research articles Journal of proteome research

Comparison of Statistical Approaches for the Analysis of Proteome Expression Data of Differentiating Neural Stem Cells

Martin H. Maurer,*,† Robert E. Feldmann, Jr.,† Jens O. Brömme,† and Armin Kalenka‡

Department of Physiology and Pathophysiology, University of Heidelberg, Heidelberg, Germany; Department of Anesthesiology and Critical Care Medicine, Faculty of Clinical Medicine Mannheim, Mannheim, Germany

Received August 30, 2004

Comparative proteomic studies often use statistical tests included in the software for the analysis of digitized images of two-dimensional electrophoresis gels. As these programs include only limited capabilities for statistical analysis, many studies do not further describe their statistical approach. To find potential differences produced by different data processing, we compared the results of (1) Student's t-test using a spreadsheet program, (2) the intrinsic algorithms implemented in the Phoretix 2D gel analysis software, and (3) the SAM algorithm originally developed for microarray analysis. We applied the algorithms to proteome data of undifferentiated neural stem cells versus in vitro differentiated neural stem cells. We found (1) 367 spots differentially expressed using Student's t-test, (2) 203 spots using the algorithms in Phoretix 2D, and (3) 119 spots using the algorithms in SAM, respectively, with an overlap of 42 spots detected by all three algorithms. Applying different statistical approaches on the same dataset resulted in divergent set of protein spots labeled as statistically "significant". Currently, there is no agreement on statistical data processing of 2DE datasets, but the statistical tests applied in 2DE studies should be documented. Tools for the statistical analysis of proteome data should be implemented and documented in the existing 2DE software.

Keywords: two-dimensional gel electrophoresis • proteomics • Phoretix • t-test

Introduction

Two-dimensional gel electrophoresis (2DE) in combination with mass spectrometry is the standard approach in proteome research,[1] although recent years have brought up electrophoresis-free methodology based solely on mass spectrometry or protein chips.[2,3]

Characteristically, 2DE resolves several hundred, or even thousands, of protein spots per gel, depending on the dynamic range of the protein staining procedure applied. Therefore, sophisticated software had to be developed to detect spots in digitized gel images. Early programs such as Flicker[4] (http://www-lmmb.ncifcrf.gov/flicker/) were solely used for spot detection and comparison of spot images, nowadays software also allows comparison of qualitative and quantitative parameters and variables in the experimental data. Third generation software development met high-throughput prerequisites for fully automated processing control not only for image analysis but also for automated spot picking, trypsinization, and spotting onto MALDI targets (so-called "proteomic workstations"). Examples of proteomic workstations may be the Ettan Spot Handling Workstation (Amersham Biosciences, Uppsala, Sweden), or Progenesis Workstation (Nonlinear Dynamics, Newcastle-upon-Tyne, U.K.).

Unfortunately, nearly all of these programs and software packages lack the description of the statistical algorithms applied for the comparison of spot parameters, or statistical functions are not implemented at all. Whereas nearly all programs provide result previews based on the display of relative expression levels as the ratio of mean values, the software uses the ratio along with a few undocumented and weak statistical functions to determine expression differences.

In the present study, we compare three different statistical approaches to analyze proteomic expression data between two experimental groups of hippocampal neural stem cells, including (i) the software-implemented algorithm for data analysis, (ii) a statistical test performed with exported data independent from the original 2DE software, and (iii) an algorithm for large scale expression analysis originating from microarray gene expression data processing.

Material and Methods

Cells and Two-Dimensional Gel Electrophoresis. Hippocampal neural stem cells were grown as neurosphere cultures in the presence of FGF-2 and EGF as described previously.[5] Cells were differentiated in vitro for 48 h as reported[6] by adding 2% fetal bovine serum and removing growth factors. 2DE was performed as described elsewhere in detail.[5,6] Briefly, cells were

* To whom correspondence should be addressed: Dr. Martin H. Maurer, Department of Physiology and Pathophysiology, Im Neuenheimer Feld 326, 69120 Heidelberg, Germany. Phone: +49-6221-544075, Fax: +49-6221-544561, E-mail: maurer@uni-hd.de.

† Department of Physiology and Pathophysiology, University of Heidelberg.

‡ Department of Anesthesiology and Critical Care Medicine, Faculty of Clinical Medicine Mannheim.

10.1021/pr049841l CCC: $30.25
Published on Web 11/23/2004

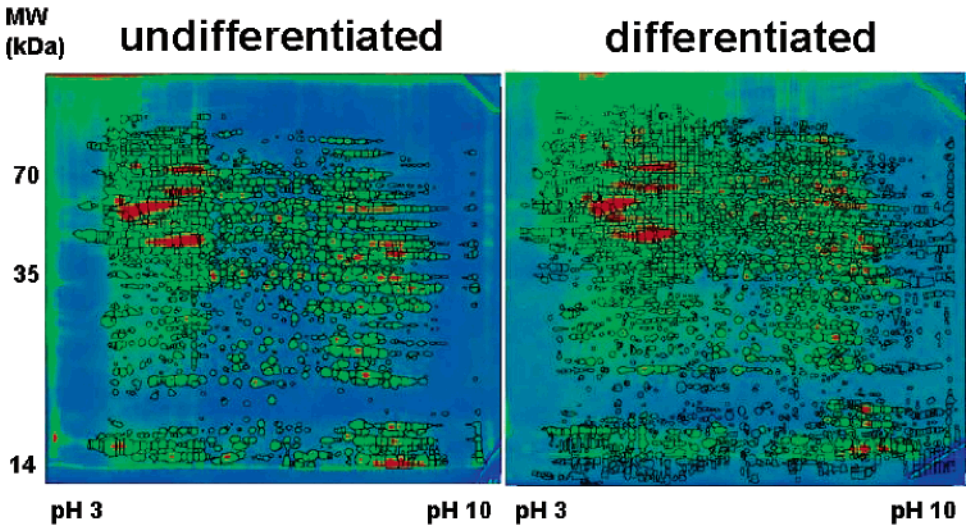

Figure 1. Original gel images of the dataset using the Progenesis Workstation (PerkinElmer Life & Analytical Sciences, Cambridge, U.K.) from undifferentiated neurospheres (left part) and in vitro differentiated neurospheres (right part). Spots detected by the software were marked by black outlines. Calculated and normalized spot volumes of 5 gels from the undifferentiated and 6 gels from the differentiated group were used for statistical calculations.

solubilized in 40 mM Tris-HCl, 7 M urea, 4% (w/v) 3-[(3-cholamidopropyl)dimethylamino]-1-propanesulfonate (CHAPS), 10 mM 1,4-dithiothreitol, and 1 mM EDTA and separated in the first dimension according to their isoelectric point using immobilized pH gradient gels (Immobiline DryStrips pH 3−10 NL, 18 cm, Amersham Biosciences, Uppsala, Sweden) at 100 310 Vh. In the second dimension, proteins were separated according to their molecular weight using 12.5% polyacrylamide electrophoresis. Gels were silver-stained and images were digitized using a desktop scanner.

Image Processing, Dataset, and Statistical Analysis. We used the Progenesis Workstation (PerkinElmer Life & Analytical Sciences, Cambridge, UK) for spot detection, gel matching, background correction (mode of nonspot, vector size 100 pixel), and normalized these data to the sum of the total spot volume as described.[6] The dataset was obtained from 5 gels for the undifferentiated cells and 6 gels for the differentiated cells. The values used for statistical analysis were derived from single gels. The dataset contained 2472 unique spots, of which 2066 occurred in the undifferentiated group vs 1697 spots in the differentiated group. We found 1291 spot matches in both groups (Figure 1). Data were exported to a spreadsheet program (Microsoft Excel 2002, Microsoft, Redmond, WA). We compared spot volumes defined as the integral of the optical density over the spot area between the control and the experimental groups using (i) Student's t-test,[7] (ii) the algorithm for comparing expression data implemented in the Phoretix 2D Database v6.0 software (Nonlinear Dynamics, Newcastle-upon-Tyne, UK; http://www.nonlinear.com), and (iii) an algorithm originally developed for microarray analysis (http://www.stat.stanford.edu/~tibs/SAM).[8]

Statistical Analysis. (i) For Student's t-test, we compared the means and standard deviations of the individual spots' optical densities from 5 gels of the undifferentiated cells and 6 gels from the differentiated cells by applying the standard algorithm.[7] We set the level of significance to $P < 0.05$ and compared data by using a two-sided t-test for unpaired samples. (ii) Functions implemented in the Phoretix 2D Database software were used to compare the dataset by a two-sided t-test with the level of significance set to $P < 0.05$. (3)

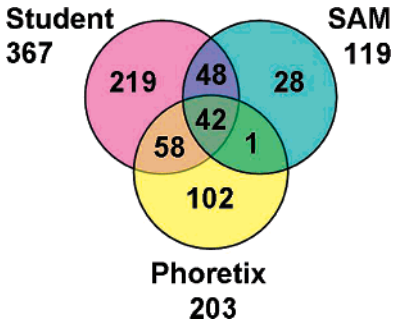

Figure 2. Venn diagram showing the numbers of differentially expressed protein spots using Student's t-test, Phoretix 2D database v6.0, and SAM-Significance Analysis of Microarrays. Numbers outside the circles summarize totals in each groups. We found 42 protein spots in all three groups (center).

For comparison of data analysis tools originally developed for microarrays, we used SAM v1.21 (Statistical Analysis of Microarrays).[8] The SAM algorithm identifies statistical significances by assimilating sets of specific t-tests. Then, it assigns a score to each dataset item (here: volume of a protein spot, calculated as the integrated optical density over the staining area) on the basis of its change in expression relative to the standard deviation of repeated measurements.[8] Specifically, we applied the following settings: 10-nearest neighbor imputer, two class unpaired data, unlogged dataset, 100 permutations, delta = 0.21577. The program requires setting a value for the fold change, which we set to 2. To identify a nonlinear bias of the dataset, we visualized the data by a M−A plot which is a graphical way-to-see ratios and spot intensity at the same time by plotting $\log_2 (A/B)$ vs 0.5 ($\log_2 A + \log_2 B$), where A and B are the normalized spot volumes in the undifferentiated and the differentiated cells, respectively.[9]

Results and Discussion

Comparing the different analytical tools based on statistical test systems, we found in the dataset of 2472 spots, (1) 367 protein spots (15%) differentially expressed using Student's t-test, (2) 203 protein spots (8%) differentially expressed using

Table 1. Applying Three Statistical Tests to the Same Dataset Revealed an Overlap of 42 Protein Spots Identified by All Three Approaches[a]

fold change undiff. vs diff. (means ± standard deviation)	protein annotation	Swiss-Prot accession no.	theoretical p*I*	theoretical MW (Da)	method of identification
12.56 ± 8.02	ATP synthase α chain, mitochondrial precursor (fragment)	P15999	9.22	58904	GM
4.59 ± 0.89	peroxiredoxin 6	O35244	5.65	24687	MALDI-TOF
4.36 ± 1.54	glutathione S-transferase P (GST 7−7) (chain 7) (GST class-pi)	P04906	7.3	23652	GM
3.65 ± 1.76	peptidylprolyl isomerase A (cyclophilin A)	P10111	8.37	18091	GM, MALDI-TOF
3.60 ± 0.52	Sep2 (fragment)	Q9ESF7	5.74	49050	GM
3.13 ± 2.25	tubulin alpha	P02551	4.94	50894	MALDI-TOF
2.40 ± 0.61	septin 2	*P*42208	6.1	41737	MALDI-TOF
2.23 ± 0.07	aldolase A, fructose-bisphosphate	P05065	8.4	39783	MALDI-TOF
2.18 ± 0.89	aldehyde reductase 1 (low Km aldose reductase)	P07943	6.28	36230	MALDI-TOF
2.12 ± 0.29	glyceraldehyde-3-phosphate dehydrogenase	Q9QWU4	8.14	36090	MALDI-TOF
2.01 ± 0.34	triosephosphate isomerase 1	P48500	6.51	27417	MALDI-TOF
0.39 ± 0.29	eukaryotic translation elongation factor 2	P05197	6.42	96192	MALDI-TOF
0.33 ± 0.28	mitochondrial aconitase (nuclear aco2 gene)	Q9ER34	7.87	86162	GM
0.29 ± 0.18	eukaryotic translation elongation factor 2	P05197	6.42	96192	MALDI-TOF

[a] The 14 spots identified by matrix-assisted laser desorption/ionization−time-of-flight (MALDI-TOF) mass spectrometry and gel matching (GM). Twenty-eight spots could not be identified. Their relative expression levels ranged from 0.13 to 13.42 (5 down-regulated, 23 up-regulated spots). The first column shows mean values ± standard deviation of fold changes in the undifferentiated vs the differentiated group as exported from the Progenesis results table.

the algorithms of Phoretix 2D Database v6.0, and (3) 119 protein spots (5%) differentially expressed using SAM-Significance Analysis of Microarrays (Figure 2). Of the spots assigned as "differentially expressed", we found an overlap of 42 protein spots with statistically significant differential expression in all three test approaches (Table 1). As the same raw data set was used for all three test systems, the differences in spots detected with a "significant" differential expression must solely be attributed to the underlying statistical approach. Of the 42 spots found in all 3 statistical approaches, we could identify 14 proteins (Table 1) by mass spectrometry and gel matching. Two-thirds of the protein spots showed low abundant expression, or no database entry was found to identify the protein unanimously.

Using an analysis method based on t-tests requires a normal distribution of the measured values. This cannot be concluded from only few data points, as in most 2DE experiments where groups do not comprehend more than 10−20 data points (e.g., individuals, gels). Thus, most tests for normal distribution find that the normal distribution of the experimental values cannot be rejected, which does not imply that a normal distribution is the true distribution of the data. Despite these problems, most proteomic studies assume a normal distribution of values, or rely on algorithms implemented in the specific software.

We performed a meta-analysis of 2DE publications investigating differential brain protein expression including 18 randomly selected studies published from 1984 to 2004 (data not shown). We found that only 9 studies (50%) sufficiently described the image analysis software used, the number of experimental groups compared, the number of gels included in each experimental group, and the statistical test applied. As a matter of course, these information should be included in every proteomic study describing (semi-)quantitative comparisons between an experimental and control group, or between several experimental groups. Reportings such as "protein spots were quantified and compared using the xyz software" as the only description of data processing do not contain sufficient information to reproduce experimental data in a reviewer's point of view. More detailed descriptions of the statistical methods used are not only desirable, but essential.

Other high-throughput analysis technologies such as microarray screening have considered statistical tests not only for data acquisition, but also for the analysis of large-scale experimental data. Tuscher et al. have provided a method for statistical evaluation of microarray experiments.[8] The software is free for academic users and can be downloaded at http://www-stat.stanford.edu/~tibs/SAM. The program can compare gene expression data using defined parameters for significance and thresholds. It can also be used for the analysis of 2DE gels, but the dataset has to be converted into the special input format. This method was the "strictest" one applied in the current study, resulting in the lowest number of spots retrieved as "significant" (Figure 3A). To exclude a nonlinear bias of the dataset, we visualized the results by a M−A plot (Figure 3B). This plot helps to explain the untypical distribution in Figure 3A of the negative values (decreased spots), which would suggest a suboptimal normalization procedure. As the values are evenly distributed around the *x*-axis, and no substantial outliers can be seen, we conclude from Figure 3A that the SAM algorithm may not be optimal for the analysis of proteome data. An explanation for this finding could be that proteome expression data with only several hundreds of protein spots are much smaller than microarray data sets with at least several thousand sequences. We used the SAM algorithm to test if it could also be applied for smaller datasets such as from 2DE studies, since there are no statistical tools tailored for proteome expression data. In conclusion, the SAM algorithm seems not very ap-

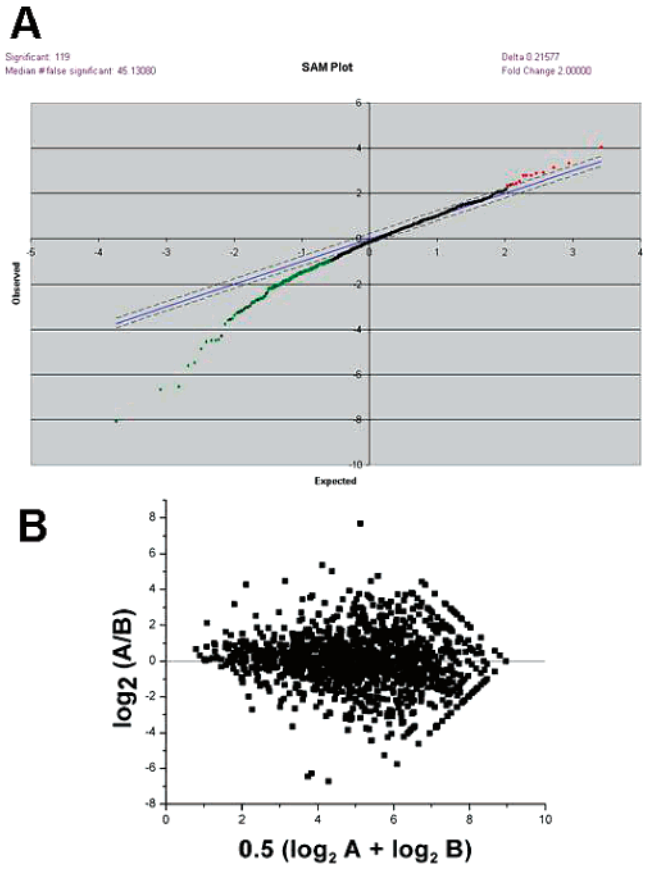

Figure 3. SAM-Significance Analysis of Microarrays plot shows expected vs observed changes in expression levels. Spots on the blue line are unchanged between the two experimental groups, whereas spots beyond the dashed lines indicate differential expression. Red spots are increased in their concentration, green spots are decreased in their concentration, black spots are not statistically different. Plotted with SAM v1.21 (ref 8).

plicable to proteome expression data, as is does not produce very stringent results due to small datasets.

Although the scientific literature for microarray analysis comprehends excellent work on statistical methods for high-troughput data analysis,[10−12] there are presently no comparable solutions for proteomics.[13] At the moment, statistical tools for platform- and software-independent statistical analysis of 2DE datasets are being developed, e.g., DeltaStat (http://deltastat.org),[14] an open source software based on the R language for statistical computing (http://www.r-project.org) and the MySQL database (http://www.mysql.com) which allow easy incorporation of additional statistical tests.

Currently, 2DE is still a major technique in quantitative proteomic comparisons,[1] but other techniques mainly based on 2D chromatographic protocols emerge rapidly.[3] Therefore, quantitative analysis of 2DE needs to establish reliable statistical tools in order to maintain and improve its value for proteomic analysis. Several studies compared the commercially available 2DE programs Melanie, PDQuest, Progenesis, and Z3 using a standardized dataset with regard to spot detection, gel matching, and spot quantitation.[15,16] The findings indicate that spot quantitation is one of the most delicate and important steps in 2DE gel analysis for finding differentially expressed proteins. Taking into account that all statistical tests applied in this study are based on t-tests, or a modificaton thereof, we surprisingly found divergent results. For example, statistical software specially tailored for large-scale microarray analysis such as the SAM algorithm may not be well-suited for proteomic applications, showing the need for developing such tools for proteome studies. On the other hand, applying statistical algorithms built in the 2DE software which are not publicly documented, may lead to nonmeaningful results due to algorithm-specific settings. Thus, for obtaining accurate results, proper statistical tests have to be developed, applied and described.

Conclusions

Our results show that applying different statistical approaches on the same dataset can result in divergent sets of protein spots labeled as statistically "significant". Currently, there is no agreement on statistical data processing of 2DE datasets, but the statistical tests applied in 2DE studies should be documented. In this context, the development of open source software could have a great impact on the field of 2D gel analysis, as open source software is collaborative and peer-reviewed by nature. This allows the users to study the source code and to create a documentation for the statistical tests used.

Acknowledgment. This work was supported in part by a grant of the German Ministry of Education and Research (BMBF) in the National Genome Research Network NGFN-2 (to M.H.M.).

References

(1) Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* **2002**, *2*, 3−10.
(2) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198−207.
(3) Righetti, P. G.; Campostrini, N.; Pascali, J.; Hamdan, M.; Astner, H. Quantitative proteomics: a review of different methodologies. *Eur. J. Mass Spectrom.* (*Chichester, Eng*) **2004**, *10*, 335−348.
(4) Lemkin, P. F.; Myrick, J. M.; Lakshmanan, Y.; Shue, M. J.; Patrick, J. L.; Hornbeck, P. V.; Thornwal, G. C.; Partin, A. W. Exploratory data analysis groupware for qualitative and quantitative electrophoretic gel analysis over the Internet-WebGel. *Electrophoresis* **1999**, *20*, 3492−3507.
(5) Maurer, M. H.; Feldmann, R. E., Jr.; Fütterer, C. D.; Kuschinsky, W. The proteome of neural stem cells from adult rat hippocampus. *Proteome Sci.* **2003**, *1*, 4.
(6) Maurer, M. H.; Feldmann, R. E., Jr.; Fütterer, C. D.; Butlin, J.; Kuschinsky, W. Comprehensive proteome expression profiling of undifferentiated vs differentiated neural stem cells from adult rat hippocampus. *Neurochem. Res.* **2004**, *29*, 1129−1144.
(7) Gossett, W. S. The probable error of a mean. *Biometrika* **1908**, *6*, 1−25.
(8) Tusher, V. G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98*, 5116−5121.
(9) Dudoit, S.; Yang, Y. H.; Callow, M. J.; Speed, T. P. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* **2002**, *12*, 111−139.
(10) Brazma, A.; Vilo, J. Gene expression data analysis. *FEBS Lett.* **2000**, *480*, 17−24.
(11) Smyth, G. K.; Yang, Y. H.; Speed, T. Statistical issues in cDNA microarray data analysis. *Methods Mol. Biol.* **2003**, *224*, 111−136.
(12) Storey, J. D.; Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9440−9445.
(13) Boguski, M. S.; McIntosh, M. W. Biomedical informatics for proteomics. *Nature* **2003**, *422*, 233−237.

(14) Dunsmore, J. DeltaStat for 2D gel Analysis v0.4. **2004**. http://deltastat.org.
(15) Raman, B.; Cheung, A.; Marten, M. R. Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie. *Electrophoresis* **2002**, *23*, 2194−2202.
(16) Rosengren, A. T.; Salmi, J. M.; Aittokallio, T.; Westerholm, J.; Lahesmaa, R.; Nyman, T. A.; Nevalainen, O. S. Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels. *Proteomics* **2003**, *3*, 1936−1946.

PR049841L