Available online at www.sciencedirect.com

ScienceDirect

Journal of Chromatography B, 849 (2007) 261–272

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Statistics for proteomics: Experimental design and 2-DE differential analysis☆

Jean-François Chich[a,*], Olivier David[b], Fanny Villers[b], Brigitte Schaeffer[b], Didier Lutomski[c], Sylvie Huet[b]

[a] INRA, Biologie Physico-Chimique des Prions, VIM 78352 Jouy-en-Josas Cedex, France
[b] INRA, Mathématiques et Informatique Appliquées, MIA 78352 Jouy-en-Josas Cedex, France
[c] Université Paris XIII, Laboratoire de Biochimie des Protéines et Protéomique, UMR CNRS BioMoCeTi, UFR SMBH Léonard de Vinci, 74 rue Marcel Cachin, F-93017 Bobigny, France

Received 6 June 2006; accepted 8 September 2006
Available online 1 November 2006

Abstract

Proteomics relies on the separation of complex protein mixtures using bidimensional electrophoresis. This approach is largely used to detect the expression variations of proteins prepared from two or more samples. Recently, attention was drawn on the reliability of the results published in literature. Among the critical points identified were experimental design, differential analysis and the problem of missing data, all problems where statistics can be of help. Using examples and terms understandable by biologists, we describe how a collaboration between biologists and statisticians can improve reliability of results and confidence in conclusions.

Keywords: Experimental design; Differential analysis; Proteomics; Two-dimensional gel electrophoresis

1. Introduction

The term "proteomics" appeared in 1995 [1,2]. The primary goal of this new discipline was to study the protein complement of a genome but it rapidly appeared that this task was far from reach even if some ambitious and international initiatives, like Human Proteome Organisation (HUPO) founded in 2001 [3], were undertaken.

Proteomics relies mainly on the separation of a complex mixture of proteins by bidimensional electrophoresis (2-DE), mass measurement of peptides generated after spot proteolysis by mass spectrometry and search in databases.

Constant technical improvements were performed over the years, in particular accuracy and easiness of use of mass spectrometers and database enrichment. However, numerous publications are now considered of questionable quality. To improve overall quality and results reliability, four weak points to consider were identified recently [4]. These points concern experimental design, analysis of protein abundance data, confidence in protein identification by mass spectrometry and analytical incompleteness. The third point has been already discussed [5,6] and we will turn toward the three other points.

The role and importance of experimental design were described for transcriptomics but less frequently for proteomics. While proteomics cannot usually handle as much data as transcriptomics, the importance of experimental design should be emphasized. We show here how statistics can help to define suitable experimental designs, using classical knowledge on this subject [7,8] and knowledge issued from transcriptomics [9–11]. We show that establishing an experimental design in a dynamic collaboration between biologists and statisticians is useful to forecast sampling or experimental biases. In particular, experimental design allows to limit systematic errors, to improve precision of subsequent statistical tests and contributes thus to reduce the number of false positives.

Differential analysis of spot volume is generally handled by using commercial software packages that propose statistical tools to help to conclude on the significance of variation. Probably most biologists consider that these packages are sufficient

☆ This paper is part of a special volume entitled "Analytical Tools for Proteomics", guest edited by Erich Heftmann.

* Corresponding author. Tel.: +33 1 34 65 29 12; fax: +33 1 34 65 26 21.
E-mail address: chich@jouy.inra.fr (J.-F. Chich).

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2006.09.033

for their purpose. We show here that statisticians propose powerful tools that can be used for improving data analysis. These tools can help to rationalize the decisions on significance and can draw attention on the imperfections of gels, spot mismatches and other artifacts of 2-DE gels.

Moreover, analytical incompleteness is encountered in proteomics, especially in spots missing (missing data) on one or more gels coming from the same series. The particular and difficult problem of these missing data on 2-DE gels is generally due to experimental problems and must be taken into account in the statistical analysis. However, the questions raised by these problems are not trivial and are discussed.

Lastly, we emphasize the interest of collaborations between biologists and statisticians at different levels of proteomics experiments, in order to draw the most robust conclusions from experimental results.

2. Experimental designs

2.1. A practical example: cell line proteomics

In this part, an example corresponding to a question of proteomics applied to cell biology is described. The situation presented here can be easily applied to other situations. Cell lines are largely used as biological tools to study effects of an infection, the effect(s) of a drug and so on. These immortalized cells are easy to grow and are a useful source of large amounts of proteins needed for proteomic studies.

However, proteomics on these cells is subject to variabilities that must be taken into account when possible. The first variability is clonal drift due to the number of passages of cells (a technique for diluting cells that enables them to be kept alive and growing under cultured conditions for extended periods of time), acquired chronic contaminations or metabolic modifications due to culture media. Unfortunately, this variability cannot be easily considered since cell lines are heterogeneous (see for example, in the prion field [12,13]) from passage to passage. Thus, cloning artifacts can induce false conclusions.

Another source of variability can be due to small biological variations (cell growth variability, etc.). If a researcher wants to study the effect of a drug on a cell line, he adds the drug in a series of flasks and a placebo in another. However, differences observed between flasks for the same treatment account for a variability that can be taken into account by experimental design and statistical method(s).

The second variability is a technical one. It involves the preparation of cells and the protein solubilization prior to the separation by 2-DE: for example, if cells are washed using a cold ice buffer, it is likely that they will express chaperones that might interfere with the studied phenomenon. A variability can be due also to the apparatus used for cell culture (variable heat or humidity of an incubator) or used for protein separation. One aim of statistical methods presented in this article is to take account of these sources of variation.

A study using cell lines, was undertook by two of us [14] and it is used as a basis to present ideas underlying experimental design. Two cell lines were used to study prion infection. The first one is GT1–7, a subclone of GT1, a highly differentiated hypothalamic cell line displaying a number of neuron functions [15]. The second line results from infection of GT1–7 clone with the Chandler strain of prion (ScGT1–7: Sc for scrapie, the prion disease of sheep) [16]. Though GT1 and ScGT1 were described to be in a steady state after, respectively, 12 passages [15] and 55 passages post-infection [16], clonal drift could occur in cell populations grown in so different laboratories for years and generate variability. In order to study how prion infection affects cellular metabolism, a proteomic approach was made on both cell lines.

2.2. Two-phase experiment

Our *two-phase experiment* was performed as schematized in Fig. 1A. The first phase consists of the cell cultures. GT1–7 and ScGT1–7 cells were grown separately for several passages in order to obtain an amount of proteins compatible with a proteomic analysis. The second phase consists of protein extraction from these cells and to separate them by 2-DE.

The objective of this two-phase experiment was to compare protein abundance according to two conditions. These conditions are defined as healthy (H), for GT1–7, and infected (Sc) for ScGT1–7. In the first phase, cell cultures, sample preparation and/or pooling represent the *biological phase*. Separation of proteins by 2-DE, organization of gel runs and staining represent the *technical phase*. Technical variability due to the electrophoresis apparatus was considered to be non-significant because running conditions (current, buffer and temperature . . .) were controlled; thus six 2-DE (IEF then SDS-PAGE) were run for H (represented by “batch 1 ”), followed by six 2-DE for Sc (“batch 2”).

After silver staining, gels were scanned and classically analyzed using the software ImageMaster 4.01 (GE-Healthcare Bioscience). After spot volume measurements and matching, differential analysis (Section 3) was performed and spots showing significant abundancy variations (genome.jouy.inra.fr/gt1) between H and Sc were identified by mass spectrometry. The expression of the proteins corresponding to these spots can be considered to be specifically affected by prion chronic infection. They can be potential markers of prion disease or targets for drug therapy.

Thus, variability arises generally from both phases, calling for rational implementation of work plans [17,18]. It should be emphasized that establishing an experimental design allows bias reduction and increased confidence in experimental results.

2.3. Constructing experimental blocks or blocking

If the researcher suspects variability during gel runs, for example, he can account for this variability by constructing *blocks*. A block is a set of experimental materials considered as consistent. The objective of blocking is to make the comparison between observed conditions, with little as possible dependent on artefacts or heterogeneities (differences between gels, etc.) and as much as possible dependent on the differences the

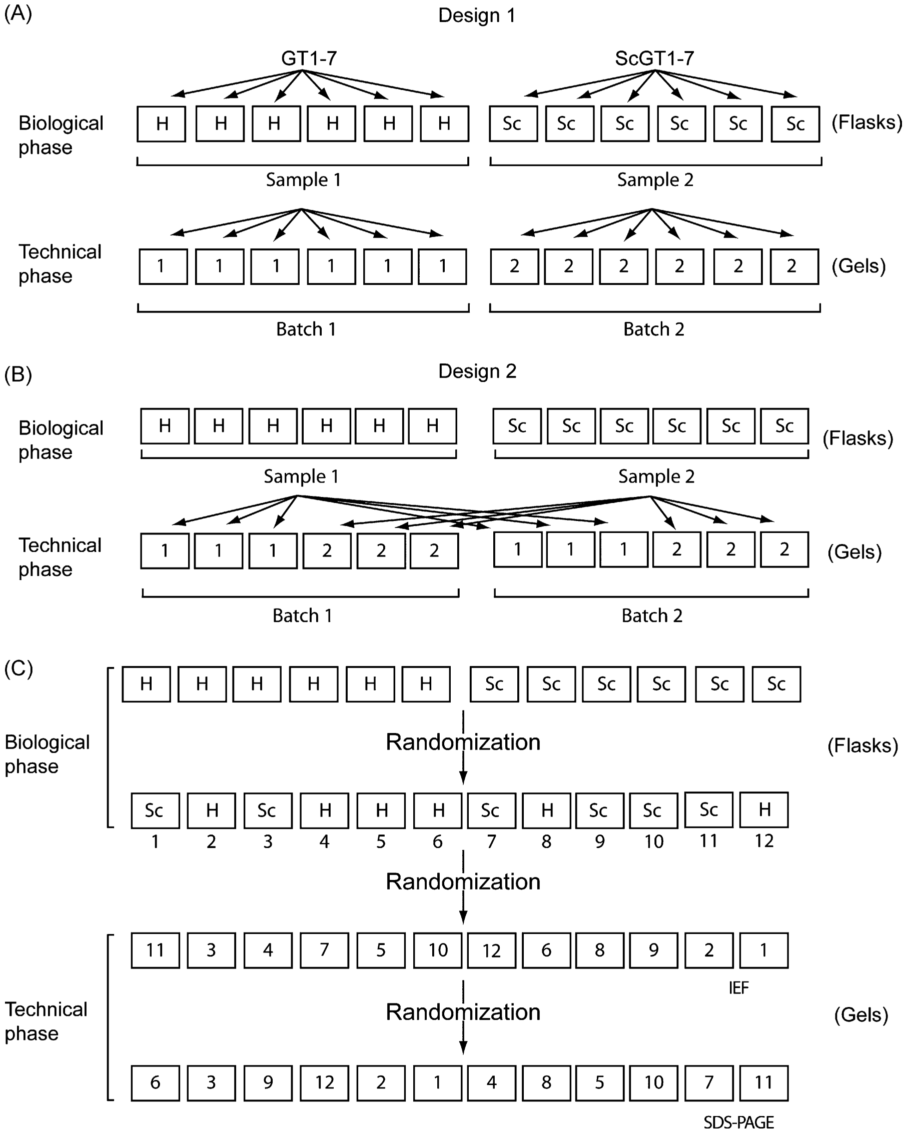

Fig. 1. Examples of experimental designs. Squares correspond to flasks in the biological phase and to gels in the technical phase. H denotes healthy GT1–7 cells and Sc denotes ScGT1–7 (chronically infected GT1–7 cells). (A) Unrandomized design for experiment on cell lines. Healthy cell (H) and infected cell (Sc) were cultured in several flasks. Samples were pooled and 2-DE were performed. (B) Unrandomized design with "blocks", the heterogeneities suspected in the technical phase can be taken into account in differential analysis. (C) Example of a randomized design without blocking. Numbers denote sample landmarks. GT1–7 cells (H or Sc) were cultured in different flasks. Heterogeneities are suspected in biological and technical phases and a randomization is performed for both phases. The effect of suspected potential biases is eliminated.

researcher needs to characterize (effect of infection in the example). Thus, blocking takes into account heterogeneities known or suspected from the beginning of the experiment and improves the precision of the following statistical analysis. When blocking is used, both the blocks and the allocation of conditions to blocks have to be chosen (see [7,8] for more information on blocking).

An example of blocking is shown in Fig. 1B. Cultures are performed as described in the previous schema. Variabilities being suspected between "batch 1" and "batch 2", the researcher constructs one block that will run in "batch 1" and a second block that will run in "batch 2". Each block is composed of three gels (isoelectric focusing followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis) with proteins from H (sample 1) and three gels with proteins from Sc (sample 2). If there is a heterogeneity due to the electrophoresis apparatus, the researcher will be able to identify it by comparing gels loaded with sample 1 from both batches and gels loaded with sample 2 also from both batches. Moreover, the researcher will be able both to quantify this effect and to improve precision in differential analysis. The remaining variabilities are due to differences between H and Sc that the researcher wants to characterize.

The technique of blocking for the technical phase was described for transcriptomics, to take into account the heterogeneities linked to arrays and dyes in microarrays experiments and complicated designs were proposed [9–11,19]. Pairing is a particular blocking with blocks of size 2 that was described for transcriptomics but also for 2D difference gel electrophoresis (2D-DIGE) experiments [20]. Blocking can also be used in the biological phase [21].

2.4. Randomization

Because it is difficult for a researcher to identify all sources of variation using his judgement or experience, the use of randomization is a common practice. Randomization was recommended for microarray experiments [9,10]. To our knowledge, randomization for proteomics was described only for experimental design in mass spectrometry [6].

We will consider again the previously described experiment but with unknown heterogeneities both in biological and in technical phases. This experiment thus calls for a randomization for both phases [18]. The schema of the experimental design is shown in Fig. 1C. The incubator is supposed to be heterogeneous (heat, CO_2 distribution, etc.) and a randomization is performed before placing the culture flasks with landmarks (six "H" and six "Sc") inside. Cells are cultured and proteins are extracted individually from each flask. To avoid biases due to, for example, preferential current run, differences in strip or gel batches, buffer composition, randomization can be performed at the different levels of the technical phase: (a) allocation of proteic samples to strips, (b) strip placement in IEF apparatus, (c) strip deposit at the top of second dimension gels, after IEF, (d) gel placement in migration tank. Subsequent gel staining, image analysis and statistical treatment are performed as usual. The example presented here is a simple randomization. However, designs with blocks can also be randomized [7,8].

In brief, randomization reduces systematic errors when comparing the conditions and estimating the precision of the results.

2.5. Replication

Although randomization and/or blocking allow control of extraneous variables, the result of a single 2-DE experiment is not satisfactory due to the intrinsic variability of the method. In proteomics, variability can be found in biological phase as well as in technical phase. Variability was estimated to be high in 2-DE [22] and it can lie in the number of spots detected, on the variance of the spot volume measured (discussed in Section 3) by software analysis. The authors showed also that a fully manual analysis is more reliable than a fully automated one. However, replication in both phases can be problematic due to the low amount of initial sample or due to low protein content in a sample (biopsy for example). Some tools are available to estimate the experiment precision a priori on the basis of the researcher objectives (www.emphron.com/) .

In order to illustrate the concept of replication, we present three examples, shown in Fig. 2. The first design (Fig. 2A) has one biological replication (one flask for H, one flask for Sc) and six technical replications (six 2-DE per flask). Its drawback is that biological variance cannot be estimated. The differential analysis will be based on the technical variance only and the precision of analysis will be over-estimated. This situation

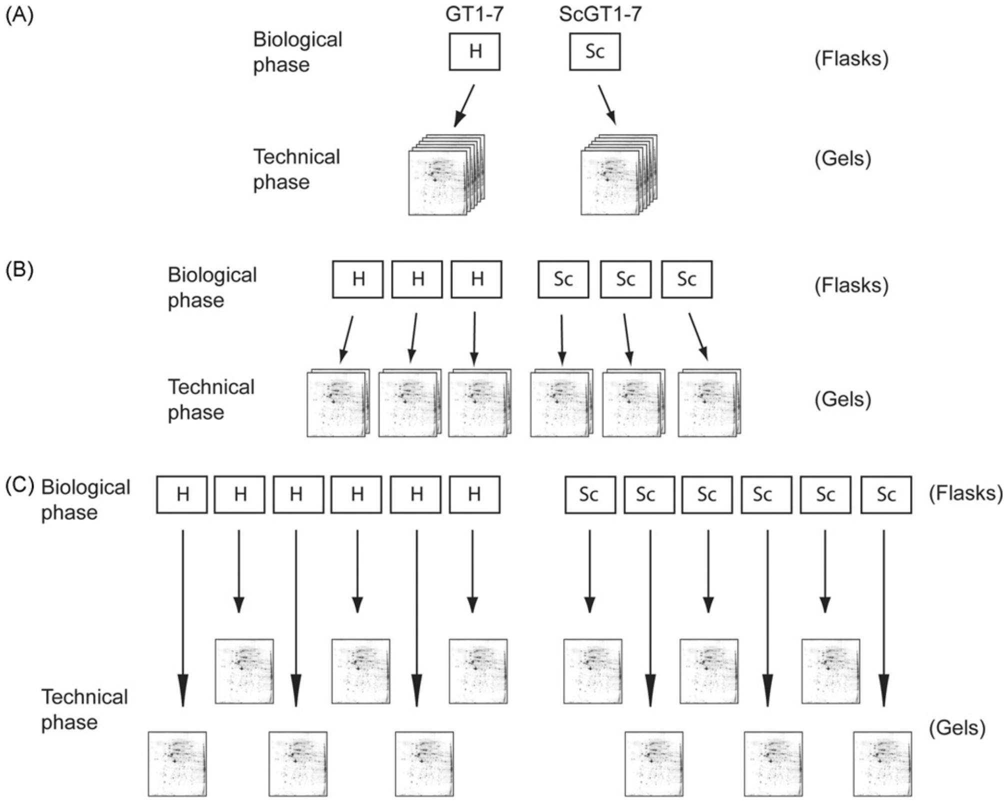

Fig. 2. Example of experimental designs using replication. Replication should take into account technical limits and the fiability of expected results. (A) Biological phase without replication and technical phase with six replications for each sample. (B) Biological phase with three replications and technical phase with two replications for each sample. (C) Biological phase with six replications and technical phase without replication.

can increase the number of false positives. Similarly, if protein extracts from several flasks are pooled into a single sample for each condition (as in Fig. 1A), the differential analysis will also be based on technical variance only and the number of false positives may increase. The use of several pools per condition avoids this problem (see [9,10] for a discussion on pooling). The second design (Fig. 2B) shows three biological replications (three flasks for H and three for Sc) and two technical replications (two 2-DE per flask). The third design has six biological replications and only one technical replication. For both designs, six 2-DE were made for each H and Sc. Both designs have thus the same ability to account for the technical variance. However, the third design (Fig. 2C) has more biological replications and the subsequent analysis will be more accurate.

Cell lines do not give the opportunity to perform *true* biological independent replications because all cells derive from a unique cell. However, several factors (clonal drift, culture media modifications, de novo latent infections, etc.) can induce, at least theoretically, variations that can be assessed by replication of flask cultures without pooling. True biological replications can be envisaged using primary culture cells [23] because these cultures can be designed to be as independent as possible from each other.

Replications are necessary to assess and increase the precision of the subsequent analysis results. From a statistical point of view, biological replications are more efficient than technical replications. In practice, biologists use technical replications due to the well known variability of the 2-DE technique. Literature studying the experimental design of microarray experiments emphasizes using replications to assess and control experimental variability [9–11]. However, these recommendations are discussed since a system of unreplicated experiments was described [11]. For proteomics, discussions on the number of observations are available in [20,24].

2.6. Discussion

In summary, experimental design is an important part of biological experiments, especially in proteomics where technical methods are long, difficult with intrinsic technical variability. The situation of two-phase experiment is often encountered in proteomics [21] and transcriptomics [10] and experimental design offers several tools that can be useful to minimize the effect of variabilities on the results. The choice of a suitable experimental design can improve the reliability of true positives detection and reduces the number of false positives in the subsequent differential analysis.

While general guidelines can be drawn from the examples shown here, it should be kept in mind that establishing an experimental design is a compromise between the availability of biological material, the technical difficulties of the approach and the reliability of the expected results.

3. Differential analysis

The aim of the differential analysis is to detect the proteins whose abundance differs according to the condition. In statistical terms, this comes to test simultaneously a large number of hypotheses: for each spot j, we have to test the hypothesis H_j that the spot volume does not differ according to the condition, or in other words that the corresponding protein is not variant in abundance. This problem has been extensively studied for the determination of the differentially expressed genes in microarray experiments [25–30]. Nevertheless, the adaptation of these works to data coming from 2-DE is not direct for the following reasons:

- The data present a great variability due to the complexity of the image analysis [31–35].
- The number of missing data is large, up to 50–60% [36].
- Generally the replication number is small, between 3 and 6.
- Some observations are irrelevant. This is the case when the image analysis process detects spots in trails or overlapping spots.

The statistical analysis provides a list of variant spots, using a procedure that is based on a statistical model and on the data. The parameters controlling the procedure, for example, the probability of deciding wrongly that a spot is variant, are estimated. In reality, because some observations are irrelevant, this list is only a list of potentially interesting spots and must be carefully examined before validation.

Detection of pertinent spots is thus an iterative procedure between the researcher and the results of the statistical analysis. Nevertheless, providing methodological tools for detecting irrelevant observations and variant spots, may help to best analyze the data.

The testing procedure consists of choosing a *test statistic* and deciding if the hypothesis H_j is rejected or not. These problems are treated in Section 3.1. The next question to consider is *what does the procedure control*? This is the object of Section 3.2.

In practice, preliminary analysis is necessary in order to verify that the statistical model is consistent with the data. Another important question is *which strategy for missing data*? All this will be discussed in Section 3.3.

3.1. Statistical models and testing methods

Statistical problems involved in gel analysis have been discussed [34,36–39]. In this paper we consider two approaches, the spot by spot approach (or univariate analysis) where the test statistic for testing the hypothesis H_j is based on the data observed for the spot j only, and the global approach (or multivariate analysis) where the test of H_j is based on the results of an analysis of variance considering all the observations together. The methods taking into account the experimental design, for example blocking, are mentioned in Section 3.1.3.

Let us denote by Y_{jcg} the response for spot j, under condition c, on gel g. The response is the percentage of volume on gel g or a suitable transformation of spot volume, that is a transformation for which the statistical assumptions needed for applying the methods described below will be reasonably satisfied. The problem of choosing a transformation is discussed in Section 3.3.3.

3.1.1. Spot by spot analysis

In the spot by spot analysis, we assume that the Y_{jcg}'s are distributed as Gaussian independent variables with mean m_{jc} and variance σ_j^2. Testing H_j comes to test that the m_{jc}'s are all equal using the classical Fisher or Student statistics test (see Box 1). Several variants of Student statistics have been proposed [40]. Non-parametric tests can also be applied such as the Mann–Whitney (or Kolmogorov) test. If we denote by F_c the distribution function of the observations of spot *j* under condition *c*, the Mann–Whitney test consists in testing that the distributions F_c are identical. It does not need to assume Gaussian distribution.

Box 1. Spot by spot analysis: test of H_j.
(A) Comparison of two conditions
Let us denote $Y_{jc.}$ as the empirical mean of the Y_{jcg}, n_{jc} the number of observations of spot *j* under condition *c*, and SCR$_{jc}$ as the residual sum of squares defined as follows:

$$\mathrm{SCR}_{jc} = \sum_{g=1}^{n_{jc}} (Y_{jcg} - Y_{jc.})^2.$$

If $n_{j1} + n_{j2} \geq 3$, the test statistic for testing H_j: "$m_{j1} = m_{j2}$" against "$m_{j1} \neq m_{j2}$" is defined as follows:

$$S_j = \frac{|Y_{j1.} - Y_{j2.}|}{\sqrt{(\mathrm{SCR}_{j1} + \mathrm{SCR}_{j2})/(n_{j1} + n_{j2} - 2)((1/n_{j1}) + (1/n_{j2}))}}.$$

Let us denote by Z_d a Student variable with *d* degrees of freedom. If H_j is true, then S_j is distributed as $|Z_{n_{j1}+n_{j2}-2}|$ and the *p*-value p_j is defined as

$$p_j = \mathrm{pr}(|Z_{n_{j1}+n_{j2}-2}| > S_j).$$

(B) Comparison of *C* conditions, $C \geq 3$. If n_j, the total number of observations for spot *j*, is greater than $C + 1$, the test statistic for testing H_j: "$m_{j1} = \cdots = m_{jC}$" against the alternative that there exist two conditions c, c' such that "$m_{jc} \neq m_{jc'}$" is defined as follows:

$$S_j = \frac{n_j - C}{C - 1} \frac{\sum_{c=1}^{C} n_{jc}(Y_{jc.} - Y_{j..})^2}{\sum_{c=1}^{C} \mathrm{SCR}_{jc}}$$

where $Y_{jc.}$, n_{jc} and SCR$_{jc}$ are defined as above, and $Y_{j..}$ is the mean of the responses for spot *j*.
Let us denote by F_{d_1,d_2} a Fisher variable with d_1 and d_2 degrees of freedom. If H_j is true, then S_j is distributed as F_{C-1,n_j-C} and the *p*-value p_j is defined as

$$p_j = \mathrm{pr}(F_{C-1,n_j-C} > S_j).$$

Box 2. Global ANOVA approach: test of H_j.
Assume that we compare two conditions. The test statistic S_j, is based on the least squares estimators of the differences $(\mathrm{SpC})_{j1} - (\mathrm{SpC})_{j2}$'s divided by their estimated standard errors. The S_j's are distributed as $|Z_{n-GC-(J-1)(C-1)}|$, where $n = \sum_{j=1}^{J} n_j$ is the total number of observations, *J* the number of spots, *G* is the number of gels for each condition.
The *p*-values p_j are defined as in Box 1.

3.1.2. Global analysis

In the global analysis, we start with an analysis of variance (ANOVA) model, where the response Y_{jcg} is modeled as follows:

$$Y_{jcg} = (G)_g + (\mathrm{SpC})_{jc} + E_{jcg}$$

where $(G)_g$ is the effect of gel *g* (the part of variability due to gel *g* in the response *Y*), and $(\mathrm{SpC})_{jc}$ is the spot × condition effect defined as the effect of spot *j* under condition *c* on the response *Y*. The random errors E_{jcg} are distributed as centered Gaussian independent variables with the same variance σ^2. Testing H_j comes to test that the differences between the spot × condition effects $(\mathrm{SpC})_{jc}$ are zero using the Student or Fisher statistic (see Box 2).

3.1.3. Analyses with block effects

The block effects in the experimental design have to be taken into account in the modeling [10,17–19,21]. Let us consider the example given by Design 2 of Fig. 1B. Let Y_{jcag} be the response of spot *j* under condition *c* measured on gel *g* in experimental apparatus *a*. In the spot by spot approach, for each spot *j*, we consider the following ANOVA model,

$$Y_{jcag} = (A)_a + (C)_c + (\mathrm{AC})_{ac} + E_{jcag}$$

where $(A)_a$ is the mean effect of apparatus *a*, $(C)_c$ the mean effect of condition *c*, and $(\mathrm{AC})_{ac}$ is an apparatus × condition effect. The last effect is called an interaction effect meaning that the condition effect may differ according to the apparatus. For each spot *j* the model parameters are estimated, and testing H_j comes to test that the $(C)_c$'s are all equal. In the global approach model, the response Y_{jcag} is modeled as follows:

$$Y_{jcag} = (G)_g + (\mathrm{SpA})_{ja} + (\mathrm{SpC})_{jc} + (\mathrm{SpAC})_{jac} + E_{jcag}$$

where $(G)_g$ is the gel effect, $(\mathrm{SpA})_{ja}$ the effect of spot *j* observed in apparatus *a*, $(\mathrm{SpAC})_{jac}$ an apparatus × condition × spot effect, and $(\mathrm{SpC})_{jc}$ is the effect of spot *j* under condition *c*. As before, testing H_j comes to test that the differences between the spot × condition effects are zero.

3.1.4. Decision rules

The decision rule for rejecting H_j is the following: for each spot *j*, we calculate the *p*-value p_j, defined as the probability for rejecting H_j when H_j is true. The hypothesis H_j is rejected when p_j is small. Therefore the set of variant spots corresponds to the smallest p-values. For example, we can choose to reject H_j when p_j is smaller than 5%.

3.2. Controlling the testing procedure

The question that arises is *how many errors are we doing when testing J hypotheses*? We commit an error in two situations:

1. When we decide that a spot is a variant when it is not. Such an error leads to a false positive. The number of such errors is denoted FPos.
2. When we decide that a spot is not a variant when it is. Such an error leads to a false negative.

The control and elimination of false positives is important in order to avoid drawing false conclusions, particularly when the conclusions are the starting point of a new costly experiment.

If we run the testing procedure with $\alpha = 0.05$, then we expect up to 5% of the total number of spots to be variants by chance alone. In other words, if the differential analysis is performed with $J = 1000$, then we expect up to 50 spots to be wrongly detected as variants. Such a control is not acceptable. Two methods described below overcome this problem.

3.2.1. The family-wise error rate or FWER

The FWER is defined as the probability of having at least one false positive. It can be shown that if each hypothesis H_j is rejected when $p_j \leq \alpha$, then FWER $\leq \alpha J$. Choosing $\alpha = 0.05/J$ leads to FWER ≤ 0.05. This choice of α is known as the Bonferroni correction. This procedure allows very few occurrences of false positives, but makes the decision rule that a spot is differentially expressed very strict.

3.2.2. The false discovery rate or FDR

If R denotes the number of rejected hypotheses, the FDR is defined as the expected value of the ratio FPos$/R$ when R is positive. Controlling the FDR at level 0.05 means that up to 5% of spots among the spots detected as variants, are identified by chance. This procedure proposed by Benjamini and Hochberg [41] is detailed in Fig. 3. Several variants and improvements of this procedure have been proposed [40,30,42].

3.2.3. Which one to choose?

The choice between FDR or FWER procedure should be made on the basis of the aim of the research. If the differential analysis is a work whose objective is to list potential proteins involved in a physiological process, FDR method provides a reliable tool. If the objective is to determine if a protein is a potential biomarker, according to [4], false positives must be totally eliminated and FWER method should be preferred. However, in this case, further investigation is needed after this step to validate the biomarker.

3.3. Preliminary analyses

3.3.1. Removing irrelevant data

Testing simultaneously a large number of hypotheses has a cost: the larger the J, the more the procedure is strict. Therefore retaining in the differential analysis spots for which the observations are not relevant may compromise the differential analysis for the other spots. The amount of protein quantified in each spot can be computed when the spots are correctly detected by the image analysis software, but 2-DE images present smears and trails corresponding to migration artifacts. Those spots are frequently located near the left or right side of the image, corresponding to zones of accumulation of protein not within the pH gradient used, and around overabundant proteins, such as actin

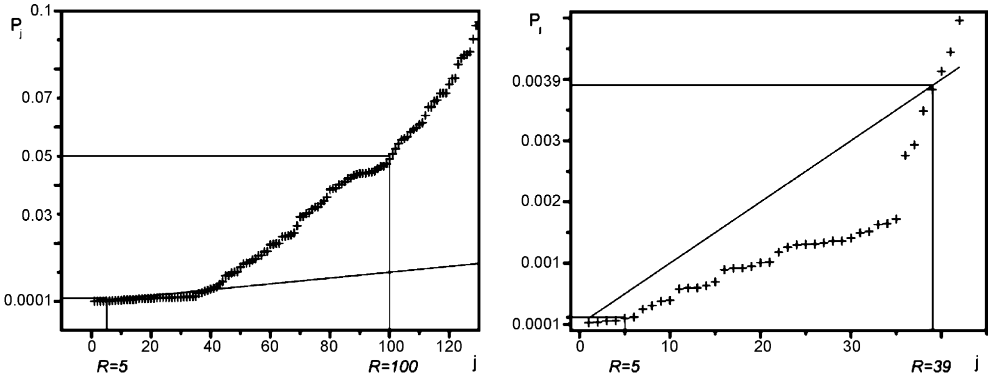

Fig. 3. Decision rules. Once the p-values p_j are calculated, it remains to define a threshold, such that the hypothesis H_j is rejected as soon as p_j is smaller than the threshold. Let us formulate the problem in another way by considering the set of ordered p-values into ascending order, $p_{(1)} < p_{(2)} < \cdots < p_{(J)}$, and a set of thresholds that may depend on j denoted τ_j. The number of rejected hypotheses H_j, R, is defined as the largest j such that $p_{(j)} \leq \tau_j$. Finally, we reject H_j if $p_j \leq \tau_R$. If all the $p_{(j)}$'s satisfy $p_{(j)} > \tau_j$, then $R = 0$: none of the hypotheses H_j is rejected.

- If τ_j is constant and equal to α, then R is simply the number of p-values that are smaller than α. We can take $\alpha = 0.05$, or apply the Bonferroni correction with $\alpha = 0.05/J$.
- The method proposed by Benjamini and Hochberg takes $\tau_j = 0.05j/J$. Then H_j is rejected if $p_j \leq 0.05R/J$. They have shown that the FDR is controlled as follows: FDR $\leq 0.05T/J$ where T is the number of spots that are not differentially expressed.

These methods are illustrated by the graphics of p-values $p_{(j)}$ in ascending order as function of j, for $j = 1, \ldots, 125$ on the left and $j = 1, \ldots, 42$ on the right. These data are coming from a simulated example with $J = 500$. The number of rejected H_j equals 100 if $\tau_j = 0.05$, 5 if $\tau_j = 0.05/500$ and 39 if the Benjamini and Hochberg's method is used with $\tau_j = 0.05j/J$.

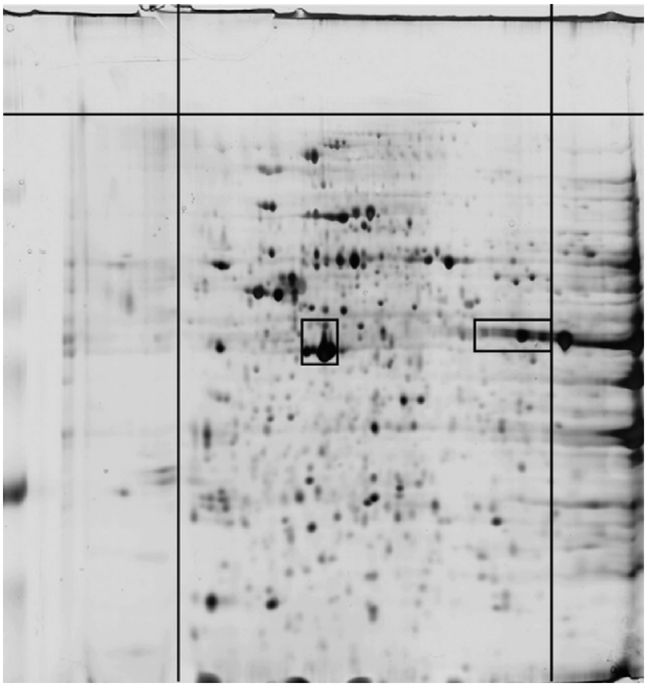

Fig. 4. Removing irrelevant spots. Spots near the left and right sides of the gel and near the top are deleted as well as spots located near actin and tubulin that are overabundant.

or tubulin for example. To improve the analysis, those spots are deleted (see Fig. 4).

3.3.2. Checking gel replications within conditions

The experiment can be used for differential analysis if within each condition, the gels can be viewed as replications of the same observation. However, the classification of gels into conditions may be uncertain because of biological or technical variability in the experiment.

Data-mining methods are suitable for checking this assumption, considering the gels as the experimental units (the cases) and the spots as the variables. Unsupervised methods such as principal component analysis (PCA) or hierarchical clustering, ignores the condition under which the gels were observed. Their aim is to discover structures from the evidence of the data matrix alone. If the structure proposed by the analysis consists of splitting the gels into conditions, then we are allowed to use the data set for differential analysis. If not, such an analysis may give information on what is going wrong in the data set.

Some of these methods such as PCA, cannot be used when many spots are missing, particularly in the context of 2-DE gels, because they are not missing at random. It is then possible to run the method only on spots observed on all the gels. Another possibility is to attribute values to missing data. This point will be discussed in Section 3.4.

3.3.3. Choice of a suitable transformation of the observed volumes

The testing procedures presented in Section 3.1 rely on statistical assumptions that should be checked.

The global approach assumes that there exists a suitable transformation of the observed volumes such that an ANOVA model is appropriate for modeling the data. The spot by spot approach assumes that there exists a transformation such that the gels within a condition are replications of the same observation, and the variance of the resulting response does not depend on the condition c. If these assumptions are not satisfied, then the testing procedures are false, that is, the calculation of the p-values is no longer valid. Usually the gel effect is eliminated by calculating for each spot the percentage of volume on the gel. Then the Student test or the Mann–Whitney test is used for testing H_j for each spot j. However, it has been observed that the larger the spot, the larger the variance [43,38]. It is therefore worthwhile to look for a transformation in order to stabilize the variance. This heterogeneity in the data variability is mainly due to a scale phenomenon, well-known when the observation (the spot volume) is a count (number of pixel × intensity).

In practice the problem is to find a transformation T of the volumes V_{jcg} or the percentage of volumes on each gel $\%V_{jcg}$, such that the transformed data $Y_{jcg} = T(V_{jcg})$ or $T(\%V_{jcg})$ satisfy the assumptions of Section 3.1. In some cases the logarithmic transformation is applied with success. In other cases, other transformations are more appropriate. The Box–Cox method allows to estimate an optimal transformation from the data [38,44]. Other normalization methods based on the data have been proposed [39,45]. In any case, graphics and statistical analyses are useful for detecting the presence of structures in the variance of the data [38,46]. Precisely let us denote by R_{jcg} the residuals defined as $R_{jcg} = Y_{jcg} - \hat{Y}_{jcg}$, where $\hat{Y}_{jcg}$ is the predicted value for spot j on gel g under condition c: in the spot by spot approach, $\hat{Y}_{jcg}$ is simply equal to $Y_{jc\cdot}$; in the global approach, $\hat{Y}_{jcg} = \widehat{(G)}_g + \widehat{(\mathrm{SpC})}_{jc}$, where $\widehat{(G)}_g$ and $\widehat{(\mathrm{SpC})}_{jc}$ are, respectively, the estimated gel and spot × condition effects. The residuals are estimating the random errors E_{jcg}. If the chosen model is correct, then their distribution is nearly the same than the errors distribution. Therefore, structures in the variance of the observations may be detected for example by examining graphics of residuals versus the predicted values, or the position on the gel. If such structures exist, they can be taken into account in the global ANOVA model. Moreover, looking carefully at spots j whose absolute residuals $|R_{jcg}|$ or empirical variances are very high, may reveal problems during the image analysis process, as mismatching for example. It gives the opportunity to correct the data if necessary.

A residual analysis for studying the variability of data coming from example of Section 2 is shown in Fig. 5. The graphic of residuals versus the predicted values shows that the residuals are increasing with the spot volume. The optimal transformation for stabilizing the variance is estimated by the Box–Cox method: we found $T(\%V) = (\%V)^{1/3}$. For that example, we did not find that the data variability was depending on the spots position on the gel.

Other sources of heterogeneity may exist in the data, and the distribution of residuals may be much more spread out than the Gaussian distribution though no particular structure was detected in the variance of the observations. Some authors [47,27] proposed in the context of differential analysis of gene expression, to use bootstrap methods to address the problem of non-Gaussian distribution of the test statistic. Nevertheless it

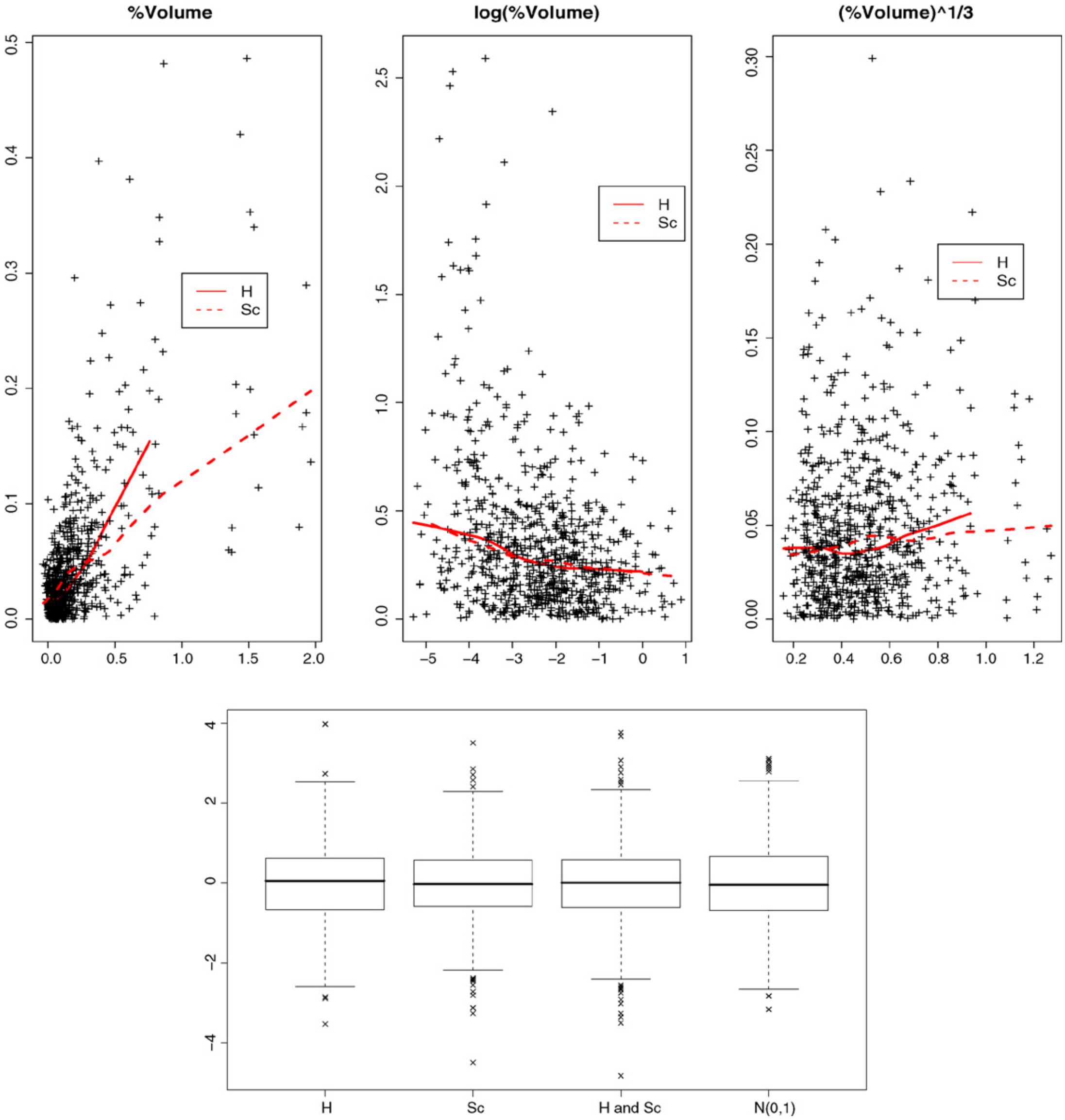

Fig. 5. Graphics for studying the data variability. The two first graphics represent the absolute values of the residuals $R_{jcg} = Y_{jcg} - \hat{Y}_{jcg}$ vs. the predicted values $\hat{Y}_{jcg}$ for the data coming from example of Section 2. Only spots whose volume ratio was greater than 2 were considered in the differential analysis. On the left, the Y_{jcg} are the percentage of volumes on each gel: $Y_{jcg} = \%V_{jcg}$. The lines are smoothed fits of the data, one for each condition. They clearly show that the residuals are increasing with the mean. The logarithmic transformation $Y_{jcg} = \log(\%V_{jcg})$, see the graphic on the middle, inverts the tendency: the smoothed fits of the residuals are decreasing functions of the predicted values. On the right, $Y_{jcg} = (\%V_{jcg})^{1/3}$. This power transformation allows to stabilize the variance of the observations, the smoothed fits being nearly horizontal. The last graphic represents the distribution of the standardized residuals after the power transformation. The standardized residuals should be distributed as independent Gaussian variables with mean 0 and variance 1. The two first boxplots consider the residuals under each condition. They do not show any particular difference between the conditions. The third boxplot represents the distribution of all the residuals, and the last one the distribution of n simulated Gaussian (0,1) variates, where n is the total number of residuals. Looking at these graphics, there is no reason to suspect that the responses $Y_{jcg} = (\%V_{jcg})^{1/3}$ are not Gaussian distributed with the same variance.

should be noted that bootstrap method is not well adapted to the spot by spot approach because of the small number of replications. Moreover applying bootstrap methods for the differential analysis of 2-DE in a global ANOVA model, leads to heavy computation. Indeed, because of missing data, the algorithm for estimating the parameters is time consuming.

3.4. Strategy for missing data

Missing data cannot be ignored in differential analysis of 2-DE, because they affect a large number of spots, and because the lack of observation may be due to proteins variant in abundance.

Several reasons lead to missing observations, for example the actual absence of a given protein, or a mismatching. In some cases, it is possible to guess the reason. For example, when the spot is not observed on any gels within a condition, the protein may be absent. But generally it is hazardous to interpret missing data without a tedious inspection of the data, spot by spot.

The usual testing procedures used for the differential analysis does not need a complete data set, but they need a minimum

Box 3. Effect of the sample size on the estimated variance variability.

Let $X_1, \ldots, X_n$ be n independent Gaussian observations with mean m and variance σ^2. The empirical variance s^2 defined as follows

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X.)^2, \quad \text{where } X. = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

is an unbiased estimator of the variance σ^2. Its coefficient of variation is equal to

$$\mathrm{CV}(s^2) = 100\frac{\text{standard-error}(s^2)}{\text{mean}(s^2)} = 100\sqrt{\frac{2}{n-1}}.$$

It follows immediately that if $n = 3, \mathrm{CV}(s^2) = 100\%$, if $n = 9$, $\mathrm{CV}(s^2) = 50\%$.

Let us now apply these results to the Student test used for testing H_j in the spot by spot approach. For the sake of simplicity, assume that $n_{j1} = n_{j2} = n_j/2$. The denominator of S_j is the square-root of the estimated variance of the difference $Y_{j1}. - Y_{j2}.$. More precisely, the variance of $Y_{j1}. - Y_{j2}.$ is estimated by $S_j^2 = 4s_j^2/n_j$ where s_j^2 is the empirical variance. Using the results given above, we get that the coefficient of variation of S_j^2 equals $100\sqrt{2/(n_j - 2)}$. For example, $n_j = 6$ leads to $\mathrm{CV}(4s_j^2/n_j) = 71\%$, $n_j = 12$ leads to 45%. These simple calculations show the importance of the number of replications in the differential analysis.

Box 4. Spot by spot approach and Student statistic: variations of the *p*-values as function of the estimated standard-error and the number of replications.

Let us consider a differential analysis of 2-DE comparing two conditions, based on a spot by spot approach, where the number of spots *J* is equal to 500. Suppose that

- After the logarithmic transformation of the volume percentages, the responses are Gaussian distributed with the same variance.
- Using a Bonferroni procedure that controls the FWER at 5%, five spots were detected as variant. Precisely, the hypothesis H_j was rejected if the *p*-value was lower than 0.0001.
- Using the procedure controlling the FDR at 5%, we found 39 variant spots. In that case, the hypothesis H_j was rejected if the *p*-value was lower than 0.0039.

Let us consider spots for which the means difference $\delta_{12} = |Y_{j1.} - Y_{j2.}|$ equals log(2). According to our experience when analyzing 2-DE data, the estimated standard-error of these δ_{12} may vary between 0.05 and 1. The table below gives the *p*-values corresponding to $\delta_{12} = \log(2)$ for several values of their estimated standard-errors, denoted S.E. (δ_{12}) and several values of the number of replications. It is assumed that the number of replications is the same under each condition: $n_{j1} = n_{j2} = n_j/2$.

S.E. (δ_{12})	$n_j = 4$	$n_j = 6$	$n_j = 8$	$n_j = 10$	$n_j = 12$
0.05	0.0052	< 0.0001	< 0.0001	< 0.0001	< 0.0001
0.1	0.020	0.00011	< 0.0001	< 0.0001	< 0.0001
0.2	0.074	0.013	0.0027	0.00059	0.00013
0.5	0.30	0.16	0.097	0.059	0.037
1	0.55	0.44	0.36	0.30	0.26

This table highlights that the decision rule is strongly dependent on the variability of the data and the number of replications.

number of observations for each spot, at least one observation for each spot under each condition. If we use the spot by spot approach, at least three observations for each spot for the comparison of two conditions (see Box 1A) are needed. But, as it is shown in Boxes 3 and 4, one should prefer to have at least five or six observations for each spot.

Some authors proposed to set the missing data to the value 0, or to the lowest observed value in the data set [46]. Such a procedure assumes that all missing data are due to lack of protein. Others proposed to replace the missing data for one spot on one gel by the mean of the observations for this spot. More sophisticated methods have been proposed as the *k*-nearest neighbour method [48,49]. Nevertheless they are not adapted to the case where values are missing on all the gels corresponding to one condition.

Another solution is to replace missing data by some simulated values, for example by drawing Gaussian variables with mean m and variance s^2. The values of m and s^2 may be chosen with the help of the data. For example, m is the smallest or one of the smallest observed values as the 0.025 quantile of the data, and s^2 is the median of the empirical variances calculated for each spot. The question is now *how many missing data must be replaced by simulation*? One possibility is to simulate missing data in order to get the minimum number of observations required for the statistical analysis. At the opposite end we could simulate data wherever they are missing. The risk is then to bias the differential analysis by introducing additional information possibly erroneous.

What is a good strategy for missing data in 2-DE analysis is an open question that needs further work.

3.5. Discussion

Whatever the approach chosen, spot by spot or global approach, it is always advantageous to carry out preliminary analyses as described in Section 3.3. The differential analysis of 2-DE gels is an iterative process. The statistical analysis will provide a list of differentially expressed spots in terms of protein abundance, based on a decision rule strongly dependent on the data variability and on the number of replications (see Box 4). This list will be confirmed or rejected by the researcher. If several spots are rejected, it may be worthwhile to suppress these spots from the data and go back to the beginning of the analysis.

One customary practice is to retain among spots detected as variants those that are biologically significant (see [39]). For example, the two-fold change rule is applied: it consists of keeping spots whose volume ratio is greater than 2. Another practice is to do the differential analysis only with spots whose volume ratio is greater than 2. Let us clarify that from a statistical point of view, those rules have no meaning. In practice, spots whose volume ratio is smaller than 2 may be observed with great precision and with a large number of replications, and thus may be detected as variant. Suppression of those spots before the differential analysis may lead to eliminate variant spots. Statistics cannot decide what is biologically pertinent or not, but can propose objective methods based on the data, to suggest both what could be interesting, and what should be moved aside or corrected.

Let us now discuss the choice between the spot by spot and the global approaches.

- The spot by spot analysis does not need sophisticated software, and is proposed by the software packages used for image analysis. It is thus very attractive. Nevertheless, as each test uses information coming from only one spot, a large number of replications is necessary (see Box 4). In Section 3.1 we assumed that the variance of the observations for one spot was identical under both conditions. This assumption could be relaxed and the test statistic adapted to the case where the variance is dependent on the condition. Therefore, the number of replications by condition should be large enough to estimate properly the variance of the test statistic (see Box 3).

  The Mann–Whitney test is attractive because it does not assume Gaussian distribution but it is based on the ranks of the observations rather than on the observations. It lacks power when the number of observations is small [29]: a minimum of seven replications by condition is needed according to [39].

  Whatever the test statistic, it is assumed that for each condition, and spot, the observations are replications. Therefore, the data normalization, as suppressing the gel effect, and more generally the block effects, must be done before the testing procedure. However, it should be noticed that including additional effects reduces degrees of freedom in the Student statistic.
- The global analysis uses information from all the data for testing each hypothesis H_j. The gel effects on the mean response, denoted $(G)_g$ in Section 3.1, are estimated together with the spot × condition effects, denoted $(\mathrm{SpC})_{jc}$. The variance has been assumed the same for all spots, but this assumption may be weakened by taking into account information on the variance structure. For example, the variance may depend on the condition, or on the spot localization on the image, or on the spot. Because of missing data, a statistical software, such as R (cran.r-project.org) or SAS (www.sas.com) is needed.

Let us finally underline that detecting *significant* differences in protein abundance relies on a statistical procedure that compares the differences of observed spot volumes to their variability. Therefore, the experimental design must guarantee the possibility to estimate properly this variability. Variability in the data may come from the biological and technical phases. Replications in the biological phase may be difficult to obtain in some situations, as for example when sample are taken on people or animals. In the technical phase, three or four replications in most proteomics studies should be possible. The statistician has to take into account these situations, to propose suitable statistical methods, as for example methods based on global ANOVA models, and to precise the limits in which the results can be handled.

4. Conclusion

Accurate differential analysis of proteomic data outcomes of rigorously designed experiments and produces reliable results. This dynamic interaction requires a close interdisciplinary collaboration at every step of the project and is beneficial for both biologists and statisticians. Further investigations using the results issued from such a collaboration can be considered with increased confidence. Statistical tools such as discriminate analysis, regression methods or supervised classification [50–55] can be further applied to accurately discriminate the status of unknown samples, normal or pathologic for instance. The interaction schema between statisticians and biologists is particularly important for the detection of differentially expressed proteins involved in pathologies since it can lead to the discovery of biomarker candidates. Another field of collaboration between both disciplines is the search for functional molecular (proteins only or proteins and mRNAs, etc.) networks. The aim of this approach is to establish the relationships existing between the different cellular actors in order to (re)-construct a causality network. Statistical methods in this field are under development and numerous fundamental mathematical researches are actively in progress [56–59]. It should be emphasized that the interactions between mathematicians, statisticians and biologists are not limited for providing increased confidence in biological results; they allow the delineation of new areas where collaborative research is needed.

Acknowledgements

A part of this work was supported by French GIS Prion. We warmly thank Dr. Patty Chen for improving the English version, Dr. Jeanne Grosclaude and Dr. Raymonde Joubert–Caron for critical reading of this manuscript.

References

[1] V.C. Wasinger, S.J. Cordwell, A. Cerpa-Poljak, J.X. Yan, A.A. Gooley, M.R. Wilkins, M.W. Duncan, R. Harris, K.L. Williams, I. Humphery-Smith, Electrophoresis 16 (1995) 1090.

[2] M.R. Wilkins, J.C. Sanchez, A.A. Gooley, R.D. Appel, I. Humphery-Smith, D.F. Hochstrasser, K.L. Williams, Biotechnol. Genet. Eng. Rev. 13 (1996) 19.

[3] S. Hanash, Drug Discovery Today 7 (2002) 797.
[4] M.R. Wilkins, R.D. Appel, J.E. Van Eyk, M.C. Chung, A. Gorg, M. Hecker, L.A. Huber, H. Langen, A.J. Link, Y.K. Paik, S.D. Patterson, S.R. Pennington, T. Rabilloud, R.J. Simpson, W. Weiss, M.J. Dunn, Proteomics 6 (2006) 4.
[5] L.S. Riter, O. Vitek, K.M. Gooding, B.D. Hodge, R.K. Julian, J. Mass Spectrom. 40 (2005) 565.
[6] J. Hu, K.R. Coombes, J.S. Morris, K.A. Baggerly, Brief. Funct. Genomics Proteomics 3 (2005) 322.
[7] D.R. Cox, Planning of Experiments, John Wiley, 1958.
[8] A. Dean, D. Voss, Design and Analysis of Experiments, Springer, New York, 1999.
[9] G.A. Churchill, Nat. Genet. Suppl. 32 (2002) 490.
[10] M.K. Kerr, Biometrics 59 (2003) 822.
[11] Y.H. Yang, T. Speed, Nat. Rev. Genet. 3 (2002) 579.
[12] P.J. Bosque, S.B. Prusiner, J. Virol. 74 (2000) 4377.
[13] P.-C. Klöhn, L. Stoltze, E. Flechsig, M. Enari, C. Weissmann, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 11666.
[14] J.-F. Chich, B. Schaeffer, A.-P. Bouin, F. Mouthon, V. Labas, C. Larramendy, J.-P. Deslys, J. Grosclaude, in press.
[15] P.L. Mellon, J.J. Windle, P.C. Goldsmith, C.A. Padula, J.L. Roberts, R.I. Weiner, Neuron 5 (1990) 1.
[16] H.M. Schatzl, L. Laszlo, D.M. Holtzman, J. Tatzelt, S.J. DeArmond, R.I. Weiner, W.C. Mobley, S.B. Prusiner, J. Virol. 71 (1997) 8821.
[17] C.J. Brien, Biometrics 39 (1983) 53.
[18] C.J. Brien, R.A. Bailey, J. R. Stat. Soc.: Ser. B 68 (2006) 571.
[19] M.K. Kerr, G.A. Churchill, Biostatistics 2 (2001) 183.
[20] N.A. Karp, K.S. Lilley, Proteomics 5 (2005) 3105.
[21] N. Bahrman, J. Le Gouis, L. Negroni, L. Amilhat, P. Leroy, A.-L. Lainé, O. Jaminon, Proteomics 4 (2004) 709.
[22] L.H. Choe, K.H. Lee, Electrophoresis 24 (2003) 3500.
[23] S. Cronier, H. Laude, J. Peyrin, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 12271.
[24] M.P. Molloy, E.E. Brzezinski, J. Hang, M.T. McDowell, R.A. Van Bogelen, Proteomics 3 (2003) 1912.
[25] P. Baldi, A. Long, Bioinformatics 17 (2001) 509.
[26] S. Dudoit, Y.H. Yang, M.J. Callow, T.P. Speed, J. Am. Stat. Assoc. 74 (2002) 829.
[27] S. Dudoit, J.P. Shaffer, J.C. Boldrick, Stat. Sci. 18 (2003) 71.
[28] A. Reiner, D. Yekutieli, Y. Benjamini, Bioinformatics 19 (2003) 368.
[29] S. Wang, S. Ethier, Bioinformatics 20 (2004) 100.
[30] J. Aubert, A. Bar-Hen, J.-J. Daudin, S. Robin, BMC Bioinformatics 5 (2004), article 125.
[31] J.S. Gustafsson, A. Blomberg, M. Rudemo, Electrophoresis 23 (2002) 1731.
[32] M. Rogers, J. Graham, R.P. Tonge, Proteomics 3 (2003) 879.
[33] M. Rogers, J. Graham, R.P. Tonge, Proteomics 3 (2003) 887.
[34] A.W. Dowsey, M.J. Dunn, G.-Z. Yang, Proteomics 3 (2003) 1567.
[35] B.-F. Liu, Y. Sera, N. Matsubara, K. Otsuka, S. Terabe, Electrophoresis 24 (2003) 3260.
[36] A. Roy, F. Seillier-Moiseiwitsch, K.R. Lee, Y. Hang, M. Marten, B. Raman, Chance 16 (2003) 13.
[37] J. Chang, H. van Remmen, W.F. Ward, F.E. Regnier, A. Richardson, J. Cornell, J. Proteome Res. 3 (2004) 1210.
[38] J.S. Gustafsson, R. Ceasar, C.A. Glasbey, A. Blomberg, M. Rudemo, Proteomics 4 (2004) 3791.
[39] B. Meunier, J. Bouley, I. Piec, C. Bernard, B. Picard, J.F. Hocquette, Anal. Biochem. 340 (2005) 226.
[40] S. Dudoit, M.J. van der Laan, K.S. Pollard, Stat. Appl. Genet. Mol. Biol. 3 (2004), article 13.
[41] Y. Benjamini, Y. Hochberg, J. R. Stat. Soc. B 57 (1995) 289.
[42] J.D. Storey, R. Tibshirani, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 9440.
[43] J. Malone, K. McGarry, C. Bowerman, in: Sixth EPSRC PREP Conference, 2004.
[44] B.P. Durbin, J.S. Hardin, D.M. Hawkins, D.M. Rocke, Bioinformatics 18 (2002) S105.
[45] W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Bioinformatics 18 (2002) S96.
[46] J.S. Almeida, R. Stanislaus, E. Krug, J.M. Arthur, Proteomics 5 (2005) 1242.
[47] M.K. Kerr, M. Martin, G.A. Churchill, J. Comput. Biol. 7 (2000) 819.
[48] K. Jung, A. Gannoun, B. Sitek, H.E. Meyer, K. Stuhler, W. Urfer, REVSTAT Stat. J. 3 (2005) 99.
[49] K. Jung, A. Gannoun, B. Sitek, O. Apostolov, A. Schramm, H.E. Meyer, K. Stuhler, W. Urfer, REVSTAT Stat. J. 4 (2006) 67.
[50] A.L. Boulesteix, G. Tutz, K. Strimmer, Bioinformatics 19 (2003) 2465.
[51] M. Dettling, P. Bühlmann, J. Multivar. Anal. 90 (2004) 106.
[52] T. Hastie, R. Tibshirani, Biostatistics 5 (2004) 329.
[53] J. Zhu, T. Hastie, Biostatistics 5 (2004) 427.
[54] A.L. Boulesteix, G. Tutz, Comput. Stat. Data Anal. 50 (2006) 783.
[55] J.J. Dai, L. Lieu, D. Rocke, Stat. Appl. Genet. Mol. Biol. 5 (2006), article 6.
[56] J. Schäfer, K. Strimmer, Bioinformatics 21 (2005) 754.
[57] A. Wille, P. Bühlmann, Stat. Appl. Genet. Mol. Biol. 5 (2006) 1.
[58] H. Li, J. Gui, Biostatistics 7 (2006) 302.
[59] N. Meinshausen, P. Bühlmann, Ann. Stat. 34 (2006) 1436.