network measurement, degrees of nodes provide insights into the important architecture of the nodes of interest in the whole network (22). Figure 3D displays degree distribution and the average degree of the nodes in each protein set. The average degree of 2317 DEPs was 22.48 and was significantly higher than that of random networks of counterpart size (ranged from 8.82 to 12.49, Wilcoxon's test, $P = 1.9 \times 10^{-18}$). This revealed that cancer proteins show an increase in the number of proteins they interact with, and also appear to participate in central hubs (23). Through sorting the degree of each DEP in the association network, we listed the top 10 proteins with highest degree in Figure 3D and found that most of them are cancer related. MYC, coding for a transcription factor, connected with 106 DEPs, is believed to regulate expression of 15% of all genes (24). As one of the most important oncogene, MYC is reported to drive cell proliferation, regulate cell growth and also play a great role in differentiation and stem cell self-renewal (25,26). Another hub protein epidermal growth factor receptor (EGFR) involved in cell proliferation and cancer progressing has been also known to associate with a number of cancers (27). Through the network tool, we provide a global view of cancer proteins and reveal their roles as central hubs to connect and regulate other proteins in cancer cells.

Enhanced protein page

In this version, the protein page of each DEP has been thoroughly revised. The enhanced page guided by a navigation bar exhibits in seven parts: Protein Summary, Cancer Profile, MS Experiment, Validation Assays, Sequence Variation, Association DEPs and Function

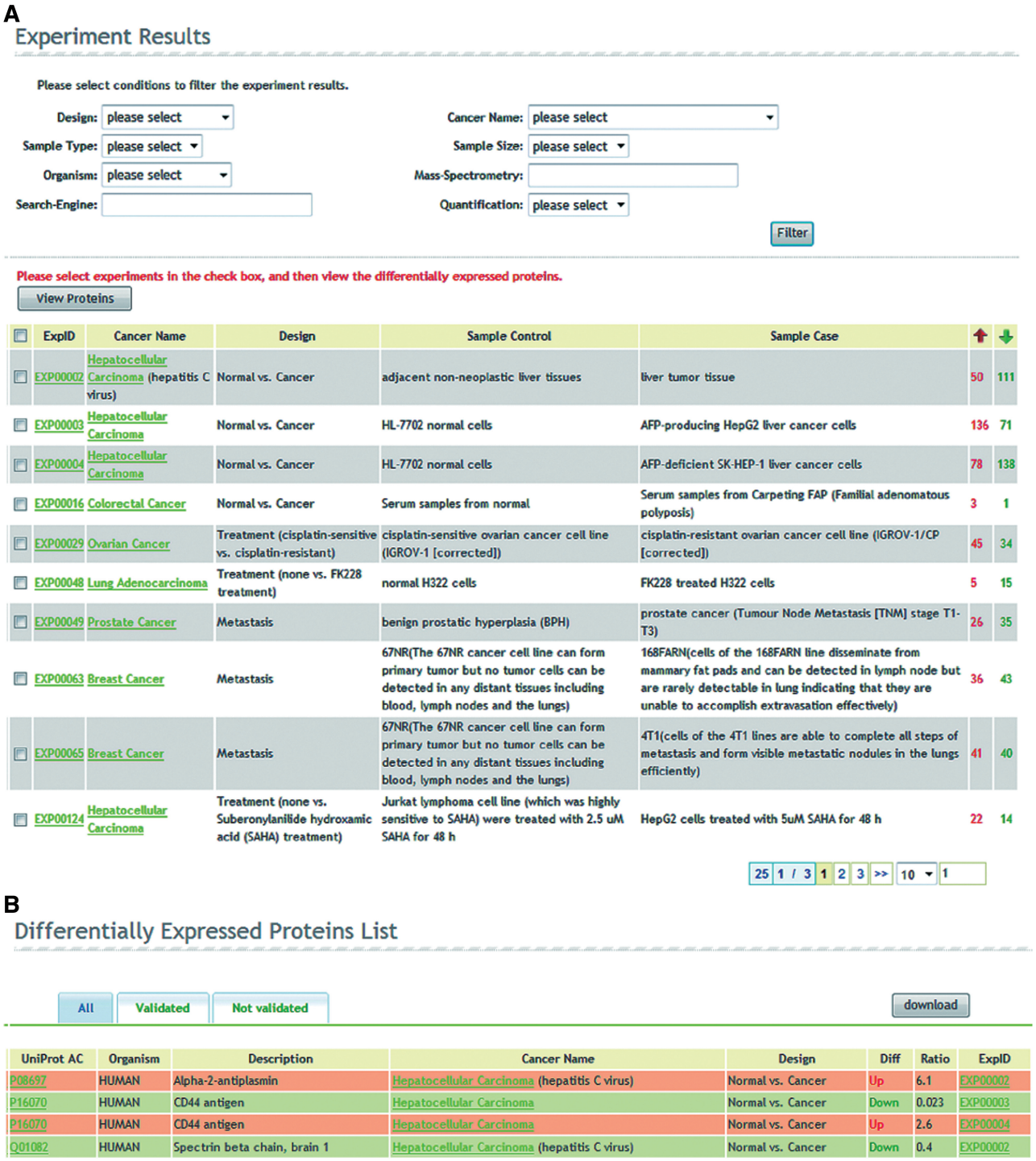

Figure 2. The web interface of advanced search and filter tools. (**A**) Experiment results page. Regardless of query from proteins, cancers or experiments, dbDEPC returns the experiment results and provides various filters in search function. After filtering the experiments, users can select them and click the button 'View Proteins' to navigate to the DEPs list. (**B**) DEP list page. The page summarizes the query result matched proteins and provides a quick overview of their differential expression in each experiment.

Annotation (Supplementary Figure S1). In the summary section, IPI IDs were replaced by UniprotKB (15) accessions due to the closure of IPI database (28). Besides protein ID and name, function and subcellular information are included to delineate the summary information for each DEP. The 'Cancer Profile' demonstrates four heatmaps of the differential expression profiles of the protein across cancers corresponding to the four types of experimental designs. The experiment details could be found in the 'MS Experiment' part. 'Validation Assays' provide confirmation by low scale assays such as western blot, immunohistochemistry, etc. Protein sequence

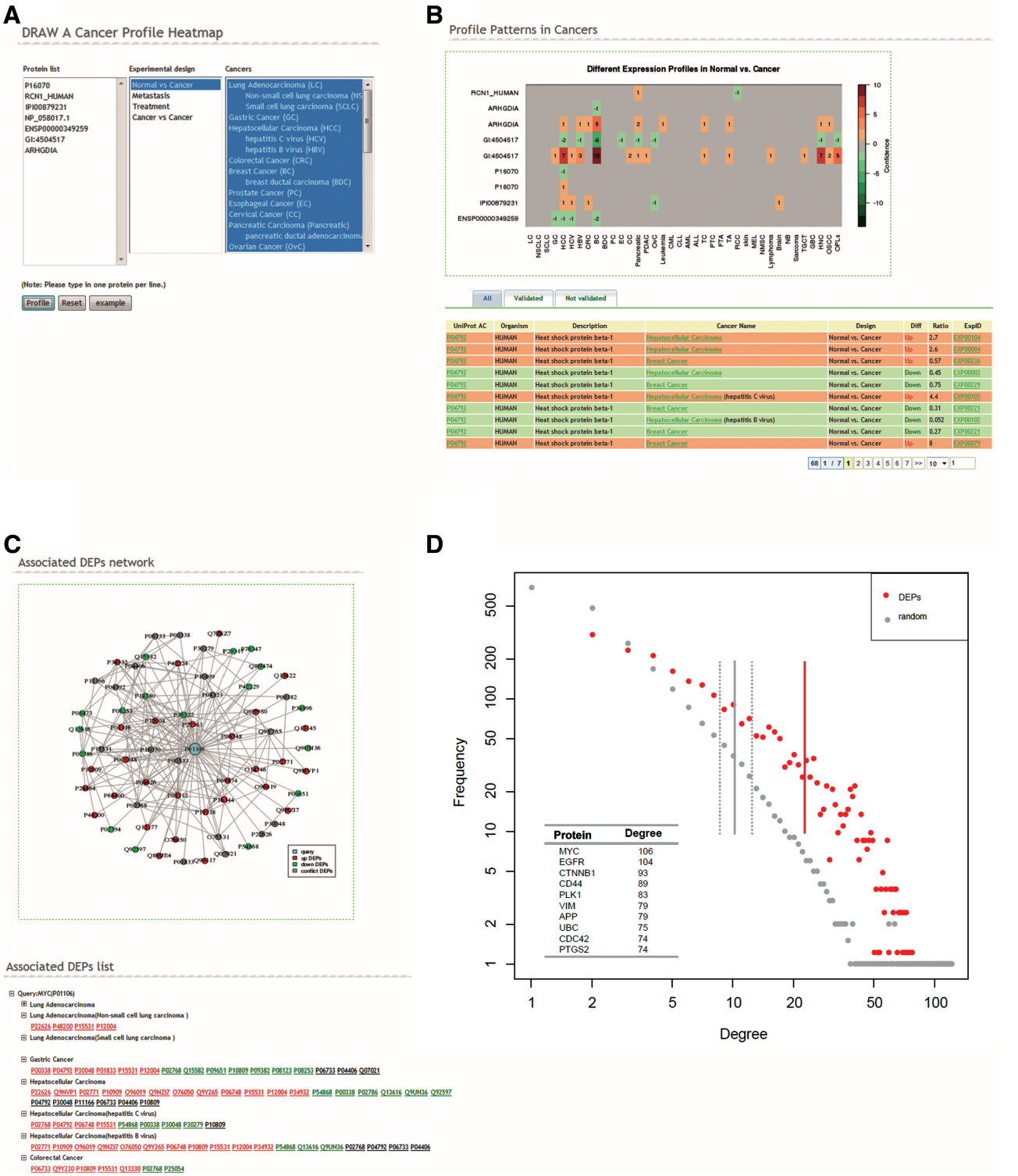

Figure 3. Enhanced profile and new network tools. (**A**) Profile page. Users input a list of proteins, select cancers of interest and pick one type of experimental design to generate a differential expression heatmap in certain type of experiments. (**B**) Example heatmap and DEPs list. The heatmap visualizes the differential expression profiles of DEPs in multiple cancers in the experiments focusing on normal versus cancer comparison. (**C**) Example network result page. The page demonstrates DEPs association network of MYC in query cancers, and the associated DEPs list. Red or green nodes indicate upregulated or downregulated, while black ones are in conflict. (**D**) Degree distribution and the average degree of the nodes in DEPs association network and in the random protein association network of counterpart size. The top 10 DEPs with the highest degree were listed at the 0.7 confidence score.

variations from CanProVar (20) are highlighted with yellow or green color indicating cancer related or just from dbSNP, respectively, which may provide users possible explanations to the protein differential expression in cancers. 'Association DEPs' section facilitates users to find interested associated DEPs for further information. The last part 'Function Annotation' displays the biological descriptions of the protein-involved KEGG pathways (18) and related Gene Ontology functions (17).

OTHER NEW FEATURES

In this version of dbDEPC, we also provide a new Venn-diagram-based experiment comparison tool in the search result page through the 'Intersection' button, which provides a straightforward comparison between upregulated or downregulated DEPs lists from two or three experiments.

In addition to license free download for all academic users, we now open a user upload system to invite other researchers to share their findings. The upload files or publications will be manually reviewed, extracted and deposited into database if the data meet our quality standards.

DISCUSSION

Proteomic technology nowadays allows researchers to view protein change quantitatively in cancer patients versus their healthy counterparts, thus lead to deeper understanding behind the protein function. With the considerable increasing of cancer proteomic data, dbDEPC is expected to provide a resource to facilitate cancer research at protein marker level.

Here is a possible scenario of how dbDEPC could benefit cancer studies. Suppose a researcher would like to find the DEPs in metastatic breast cancer; as he/she queries by breast cancer and filters experimental design by 'Metastasis', he/she can see 24 experiments at current dbDEPC. Further, he/she wants the samples to be human tissues so he/she filters the sample type by 'tissue' and organism by '*Homo sapiens*', eight experiments would meet such criteria. Clicking experiment ID, more information about experimental design and biological background would show up. Now the eight experiments can be viewed by categories. For example, focusing on breast cancer concerning lymph node metastases, he/she selects three experiments (EXP00095, EXP00213 and EXP00215). The results can be displayed by two choices: 'View Proteins' or 'Intersection'. (i) Click the 'View Proteins' button, all DEPs identified in three experiments are provided as a downloadable tab separate text file, in this case, 94 proteins altogether. Each is marked with upregulated or downregulated in metastasis of that experiment. Clicking 'Validated' tag, the user can see that only eight proteins (O00299, O75083, P07339, P13796 downregulated and P00918, P01011, P09211, P50454 upregulated) were validated by traditional biochemical assays. For each DEP, the detailed annotation can be found on the protein information page. (ii) Back to 'Intersection' button. It is an experiment comparison tool; in this case, the three experiments come up with two venn-diagrams showing the intersection of upregulated and downregulated proteins, respectively, in these experiments. Protein numbers that were identified by multiple experiments can be seen such as, one upregulated protein (P20774) identified by two studies (EXP00095 and EXP00213), etc.

Moreover, even if the user-queried proteins are not among the DEPs in dbDEPC, the network tool can help link them to the associated DEPs in cancers, thus to provide clues for their involvement in cancer networks. The network tool also provides a systematic view of differential protein interaction network in cancers. From the above analysis on degree distribution of DEPs network, we confirmed that proteins in cancer-specific network appear to participate in central hubs (e.g. the most famous MYC).

dbDEPC is committed to be an potential reference database forDEPs in human cancers. Such layered curation and annotation of cancer-related proteins could be useful in better understanding of both the value of proteomics study in cancer research and the biological meaning of protein expression change in certain cancer stages. Metaanalyses are also possible by comparing the experiments of the like, and thus to ensure possible biomarker selection and predictive model construction.

FUTURE DIRECTIONS

Future update planned for dbDEPC will include refinement of data sets from treatment design experiments. Currently, the data sets from treatment experiments are heterogeneous. Samples were compared under various conditions such as no treatment versus drug treatment, small molecule sensitive versus small molecule resistant and so forth. We plan to redescribe the data sets and provide a more delicate description for each treatment experiment. Another feature planned is a profile similarity algorithm to compare the query differential expression profiles with the known DEPs profiles to find the similar profile results. We will keep on documenting the validation assays on the roles of particular proteins in certain cancers from external low-throughput biochemical experiments and functional analyses. More available high-throughput protein expression data sets on same cancer may cross validate each other. Currently, dbDEPC mainly focuses on human cancers. Collecting protein expression data of other model species like mouse and rat may be under future construction.

dbDEPC has been continuing to grow and as always we encourage users' feedbacks including error reports and feature requests, we hope to make dbDEPC a comprehensive resource to facilitate cancer proteomic research and may be in the end contribute to cancer treatment.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figure S1.

ACKNOWLEDGEMENTS

The authors acknowledge the Shanghai Guidance of Science and Technology for the offer of abstracts and full texts of publications collected in dbDEPC. We also acknowledge Drs Jia Jia, Yuchen Shen and Yangfan Guo from Shanghai Center for Bioinformation Technology for help with the revision of the database.

FUNDING

Funding for open access charge: National Key Basic Research Program; High Technology Development Project from the Ministry of Science and Technology (2010CB912702, 2009AA02Z304); Key Infectious Disease Project (2012ZX10002012-014); National Natural Science Foundation of China (31070752).

Conflict of interest statement. None declared.

REFERENCES

1. Jemal,A., Bray,F., Center,M.M., Ferlay,J., Ward,E. and Forman,D. (2011) Global cancer statistics. *CA Cancer J. Clin.*, **61**, 69–90.
2. Chari,R., Thu,K.L., Wilson,I.M., Lockwood,W.W., Lonergan,K.M., Coe,B.P., Malloff,C.A., Gazdar,A.F., Lam,S., Garnis,C. *et al.* (2010) Integrating the multiple dimensions of genomic and epigenomic landscapes of cancer. *Cancer Metastasis Rev.*, **29**, 73–93.
3. Caprioli,R.M. (2005) Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy. *Cancer Res.*, **65**, 10642–10645.
4. Kulasingam,V., Pavlou,M.P. and Diamandis,E.P. (2010) Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. *Nat. Rev. Cancer*, **10**, 371–378.
5. McLendon,R., Friedman,A., Bigner,D., Van Meir,E.G., Brat,D.J., Mastrogianakis,G.M., Olson,J.J., Mikkelsen,T., Lehman,N., Aldape,K. *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.
6. Strausberg,R.L., Buetow,K.H., Emmert-Buck,M.R. and Klausner,R.D. (2000) The cancer genome anatomy project: building an annotated gene index. *Trends Genet.*, **16**, 103–106.
7. Rhodes,D.R., Kalyana-Sundaram,S., Mahavisno,V., Varambally,R., Yu,J., Briggs,B.B., Barrette,T.R., Anstet,M.J., Kincead-Beal,C., Kulkarni,P. *et al.* (2007) Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*, **9**, 166–180.
8. Everley,P.A., Krijgsveld,J., Zetter,B.R. and Gygi,S.P. (2004) Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. *Mol. Cell. Proteomics*, **3**, 729–735.
9. Lin,Z., Crockett,D.K., Jenson,S.D., Lim,M.S. and Elenitoba-Johnson,K.S.J. (2004) Quantitative proteomic and transcriptional analysis of the response to the p38 mitogen-activated protein kinase inhibitor SB203580 in transformed follicular lymphoma cells. *Mol. Cell. Proteomics*, **3**, 820–833.
10. Service,R.F. (2008) Proteomics ponders prime time. *Science*, **321**, 1758–1761.
11. Oh,J.M., Brichory,F., Puravs,E., Kuick,R., Wood,C., Rouillard,J.M., Tra,J., Kardia,S., Beer,D. and Hanash,S. (2001) A database of protein expression in lung cancer. *Proteomics*, **1**, 1303–1319.
12. Kondo,T. (2010) Cancer proteome-expression database: Genome Medicine Database of Japan Proteomics. *Expert Rev. Proteomics*, **7**, 21–27.
13. Uhlen,M., Oksvold,P., Fagerberg,L., Lundberg,E., Jonasson,K., Forsberg,M., Zwahlen,M., Kampf,C., Wester,K., Hober,S. *et al.* (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotech.*, **28**, 1248–1250.
14. Li,H., He,Y., Ding,G., Wang,C., Xie,L. and Li,Y. (2009) dbDEPC: a database of Differentially Expressed Proteins in human Cancers. *Nucleic Acids Res.*, **38**, D658–D664.
15. Magrane,M. and Consortium,U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database*, doi:10.1093/database/bar009.
16. Seal,R.L., Gordon,S.M., Lush,M.J., Wright,M.W. and Bruford,E.A. (2011) genenames.org: the HGNC resources in 2011. *Nucleic Acids Res.*, **39**, D514–D519.
17. Berardini,T.Z., Khodiyar,V.K., Lovering,R.C. and Talmud,P. (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res*, **38**, D331–D335.
18. Kanehisa,M. (2002) The KEGG database. *Novartis Found. Symp.*, **247**, 91–101; discussion 101–103, 119–128, 244–252.
19. Jensen,L.J., Kuhn,M., Stark,M., Chaffron,S., Creevey,C., Muller,J., Doerks,T., Julien,P., Roth,A., Simonovic,M. *et al.* (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, D412–D416.
20. Li,J., Duncan,D.T. and Zhang,B. (2010) CanProVar: a human cancer proteome variation database. *Hum. Mutat.*, **31**, 219–228.
21. Vizcaino,J.A., Cote,R., Reisinger,F., Barsnes,H., Foster,J.M., Rameseder,J., Hermjakob,H. and Martens,L. (2009) The Proteomics identifications database: 2010 update. *Nucleic Acids Res.*, **38**, D736–D742.
22. Barabasi,A.-L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.
23. Jonsson,P.F. and Bates,P.A. (2006) Global topological features of cancer proteins in the human interactome. *Bioinformatics*, **22**, 2291–2297.
24. Gearhart,J., Pashos,E.E. and Prasad,M.K. (2007) Pluripotency redux–advances in stem-cell research. *N. Engl. J. Med*, **357**, 1469–1472.
25. Cole,M.D. (1986) The myc oncogene: its role in transformation and differentiation. *Annu. Rev. Genet.*, **20**, 361–384.
26. Meyer,N. and Penn,L.Z. (2008) Reflecting on 25 years with MYC. *Nat. Rev. Cancer*, **8**, 976–990.
27. Normanno,N., de Luca,A., Bianco,C., Strizzi,L., Mancino,M., Maiello,M.R., Carotenuto,A., De Feo,G., Caponigro,F. and Salomon,D.S. (2006) Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*, **366**, 2–16.
28. Kersey,P.J., Duarte,J., Williams,A., Karavidopoulou,Y., Birney,E. and Apweiler,R. (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics*, **4**, 1985–1988.