Identifying Differentially Expressed Proteins in 2-D DIGE Experiments

Yan Ma* E. James Harner
Department of Statistics
West Virginia University, Morgantown, WV

Abstract

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is the most common technique for comparing protein abundance in different samples. Traditionally, each sample is separated on a single gel, which makes it difficult to detect differentially expressed proteins due to gel-to-gel variability. CyDye difference gel electrophoresis (DIGE) Fluor dyes allow the co-migration of more than one sample per gel. Thus, an internal standard can be included along with a control and treatment in the 3-sample per gel case. This allows efficient experimental designs to be developed, which essentially eliminate the gel-to-gel variability. Specialized software, e.g., Amersham's DeCyder and Nonlinear's Cross-Stain Analysis (CSA), is required to analyze the resulting stains. The ultimate goal of these analysis systems is to find differentially expressed proteins, which can then be selected (by a spot picker) and sent to a MALDI-MS for identification and analysis. Unfortunately, these software systems are proprietary and the methods of selecting differentially expressed proteins, e.g., using multiple testing, are unknown. However, both DeCyder and CSA allow the image-analysis data to be exported using XML and Oracle, respectively. The analyst can then intervene to do their own analysis using the import facilities of R. This allows the statistician to use state-of-the-art procedures, e.g., random forests and modern multiple testing algorithms, for identifying differentially expressed proteins.

1 Introduction

Proteomics is the science of performing analyses on proteomes under differing experimental conditions (e.g., cancerous vs. normal tissues). Two-Dimensional Poly-Acrylamide Gel Electrophoresis (2-D PAGE), coupled with mass spectrometry, has had a profound impact on the development of proteomics, and these technologies continue to be the foundation of proteomic studies.

2-D PAGE is the most common technique for isolating hundreds to thousands of proteins with high resolution (O'Farrell, 1975). The main objective of 2-D PAGE is to do protein profiling so that the protein expression levels induced by different treatments or conditions can be qualitatively and quantitatively compared. The appearance or disappearance of specific protein spots together with the degree of spot intensity provides information about protein expression (Lau *et al.*, 2003).

After preparation, each sample is run on a single 2-D gel, which separates proteins according to their isoelectric points in one dimension (using an immobilized pH

*Contact: yma@stat.wvu.edu

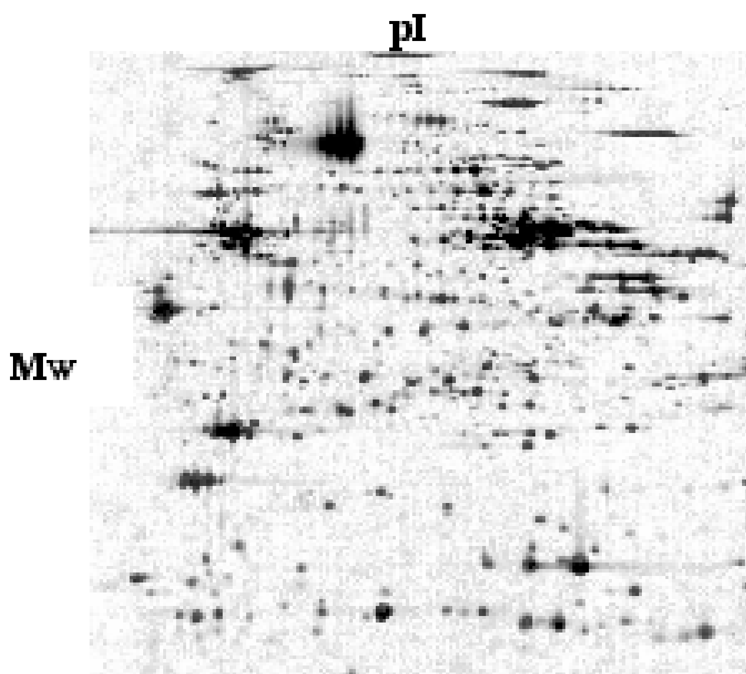

Figure 1: 2-D Gel Electrophoresis: proteins are separated along two dimensions—isoelectric point (pI) and molecular weight (Mw).

gradient) and by their molecular weights in the second dimension (see Figure 1). The separated proteins are stained prior to doing an image analysis.

Image analysis is done using software to perform spot detection and spot quantification. Once gels are imaged, a reference gel is selected (or imported) in order to align the corresponding protein spots across all the sample gels. Due to gel and treatment differences, gel matching is a complex and error-prone process. Also, within-treatment biological gel variation is still present even after matching (Roy *et al.*, 2003), and thus replication within treatments is necessary. Many factors cause gel-to-gel variation: differences in sample preparation; discrepancies in experimental conditions and electrophoretic conditions; differential deformations of gel materials, etc.

The classical one-sample-per-gel method is illustrated in Figure 2. This methodology is susceptible to the pitfalls of gel variability and distortion, which can mask differential treatment protein expression levels. For example, if the abundance of one particular protein appears to differ between cancerous and normal tissues, it may be hard to confirm that this is due to an induced biological change if the inter-gel variability is large. The recent introduction of two-DImensional Difference Gel Electrophoresis (2-D DIGE) technology removes certain types of gel-to-gel variability, which allows investigators to estimate differences in mean protein abundances more precisely. Differences in protein abundance among samples of less than 10% can often be detected using DIGE (DeCyder Differential Analysis User Manual,

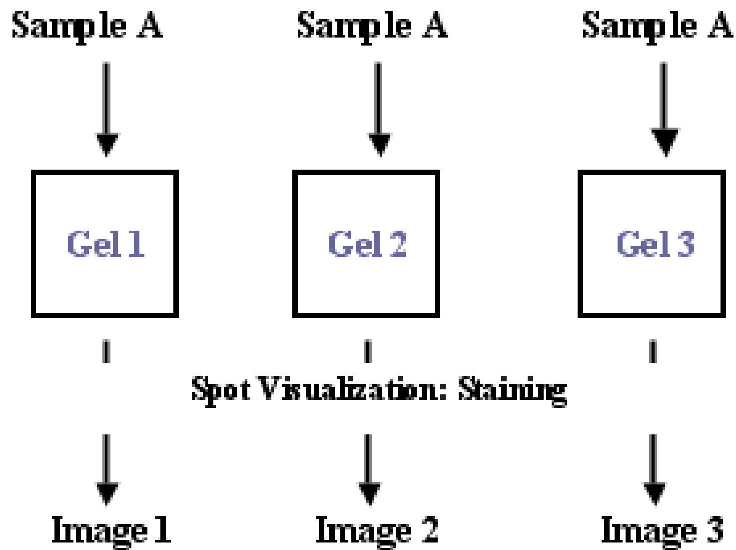

Figure 2: 2-D PAGE: each replicate of treatment A is run on a single gel resulting in 3 images.

Download at:
http://www1.amershambiosciences.com/aptrix/upp00919.nsf/Content/
Proteomics+DIGE%5CProteomics+DIGE+Protocols)

This paper is organized as follows: basic 2-D DIGE technology is introduced in Section 2; experimental designs under classical 2-D PAGE and 2-D DIGE systems are contrasted in Section 3; problems associated with proprietary software packages are discussed in Section 4; statistical and data mining techniques for identifying differentially expressed proteins are summarized in Section 5.

2 2-D DIGE Technology

Traditional 2-D PAGE technology requires each sample to be run on a single gel. As discussed previously, gel-to-gel variability can mask treatment effects, i.e., experimentally induced biological changes can be missed. With the introduction of 2-D DIGE technology, gel variation is removed as a blocking factor and the resulting comparisons are more precise.

DIGE technology was developed by Ünlü *et al.* (1997) and marketed by Amersham Biosciences. It is a method for pre-labelling protein samples prior to 2-D electrophoresis (Ettan DIGE User Manual). DIGE technology has significant benefits over conventional 2-D electrophoresis by providing a very sensitive and accurate method for measuring differences in protein expression.

DIGE technology utilizes up to three labels (dyes) that are matched in size and charge. The labels are fluorescent and spectrally distinct, and they have bright and intense colors with narrow excitation and emission bands. These properties

enable the multiplexing of up to three treatment samples (or replicates of the same treatment) on the same gel, thus allowing these samples to have identical first and second-dimension running conditions. Since treatment comparisons can be made within a gel, the gel effect is statistically removed as a blocking factor.

Multiplexing also allows the inclusion of a common mixed sample internal standard to be present in each gel, i.e., one of the three dyes can be allocated to the standard. The mixed sample internal standard has equal amounts of all the biological samples involved in the experiment and thus it contains proteins from all the biological samples. One of the internal standards is chosen as a reference and the other internal standards (one from each gel) are matched to this reference standard, i.e., protein spots are matched. Due to the identical migration paths, the treated sample spots are automatically aligned with the internal standard spots within a gel. As a result, a spot on the reference gel can be mapped to all samples once matching is done. The data is recorded as a ratio of the abundance level for each treatment to the internal standard within each gel (DeCyder Differential Analysis User Manual). In effect, the gel effect is removed from the treatment comparisons error term.

A simplified representation of 2-D DIGE technology is shown in Figure 3. Three different dyes are utilized to label different protein samples prior to 2-D electrophoresis; generally, one of them is used to label the mixed-sample internal standard. The dyes are charge-matched and have the same molecular mass. The samples are then imaged separately and the gel images are superimposed to identify differences (Lau *et al.*, 2003). The comparison of the treatment samples with the mixed-sample internal standard within each gel enables protein abundance in the diseased and normal samples (for example) to be compared without interference from gel variation (Friedman *et al.*, 2004).

DIGE technology is preferred over the 2-D PAGE because:

- the gel-to-gel variation is reduced;
- gel matching is faster;
- the number of gels required for a specified level of precision is reduced;
- experiments with large sample sizes are feasible;
- accurate quantitation is possible.

Amersham Bioscience is the pioneer of the 2-D DIGE technology and they sell hardware along with their DeCyder software. Amersham's Ettan DIGE system (see Figure 4), uses spectrally resolvable CyDye fluors (Cy2, Cy3 and Cy5), which allows the simultaneous separation of up to three samples on a single gel. After 2-D gel electrophoresis and scanning on a Typhoon Variable Mode Imager, DeCyder Differential Analysis (DIA) software analyzes the resulting protein spots. Nonlinear's Phoretix 2-D Evolution quickly followed suit with their CSA software.

3 Design of Experiments in 2-D DIGE

Results from 2-D gel studies may be inconclusive due to excessive variability or confounding if proper treatment and experimental design principles are not followed. A single factor, e.g., disease state, is most common for 2-D gel experiments. However, when treatments involve more than a single factor, the researcher must then specify their relationships—factorial, hierarchical, or repeated. For example, drug

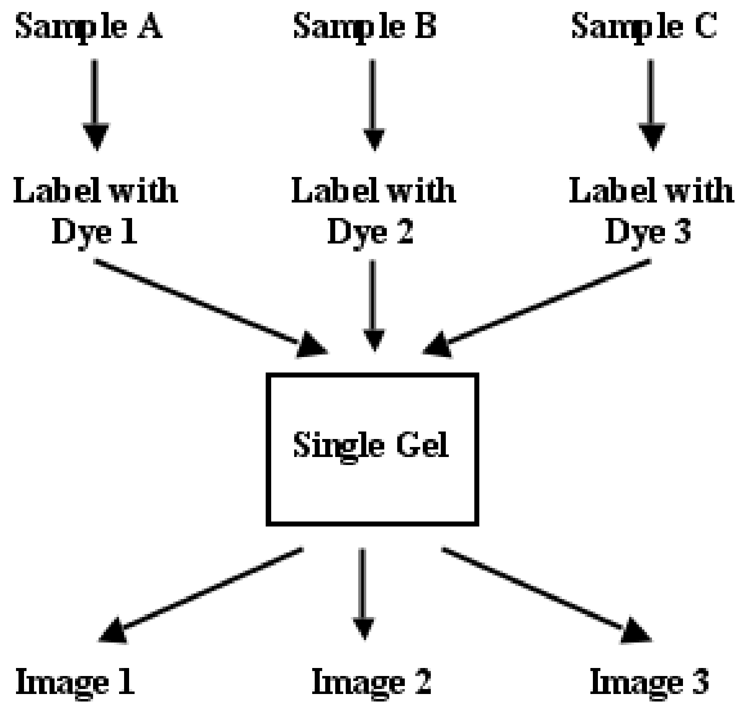

Figure 3: 2-D DIGE: 3 samples are labeled with a different dye, mixed together, and then co-elecrophoresed on a single gel.

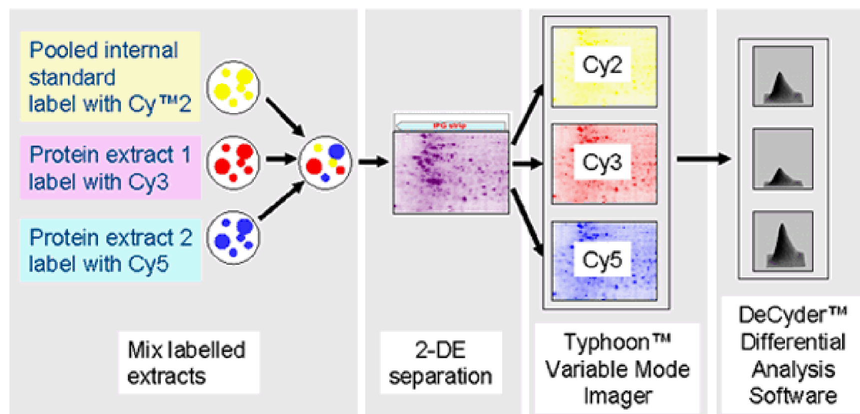

Figure 4: Workflow for the Ettan DIGE System

treatment may be hierarchically arranged relative to disease state and sampling may be repeated over the life cycle of the disease. The researcher must also choose an experimental design that is consistent with the treatment design.

The experimental design must ensure that variability is controlled and that valid error terms are available for the treatment effects of interest. The principal difference between 2-D PAGE and 2-D DIGE is that gel-to-gel variation is controlled in the latter. This is done by blocking, which removes the gel variation from the error term used for testing the treatment effects in 2-D DIGE. The completely randomized design in 2-D PAGE experiments is contrasted with the (incomplete) block designs of 2-D DIGE experiments. Either of these designs might be extended by repeated sampling over the life cycle of a disease, which incorporates a split plot into the above whole-plot design (either completely randomized or blocked).

Suppose we are designing an experiment for comparing 2 disease states, e.g., by sampling cancerous and normal tissues. In order to assess the induced biological changes due to cancer, an appropriate number of biological replicates must be used, which can be determined by power calculations. With the conventional one-sample-per-gel methods, the experiment can be designed as in Table 1—with 4 replicates per treatment in this case. The tissue-to-tissue variation of this approach is typically very high, often much higher than the inherent biological variation within a tissue sample. The tissue is the appropriate experimental unit if inferences are to be made to a population of patients (assuming one tissue sample per patient). Subsampling within a tissue sample is probably of little value.

This design is a completely randomized design with all gels aligned to a reference gel—typically one of the control samples. This process aligns the gels so that corresponding proteins (spots) are identifiable, but it does not remove all sources of gel-to-gel variation. Unfortunately, time-consuming manual alignment is often required and accuracy is difficult to achieve.

Table 1: A Classical 2-D Experimental Design

Gel	Treatments
1	Control A1
2	Control A2
3	Control A3
4	Control A4
5	Treated B1
6	Treated B2
7	Treated B3
8	Treated B4

A more precise design (by including a common pooled internal standard sample in every gel) can be obtained by using a two-dye DIGE experiment (Table 2). The internal standard is labeled with one of the dyes and the experimental samples are labeled with another dye. This design is still a completely randomized design since multiple treatments are not multiplexed on the same gel. However, normalization of the pooled internal standard across gels allows the ratios of relative expressions of the same protein across gels to be compared directly.

The normalization of pooled samples in 2-D DIGE experiments is easier to do than the normalization required for 2-D PAGE since the pooled samples are from common stock and all proteins are present. Alignment is not required and the data is obtained as the ratio of treated to pooled abundances within each gel.

Table 2: A 2-D DIGE Experimental Design with 2 Dyes		
Gel	Cy2	Cy3
1	Pooled	Control A1
2	Pooled	Control A2
3	Pooled	Control A3
4	Pooled	Control A4
5	Pooled	Treated B1
6	Pooled	Treated B2
7	Pooled	Treated B3
8	Pooled	Treated B4

The above experiment can be further improved if one more dye is utilized in the experiment (Table 3). Here, both the control and treated samples, as well as the pooled internal standard, are run on the same gel. These samples are then imaged separately. However, since they were run on the sample gel, the images can be perfectly overlaid without warping. With the same treatment groups and the same biological replicates, the number of gels required is halved.

Table 3: A 2-D DIGE Experimental Design with 3 Dyes			
Gel	Cy2	Cy3	Cy5
1	Pooled	Control A1	Treated B3
2	Pooled	Treated B1	Control A3
3	Pooled	Treated B4	Control A4
4	Pooled	Control A2	Treated B2

With 2 treatments and an internal standard, a 2-D DIGE experiment can be modeled as a block design with blocks corresponding to gels. With 3 or more treatments, a 2-D DIGE experiment could be modeled as an incomplete block design. A large statistical literature is available for designing incomplete block designs.

4 Problems Associated with Specialized Software Packages

Specialized software packages, e.g., Amersham's DeCyder differential analysis software and Nonlinear's Cross-Stain Analysis (CSA), can be used to analyze the stains resulting from DIGE experiments. The ultimate goal of these analysis systems is to find differentially expressed proteins, also called the biomarkers. These relevant proteins can then be selected (by a spot picker) and sent to a MALDI-MS for identification and analysis. Generally, these are fully automated gel-processing and analysis systems. However, these systems are proprietary and methods of selecting differentially expressed proteins are not documented. These systems are easy to use, but they only have limited statistical analysis techniques available.

Fortunately, both DeCyder and CSA allow the image-analysis data to be exported using XML and Oracle, respectively. Once the expression data is extracted, state-of-the-art statistical tools, e.g., those found in R packages (http://www.r-project.org), can then be used. The analyst can then intervene and do their own analyses using various analytical methods, including recent inventions such as random forests.

5 Statistical Analysis of Differentially Expressed Proteins

Recent advances in 2-D technology imply that thousands of proteins can be separated, resulting in large data sets containing protein expression and other information.

Generally, analyses of 2-D data is not much different from that of gene expression data since both have the so-called "small n, large p" problem which means, a large number of variables (proteins or genes) relative to a small number of cases (samples). Therefore, the gene expression data analysis tools can also be applied to protein expression data. To find the proteins or groups of proteins that express themselves differentially (significantly) across experimental groups, hypothesis tests and various classification methods come into play. The latter can also be used for the purpose of class prediction. To find natural groups or patterns in the data, which may represent unknown classes, clustering and pattern recognition techniques can be used.

The algorithms to identify differentially expressed proteins can be classified as filters and wrappers. Filter techniques include simple filters and refined filters. Simple filters use univariate test statistics such as t and F tests, whereas refined filters take into account the joint distribution of the protein expression measures. Wrappers employ classification methods implicitly to learn how to distinguish experimental groups. Two methodologies undergoing rapid development—multiple testing procedures and random forests (Breiman, 2001), are discussed in the following sections.

5.1 Multiple Testing Procedures

To identify differentially expressed proteins, a simple solution is to consider each protein individually and to select significant proteins based on univariate statistical approaches (simple filters), which include the t-test, analysis of variance (ANOVA) F-test, and nonparametric tests.

For each protein, the null hypothesis is: no (mean) differential expression among treatment levels. Two types of errors can occur: false positives (Type I errors) and false negatives (Type II errors). False positives occur if the test conclusion is that a protein is differentially expressed when it is not. False negatives take place if a differentially expressed protein is not detected.

Univariate approaches are easy-to-use. However, in practice, hundreds to thousands of hypotheses are tested simultaneously which results in too many false positives assuming independent tests. Thus, from a multivariate point of view, adjustments for multiple testing are necessary.

The biological question of differential expression can be restated as a problem in multiple testing: the simultaneous test for each protein with the null hypothesis of no association between the expression levels and covariates of interest, e.g., treatment/control status or cancerous/non-cancerous status. Multiple testing procedures are refined filters. The differentially expressed proteins are identified based on the adjusted p-values for a multiple testing procedure which takes into account the dependence structure among the protein expression levels (Dudoit *et al.*, 2002). Multiple testing procedures based on p-values that control the Type I family-wise error rate (FWER) are defined as:

$$p_j^{adj} = inf\{\alpha | H_j \text{ is rejected at FWER} = \alpha\}$$

for the null hypothesis H_j.

The FWER and the false discovery rate (FDR) are two ways to measure false positives. FWER measures the chance of having one or more false positives:

$$\text{FWER} = P(\text{number of false positives} \geq 1)$$

The FDR measures the expected percentage of false positives among all significant hypotheses:

$$\text{FDR} = E(\frac{\text{number of false positives}}{\text{number of true and false rejects}} | \text{at least one reject of null}).$$

In R (http://www.r-project.org), multiple testing procedures include resampling-based multiple testing procedures for controlling the FWER, such as Bonferroni or Westfall & Young (1993), as well as procedures for controlling the FDR, such as Benjamini & Hochberg (1995). These procedures are implemented for tests based on t-statistics, F-statistics, paired t-statistics, Wilcoxon statistics and adjusted p-values are reported as the results. FWER-based multiple testing procedures are conservative for proteomic experiments, whereas FDR-based testing procedures generally are more powerful.

5.2 Random Forests

The random forest algorithm was developed by Leo Breiman (2001). It is an aggregated version of the classification tree algorithm. A classification tree is a tree-structured classifier in which each child node represents a choice between two alternatives, and each terminal node represents a classification decision. Tree classifiers are easy to construct and interpret, but they have disadvantages. One of the biggest problems associated with classification trees is its instability. One way to reduce the variability and improve accuracy is to use bagging, i.e., bootstrap aggregating, a method to "average" over many classifiers to create an aggregated classifier. The random forests algorithm (Breiman & Cutler, 2004) has been gaining attention because it:

- is unexcelled in accuracy;
- has excellent performance on large data sets;
- is efficient in handling thousands of features without variable deletion

In proteomic studies, the datasets usually contain thousands of features, most of which are not very informative. Classification techniques such as classification trees, support vector machines or discriminant functions can be used to perform some type of feature selection. However, with a large number of irrelevant features, the classification performance might degrade substantially (Dudoit & Fridlyand 2003). Random forests overcome this problem.

The random forests algorithm combines filtering and wrapping. Like the multiple testing procedures, random forests also take into account the joint distribution of the protein expression measures. In this way, all the features are considered simultaneously, which allows the detection of proteins with weak main effects but strong interactions (Dudoit & Fridlyand, 2003). This algorithm provides feature importance ordering using a single run, which provides information on the differentially expressed proteins among the experimental groups.

The random forests algorithm works as follows: A forest, consisting of a large number of unpruned classification trees, is grown. Each single tree is based on

a bootstrap sample from the original data set. Therefore, by taking bootstrap samples, multiple versions of the classifier are generated (Breiman 1996). While growing a tree, a subset of variables, rather than all the variables, is selected at random to decide the best split at each node. The bootstrap aggregating, as well as random feature selection, make the trees in the forest look different and lower the correlation among the single tree classifiers (Breiman 2001). Each tree constitutes a decision rule. The decisions from all the trees are integrated to make the final determination. Usually, the majority rule is followed, but a more flexible threshold can also be chosen. To classify a new case with input vector x, we can classify x using each of the trees in the forest. Each tree makes a classification decision for this new case. The forest then chooses the classification having the most votes.

The random forest algorithm has a built-in mechanism for obtaining unbiased estimates of error rates. For each bootstrapped sample, about one-third of cases are left out and not used in the tree construction. The out-of-bag cases, also called OOB cases, play an important role in performance assessment.

REFERENCES

Benjamini, V. & Hochberg, V. (1995) "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Statist. Soc. B*, **57**, pp. 289–300.

Breiman, L. (2001) "Random forests," *Machine Learning*, **45**, pp. 5–32.

Breiman, L. & Cutler, A. (2004) "Random forests: classification/ clustering." Available at: http://www.stat.berkeley.edu/users/breiman/RandomForests

Breiman, L. (1996) "Bagging predictors," *Machine Learning* **24**, pp. 123–140.

Dudoit, S. & Fridlyand, J. (2003) "Classification in microarray experiments," in Speed, T. (ed) *Statistical analysis of gene expression microarray data*, Chapman&Hall/CRC.

Dudoit, S. et al. (2002) "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments, *Statistica Sinica* **12**, pp. 111–139.

Friedman, D.B., *et al.* (2004) "Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry," *Proteomics*, **4(3)**, pp. 793–811.

Lau, Andy, He, Qing-Yu & Chiu Jen-Fu (2003) "Proteomic technology and its biomedical application," *Acta Biochimica et Biophysica Sinica*, **35(11)**, pp. 965–975.

O'Farrell, P.H. (1975) "High resolution two-dimensional electrophoresis of proteins," *J. Biol. Chem.*, **250**, pp. 4007–4021.

Roy, A. *et al.* (2003) "Analyzing two-dimensional gel images," Technical Report, Bioinformatics Research Center, UMBC. Available at: http://www.brc.umbc.edu/research/TechReport1.pdf

Westfall, P.H. & Young, S.S. (1993) *Resampling-based multiple testing: examples and methods for p-value adjustment*," John Wiley & Sons, New York.

Ünlü, M.E., Morgan, M.E. & Minden, J.S. (1997) "Difference gel electrophoresis: a single gel method for detecting changes in protein extracts," *Electrophoresis*, **18**, pp. 2071–2077.