DOI 10.1002/pmic.200700919

VIEWPOINT

Déjà vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins

Jiri Petrak[1, 2], Robert Ivanek[3], Ondrej Toman[1], Radek Cmejla[1], Jana Cmejlova[1], Daniel Vyoral[1], Jan Zivny[2] and Christopher D. Vulpe[4]

[1] Institute of Hematology and Blood Transfusion, Prague, Czech Republic
[2] Department of Pathophysiology, First Faculty of Medicine, Charles University, Prague, Czech Republic
[3] Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
[4] Department of Nutritional Science and Toxicology, University of California, CA, USA

After reading many 2-DE-based articles featuring lists of the differentially expressed proteins, one starts experiencing a disturbing déjà vu. The same proteins seem to predominate regardless of the experiment, tissue or species. To quantify the occurrence of individual differentially expressed proteins in 2-DE experiment reports, we compiled the identities of differentially expressed proteins identified in human, mouse, and rat tissues published in three recent volumes of Proteomics and calculated the appearance of the most predominant proteins in the dataset. The most frequently identified protein is a highly abundant glycolytic enzyme enolase 1, differentially expressed in nearly every third experiment on both human and rodent tissues. Heat-shock protein 27 (HSP27) and heat-shock protein 60 (HSP60) were differentially expressed in about 30 percent of human and rodent samples, respectively. Considering protein families as units, keratins and peroxiredoxins are the most frequently identified molecules, with at least one member of the group being differentially expressed in about 40 percent of all experiments. We suggest that the frequent identification of these proteins must be considered in the interpretation of any 2-DE studies. We consider if these commonly observed changes represent common cellular stress responses or are a reflection of the technical limitations of 2-DE.

Received: October 1, 2007
Revised: December 28, 2007
Accepted: January 11, 2008

Keywords:
2-D PAGE / Differential expression / Enolase / Meta-analysis / Most abundant

1 Introduction

Conventional 2-DE remains a fundamental tool in expression proteomics, despite its limitations. Among its most criticized weaknesses are low dynamic range and relatively low resolution. Usually only a few hundred proteins can be detected on one gel representing the most abundant soluble cytosolic proteins. A typical published 2-DE-based expression proteomics experiment features 400–1500 spots and reports between 10 and 40 identified up- or down-regulated proteins. After reading many 2-DE-based articles presenting lists of the differentially expressed proteins, one starts experiencing a disturbing sense of déjà vu. Heat shock proteins (HSP) again? Elongation factors, proteasome subunits or peroxiredoxins once more? The same proteins seem to predominate regardless of the experiment, tissue, and species. To explore this observation and to quantify the occurrence of individual differentially expressed proteins in 2-DE experiment reports, we performed a proteomic meta-analysis.

Correspondence: Dr. Jiri Petrak, Institute of Hematology and Blood Transfusion, U Nemocnice 1, 128 20 Prague, Czech Republic
E-mail: petra@uhkt.cz
Fax: +420-221-977-370

Abbreviations: GAPDH, glyceraldehydes-3-phosphate dehydrogenase; **MBP-1,** c-myc promoter-binding protein; **PDI,** protein disulfide isomerase

We compiled the identities of differentially expressed proteins identified in human, mouse, and rat tissues by 2-DE-based experiments from three recent volumes of *Proteomics* (volumes 4–6, 2004–2006) and calculated the appearance of each protein in the dataset. Added to the dataset were all experiments performed with total cellular homogenates. Experiments with body fluids, tissue cultures supernatants, and subfractionated tissues were not included. Proteins that were identified in several forms or as fragments of one polypeptide molecule in one experiment were entered into our dataset only once for the particular experiment. The identity of the proteins (often confused by incomplete or multiple protein names) was verified through provided database accession number. We did not consider whether the protein was up- or down-regulated. The resulting dataset contained approximately 4700 protein identifications presented in 169 articles encompassing 186 individual 2-D PAGE experiments. Seventy-four articles deal with cancer while 94 studies addressed other biological questions. Human cells were studied in 99 articles (108 experiments total, 38 noncancer), while the remaining 70 articles were focused on mouse or rat studies (78 experiments total, 55 noncancer). On average each 2-DE experiment reported 25 identified differentially expressed proteins. We calculated the frequency of occurrence for the most often identified proteins and protein families in the dataset and assembled the "TOP 15" charts for human and rodent tissues (Tables 1 and 2).

2 TOP 15 – Individual proteins

Results of our survey show that some proteins appear on the lists of the differentially expressed proteins identified by 2-DE MS very often, regardless of experiment type in human and rodent tissues. The TOP 15 charts for individual proteins include mostly glycolytic enzymes, heat shock, and stress proteins as well as cytoskeletal components. The human and rodent TOP 15 charts are very similar, sharing seven out of the 15 proteins. We also found that approximately 70 percent of articles reporting research with human samples identified at least two of the TOP 15 proteins as differentially expressed. Every fourth article reported at least five of the TOP 15. The most frequently identified differentially expressed protein is enolase 1 (enolase alpha). This highly abundant glycolytic enzyme has been identified as differentially expressed in about 30 percent of all 2-DE-based experiments in human and rodent tissues.

Table 1 reflects each identified protein as an individual polypeptide. However, many proteins belong to protein families of structurally very closely related molecules, often of similar, or overlapping functions such as annexins, tubulins, or peroxiredoxins. To take this fact into account we also calculated the TOP 15 charts considering protein families as individual units (Table 2).

3 TOP 15 – Protein families

When we grouped individual proteins into protein families the resulting chart changed considerably, the most often identified proteins now being keratins in human samples and peroxiredoxins in rodent tissues. Again the human and rodent charts are very similar sharing eight out of the 15 protein families. At least one member of keratin family was reported as differentially expressed in over 40 percent of all experiments performed on human tissues. Similarly, antioxidant proteins peroxiredoxins were identified in almost 40 percent of the experiments on rodent tissues. The next topmost positions belong to cytoskeletal proteins (tubulins, annexins, actins, and tropomyosins), stress and antioxidant proteins (HSP27, protein disulfide isomerases (PDIs), glutathione-S-transferases (GSTs)), and once more enolases, represented mostly by enolase 1.

4 Pseudo-groups and subunits

During our meta-analysis it became apparent that in addition to individual proteins and structurally closely related protein families there are several groups of functionally related proteins that are highly represented in the data set. Three such groups clearly stand out:

(i) Proteasome subunits (63 identifications in 108 experiments on human samples)

(ii) Heterogeneous ribonucleoprotein particle subunits-hnRNPs (53 identifications in 108 experiments on human samples)

(iii) Elongation factors (49 identifications in 108 experiments on human samples)

5 Artifacts or universal sensors?

Our simple meta-analysis of published 2-DE experiments on human, mouse, and rat tissues demonstrated that some individual proteins or protein families are strikingly overrepresented as differentially expressed, regardless of the tissue used and experiment performed. The most recurrent protein, enolase 1, was identified in every third experiment. Indeed, all of the commonly identified proteins in our meta-study are highly abundant soluble proteins. That raises the concern that their frequent identification represents a technical artifact, limitation or bias of the method. Alternatively,

Table 1. TOP 15 most often identified differentially expressed proteins

Individual proteins							
	Humans				Rodents		
	Protein name	Number of identifications	Identified in percents of experiments (%)		Protein name	Number of identifications	Identified in percents of experiments (%)
1	HSP27	34	31	1	Enolase 1	25	32
2	Enolase 1	31	29	2	HSP60	16	21
3	Triosephosphate isomerase	22	20	3	ATP synthase beta subunit	14	18
4–6	Pyruvate kinase M1/M2	21	19	4–8	Vimentin	13	17
4–6	Peroxiredoxin 1	21	19	4–8	Grp75	13	17
4–6	Peroxiredoxin 2	21	19	4–8	Apolipoprotein A1	13	17
7	Vimentin	20	19	4–8	Dihydropyrimidinase-like 2 protein	13	17
8	Annexin A4	19	18	4–8	Peroxiredoxin 6	13	17
9	HSC71	18	17	9–10	Phosphoglycerate mutase 1	12	15
10–11	Peptidyl-prolyl isomerase A	17	16	9–10	HSC71	12	15
10–11	Cytokeratin 8	17	16	11–12	Triosephosphate isomerase	10	13
12	Cathepsin D	16	15	11–12	Calreticulin	10	13
13	ATP synthase beta subunit	15	14	13–15	RhoGDI 1	9	12
14–15	Grp78/Bip	14	13	13–15	Grp78/Bip	9	12
14–15	RhoGDI 1	14	13	13–15	GAPDH	9	12

Grp75, glucose regulated protein 75 kDa; Grp78/Bip, glucose regulated protein 78 kDa; HSC71, heat shock cognate 71 kDa protein; RhoGDI 1, Rho GDP-dissociation inhibitor 1.

Table 2. TOP 15 most often identified differentially expressed proteins

Protein families							
	Humans				Rodents		
	Protein name	Number of identifications	At least one member identified in percent of experiments (%)		Protein name	Number of identifications	At least one member identified in percent of experiments (%)
1	Keratins	70	41	1	Peroxiredoxins	34	38
2	Annexins	67	40	2	Enolases	33	42
3	Peroxiredoxins	61	46	3–4	Tubulins	24	20
4	Actins	36	30	3–4	PDIs	24	26
5–6	HSP27	34	31	5	Annexins	22	26
5–6	Tropomyosins	34	23	6–7	Actins	21	23
7	GSTs	33	29	6–7	GSTs	21	19
8–10	Enolases	32	30	8–9	Tropomyosins	17	16
8–10	PDIs	32	26	8–9	Dihydropyrimidinase-like proteins	17	19
8–10	Tubulins	32	21	10	HSP60	16	21
11	Cathepsins	26	22	11–12	Carbonic anhydrases	15	18
12	TCP-1	24	22	11–12	Apolipoproteins	15	17
13–14	Triosephosphate isomerase	22	20	13–15	ATP synthase beta subunit	14	18
13–14	Pyruvate kinases	22	20	13–15	Malate dehydrogenases	14	18
15	Vimentin	20	20	13–15	14-3-3 proteins	14	14

GSTs, glutathione-S-transferases; PDIs, protein disulfide isomerases; TCP-1, chaperonin containing TCP-1 family.

these proteins could represent universal cellular sensors that respond to multiple different stimuli. However, many of these proteins, such as the conserved glycolytic enzyme, enolase 1, would not be obvious candidates for such a role. In order to determine if the genes coding the most "notorious" proteins (*ENO1*, *HSP27*, and *HSP60*) are also commonly differentially expressed at the mRNA level we performed a meta-analysis of publicly available microarray data.

6 Meta-analysis of transcriptomic data

Published mRNA expression data use a wide variety of platforms and analysis methods and are not directly comparable. We therefore had to perform a complete computational reanalysis of the original data. We separately analyzed 50 human and 50 mouse randomly selected experiments performed using Affymetrix array platform (selection criteria: CEL file available through the GEO database, at least three replicates for each individual sample available). The raw microarray data were analyzed with Bioconductor 2.1 [1] and the R project for statistical computing (version 2.6; http://www.r-project.org). The probes were annotated to Affymetrix probe set IDs using the default chip description files (cdf) "hgu133plus2" or "mouse4302," from the Bioconductor repository. The data were normalized using gcRMA [2]. We used Linear Models for Microarray Data Package, limma version 2.12 for the statistical evaluations of expression differences [3]. We considered genes to be differentially expressed if the adjusted *p*-value was <0.05 for particular comparison. We focused our attention to the transcripts of *ENO1*, *HSP27*, and *HSP60*.

In the human array experiments expression of *HSP27* and *ENO1* mRNAs was frequently altered. *HSP27* mRNA was differentially expressed in surprising 42% of all analyzed human microarray experiments and *ENO1* mRNA expression was altered in 28% of experiments on human tissues. The observed transcriptional changes for these genes correlate surprisingly well with the frequencies calculated in the presented proteomic meta-analysis. In the mouse samples, the correlation is not apparent, expression of the mRNAs coding the most often differentially expressed proteins enolase 1 and HSP60 were altered only in 9 and 7%, respectively. We are aware that our limited meta-analysis of microarray experiments may not be sufficiently representative but does suggest that the transcript levels of the two most often identified differentially expressed human proteins are also commonly altered in microarray experiments.

7 Is enolase 1 a universal sensor, regulator, or a stress chaperone?

A growing body of evidence makes clear, that enolase 1 is a multifunctional protein. Enolase 1 (P06733, P17182) (also known as alpha-enolase or Plasminogen-binding protein) is a highly abundant 48 kDa cytosolic enzyme which catalyzes the reversible dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate as a part of the glycolytic and gluconeogenesis pathways. Aside from its primary cytosolic role, enolase 1 can also be found on the cell surface serving as a plasminogen receptor [4] and in the nucleus, as an alternatively translated 37 kDa tumor suppressor c-myc promoter-binding protein (MBP-1) [5, 6]. Tau-crystallin, the principal component of reptile and some avian eye lens is also coded by *eno1* gene [7]. In addition, enolase 1 was reported to be a hypoxic stress protein and an HSP [8, 9]. This molecule has also been considered to be a diagnostic tumor marker [10], and autoantibodies against enolase 1 have been found in numerous autoimmune diseases [11]. A recent work demonstrates the existence of a regulatory circuit between c-myc, MBP-1, and enolase 1 that connects cellular energy metabolism and proliferation [12]. This regulatory data and our analysis showing differential expression of enolase 1 in nearly every third 2-DE experiment as well as the responsiveness of *ENO1* mRNA in human tissues together suggests, that enolase1/MBP-1 could play an important sensor or regulator role in multiple stress situations.

8 How good a housekeeper is GAPDH?

Enolase 1 was accompanied by two or three other highly abundant glycolytic enzymes, triosephosphate isomerase, and pyruvate kinase (in human) and triosephosphate isomerase, phosphoglycerate mutase, and glyceraldehydes-3-phosphate dehydrogenase (GAPDH) (rodents) in our TOP 15 charts. Appearance of GAPDH among the proteins highly responsive to various stimuli (15th in rodent, 17th in human TOP charts) is somehow disturbing. GAPDH protein is generally viewed and used as a housekeeping protein and therefore a reliable internal standard. In our microarray meta-study *GAPDH* mRNA was differentially expressed in 18% of human and 14% of mouse experiments. Therefore the general responsiveness of GAPDH demonstrated by our meta-analyses suggests that we should re-evaluate its image of a housekeeper gene and reconsider using GAPDH as an internal standard in expression studies. Similarly, our protein charts also advocate caution in using beta-actin or tubulin as internal standards.

9 Heat-shock proteins

Along with enolase 1, two HSPs occupied the first two positions in the TOP 15 charts. HSP27 was the most responsive protein molecule in humans whereas HSP60 occupied the second position of the chart in rodents. Presence of the HSPs and other stress proteins among the TOP 15 can be explained by the fact that a significant portion of the experiments studied tissues that were stressed under non-physiological conditions (tumor tissue, tumor derived cell lines, diseased tissue, or experimentally induced toxic states).

10 Keratins – differentially expressed molecules or investigator's signatures?

Interestingly, keratins – the most notorious protein family identified in human analyses (70 keratins identified *per* 108 experiments) were relatively under-represented in rodent experiments (11 identified keratin molecules *per* 78 experiments). Since numbers of experiments addressing epithelial

tissues in human and mouse were comparable and there is also no reason to expect that mouse keratins are less abundant than human keratins, we believe that at least some of the "differentially expressed" keratins identified in human tissue samples are in fact contaminants originating from the bodies of investigators involved. Our hypothesis is supported by the fact that nearly half of the identified (30 out of 70) human keratin molecules (cytokeratins 1, 2, 6, 9, 10, 13, 16, 17, and 19) are expressed in human skin and mucosa and can thus be potentially attributed to crosscontamination.

11 Peroxiredoxins – is more attention warranted?

Peroxiredoxins are a family of ubiquitous thiol-specific antioxidant enzymes with a molecular weight of approximately 25 kDa. Peroxiredoxins detoxify hydrogen peroxide, peroxynitrite, and organic hydroperoxides. Although found mainly in cytoplasm, peroxiredoxins can be also located in mitochondria and peroxisomes and associated with nuclei and membranes [13]. In addition to peroxide detoxification, peroxiredoxins also control cytokine-induced peroxide levels that can modulate signal transduction in mammals. Peroxiredoxins are thus involved not only in oxidative stress but also in cellular proliferation, differentiation, and apoptosis and have also been showed to function as molecular chaperones during oxidative stress [13, 14] and tumor suppressors [15]. The abundance and multifunctionality of peroxiredoxins has aroused the interest of researchers in last decade and we believe that our survey, demonstrating their responsiveness to the plethora of stimuli, should augment this enthusiasm.

12 Over-interpreting of 2-DE results

As seen from the results of our analysis a typical published 2-DE experiment demonstrates differential expression of several cytoskeletal and stress proteins, proteasome subunits, glycolytic enzymes, elongation factors, and heterogeneous ribonucleoprotein particle subunits or glutathione transferases. These proteins are (in some cases) accompanied by less abundant regulatory proteins. This result suggests that we should use extreme caution in the interpretation of differential expression of the most frequently identified proteins. We suggest that our TOP 15 charts could serve a help for other researchers to prevent unjustified overinterpretation of 2-DE-based studies.

13 Future proteomic meta-analyses and systems biology

We believe that our study demonstrates that meta-analyses of proteomic data can provide invaluable information pertinent to various biological processes (or methods involved). Targeted, statistically robust proteomic meta-analyses could provide invaluable information for biomedical research. For instance a global analysis of "cancer specific" expression patterns based on a large dataset that also consider whether the protein was up- or down-regulated, could provide a brand new tool for dissecting the complex processes of tumorigenesis. Similar meta-analytical approach was recently used to compile and prioritize a database of candidate tumor biomarkers [16]. We also believe that such crosssectional views of proteomic data have the potential to discover limitations or weaknesses of 2-DE as a method and consequently help to improve and further develop this technique.

This study was supported by grants from the Czech Science Foundation (GACR) no. 204/07/0830, from Ministry of Health of the Czech Republic MZCR/UHKT no. 023736, IGA MZCR no. NR8930-4, from Ministry of Education, Youth and Sports of the Czech Republic project nos. LC06044 and 0021620806. C. V. acknowledges support of NIH grant R01-DK056376. The authors wish to thank Mrtva Ryba.

The authors have declared no conflict of interest.

14 References

[1] Gentleman, R., Carey, V. J., Bates, D. M., Bolstad, B. *et al.*, Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* 2004, *5*, R80.

[2] Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D. *et al.*, Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003, *4*, 249–264.

[3] Smyth, G., in: Gentleman, R., Carey, V. J., Huber, W., Irrizary, R. A. (Eds.), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer, New York 2005, pp. 397–420.

[4] Pancholi, V., Fischetti, V. A., Alpha-enolase, a novel strong plasmin(ogen) binding protein on the surface of pathogenic streptococci. *J. Biol. Chem.* 1998, *273*, 14503–14515.

[5] Feo, S., Arcuri, D., Piddini, E., Passantino, R. *et al.*, ENO1 gene product binds to the c-myc promoter and acts as a transcriptional repressor: Relationship with Myc promoter-binding protein 1 (MBP-1). *FEBS Lett.* 2000, *473*, 47–52.

[6] Subramanian, A., Miller, D. M., Structural analysis of alpha-enolase. Mapping the functional domains involved in down-regulation of the c-myc protooncogene. *J. Biol. Chem.* 2000, *275*, 5958–5965.

[7] Wistow, G. J., Lietman, T., Williams, L. A., Stapel, S. O. *et al.*, Tau-crystallin/alpha-enolase: One gene encodes both an enzyme and a lens structural protein. *J. Cell. Biol.* 1988, *107*, 2729–2736.

[8] Aaronson, R. M., Graven, K. K., Tucci, M., McDonald, R. J. *et al.*, Non-neuronal enolase is an endothelial hypoxic stress protein. *J. Biol. Chem.* 1995, *270*, 27752–27757.

[9] Iida, H.,Yahara, I., Yeast heat-shock protein of M_r 48 000 is an isoprotein of enolase. *Nature* 1985, *315*, 688–690.

[10] He, P., Naka, T., Serada, S., Fujimoto, M. *et al.*, Proteomics-based identification of alpha-enolase as a tumor antigen in non-small lung cancer. *Cancer Sci.* 2007, *98*, 1234–1240.

[11] Gitlits, V. M., Toh, B. H., Sentry, J. W., Disease association, origin, and clinical relevance of autoantibodies to the glycolytic enzyme enolase. *J. Investig. Med.* 2001, *491*, 38–45.

[12] Sedoris, K. C., Thomas, S. D., Miller, D. M., C-myc promoter binding protein regulates the cellular response to an altered glucose concentration. *Biochemistry* 2007, *46*, 8659–8668.

[13] Wood, Z. A., Schroder, E., Robin Harris, J., Poole, L. B., Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem. Sci.* 2003, *28*, 32–40.

[14] Jang, H. H., Lee, K. O., Chi, Y. H., Jung, B. G. *et al.*, Two enzymes in one; two yeast peroxiredoxins display oxidative stress-dependent switching from a peroxidase to a molecular chaperone function. *Cell* 2004, *117*, 625–635.

[15] Neumann, C. A., Krause, D. S., Carman, C. V., Das, S. *et al.*, Essential role for the peroxiredoxin Prdx1 in erythrocyte antioxidant defence and tumour suppression. *Nature* 2003, *424*, 561–565.

[16] Polanski, M., Anderson, N. L., A list of candidate cancer biomarkers for targeted proteomics. *Biomark. Insights* 2006, *2*, 1–48.

