research articles Journal of proteome research

Alternative Experimental Design with an Applied Normalization Scheme Can Improve Statistical Power in 2D-DIGE Experiments

Kristof Engelen,[†,‡] Alejandro Sifrim,[†,‡] Babs Van de Plas,[§] Kris Laukens,[||] Lutgarde Arckens,[§] and Kathleen Marchal*[,‡]

Department of Microbial and Molecular Systems, Katholieke Universiteit Leuven, Kasteelpark Arenberg 20, 3001 Leuven, Belgium, Research group of Neuroplasticity and Neuroproteomics, Katholieke Universiteit Leuven, Naamsestraat 59, 3000 Leuven, Belgium, and Intelligent Systems Laboratory, University of Antwerp, Middelheimlaan 1 B, 2020 Antwerp, Belgium

Received January 6, 2010

2D-DIGE experiments are a high-throughput technique for measuring protein abundances based on gel separation. Traditionally three samples are multiplexed per gel: two biological test samples and a third internal standard sample consisting of a pool of all test samples. We demonstrate that the use of an internal standard helps to account for technical variation caused by spatial intensity biases that exist in the gels and propose a novel data-preprocessing technique, a spatial intensity bias removal (SIBR), which can approximate these biases using only the data of biological replicates loaded on the gel. Using this technique, we show that by replacing the internal standard with additional biological replicates, a significant increase in statistical power can be achieved compared to traditional 2D-DIGE designs. This boost in statistical power can be used to reduce the false positive rate for identifying differential protein abundances without compromising sensitivity, or to improve sensitivity without compromising false positive rate. A software implementation of SIBR can be downloaded at http://ibiza.biw.kuleuven.be/SIBR.

Keywords: bioinformatics • difference gel electrophoresis • proteomics • experimental design • normalization • spatial bias • statistical power

Introduction

High-throughput measuring of differences in protein abundance is customarily done by either liquid chromatography or gel-based techniques. Both categories are complementary in getting a global picture of protein expression[1−4] as they differ in the proteins that can be separated. Two-dimensional difference gel electrophoresis (2D-DIGE)[5] can be considered the gold standard of gel-based techniques. Just as with classical 2D gels, proteins are separated in two dimensions, by isoelectric point (p*I*) and by molecular weight.[6] In contrast to classical 2D gels, however, DIGE makes use of multiplexing samples from different conditions by labeling them with distinct cyanine dyes. Usually test samples are labeled with Cy5 and Cy3, while Cy2 is reserved for a third sample consisting of a pool of all test samples. This pool serves as an internal standard, used to normalize for experimental differences between gels, and in doing so to reduce technical variance, and also serves to facilitate spot matching across gel images.[7,8] The downside of using an internal standard is that one-third of the loaded samples are of no biological interest to the researcher. Nevertheless, once the spots have been quantified and matched, the internal standard is considered to be an important means to remove systematic biases in the data. These biases confound the true biological responses of interest and are the direct consequence of certain experimental factors such as the fluorescent dye used or the location of the spot on the gel. The microarray community has already developed a multitude of methods, which are directly implementable in DIGE experiments.[9] Most of these methods focus on the removal of intensity-dependent and/or spatial dye-effects,[10,11] but little research has been done on the spatial bias caused by the inherent heterogeneity of the gels. In this paper, we show that this kind of spatial bias exists and is generally handled by normalizing against the internal standard that is traditionally incorporated in the experiment design. We further propose a different way of normalizing for these spatial effects using only sample measurements, a normalization strategy called spatial intensity bias removal (SIBR). SIBR does not require an internal standard to be incorporated in the design, so that the Cy2 channel can be allocated for additional biological replicates (Figure 1). We show that such an alternative experimental design, in combination with SIBR normalization to account for

* To whom correspondence should be addressed. Centre for Microbial and Plant Genetics, Katholieke Universiteit Leuven, Kasteelpark Arenberg 20, 3001 Leuven, Belgium, e-mail: kathleen.marchal@biw.kuleuven.be. Tel.: +32 16329685.

† These authors contributed equally to this work.

‡ Department of Microbial and Molecular Systems, Katholieke Universiteit Leuven.

§ Research group of Neuroplasticity and Neuroproteomics, Katholieke Universiteit Leuven.

|| University of Antwerp.

10.1021/pr100010u

Published on Web 08/03/2010

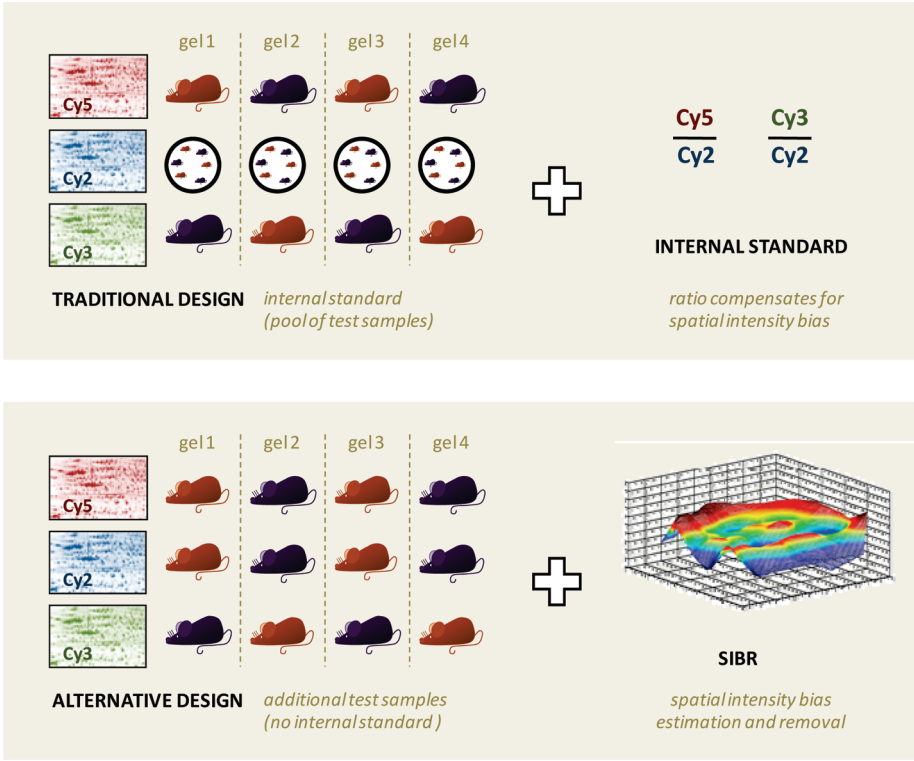

Figure 1. Alternative experimental design with an applied normalization. The approach is illustrated for a simple case where two distinct biological conditions (represented by the purple and orange mice) are compared. The traditional design uses the Cy2 channels to allocate an internal standard (represented by the circles encompassing a pool of purple and orange mice) and ratios of sample over standard compensate for spatial intensity biases inherent to the gel (top panel). The alternative design makes no use of an internal standard and relies on a spatial intensity bias removal (SIBR) normalization to remove spatial effects inherent to the gel (bottom panel). In this paper, we show that the combination of an alternative experimental design and SIBR normalization leads to a higher statistical power for identifying differential expression using the same number of gels.

spatial intensity bias, can achieve a higher statistical power for identifying differential expression using the same number of gels.

Material and Methods

2D-DIGE Data Sets. To evaluate and quantify spatial intensity biases we used the gel images of 6 independently performed experiments using the standard 2D-DIGE design as proposed by Alban et al. (2003): 3 on *Mus musculus*[12,13] (respectively 5, 8, and 8 gels), 1 on *Felis catus* (23 gels), 1 on *Apis mellifera* (15 gels), and 1 on *Salmonella typhimurium*[14] (18 gels). These experiments were conducted by four different researchers from two different laboratories. An additional experiment (*Mus* musculus, 8 gels) was set up to show that applying a SIBR-based normalization scheme on an experimental design where the internal standard was omitted in favor of additional biological replicates leads to an increase in power for detecting differential expression when compared to the traditional design. Four gels of this experiment used the traditional design (with internal standard) and 4 gels used the alternative design (without internal standard). This experiment is referred to in the text as the validation experiment. Details regarding all animals and sample preparation can be found in Supplementary File 1 (Supporting Information).

After electrophoresis, gel images were acquired using an Ettan DIGE Imager and/or a Typhoon 9400 scanner (both from GE Healthcare) following the standard protocols. The Typhoon 9400 (laser scanner) can achieve a larger dynamic range due to the adaptability of the gain on the photomultiplier tube. The Ettan DIGE (CCD scanner) on the other hand is more versatile with respect to illumination wavelength and can accommodate a broader range of dyes.[15]

Image Analysis. The resulting three 16-bit images per gel (one for each channel) were subsequently analyzed to match spots across gel images (a necessary step as gel specific separation patterns are subject to geometrical distortions[16]) and to obtain a per-spot quantification measure of protein abundance. Several studies[17–20] already demonstrated that the end results of proteomics studies can be affected by the used analysis strategy. To obtain fully matched spots, necessary for the application of the new data processing method, we chose a "warp first" workflow and used the Delta2D 3.6 software package (DECODON). Delta2D first aligns the gel images (i.e., defining warp transformation vector) and fuses them into a synthetic "master" image.[21] This fused image contains all the spots present on all initial gels. Spot detection is subsequently performed on the fusion image and detected spots are transferred back to the original images using the identified warp transformation vectors. The advantage of this "warp first" strategy detecting spots on the fused image compared or a "spot detection first" workflow detecting spots directly on the original images (e.g., DeCyder of GE Healthcare) is that spots are 100% matched and no missing values exist. Indeed, if a given biological sample does not contain a particular protein, the spot is still quantified using the background signal of the corresponding spot position.[22] Images were warped using the in-gel standard warping strategy of Delta2D. Analysis of the images for the gels containing the pooled standard ($n = 4$)

was performed independently from the images containing no such standards ($n = 4$). Automatically generated image transformation vectors were verified and additional vectors were added manually where needed. Fusion images for each treatment group were generated using the "average" mode to average out noise. These fusion images were subsequently fused using the "union" mode so all spots of the original images were present in the global fusion image. Spot detection was based on the default generated parameters within Delta2D and was manually corrected if spots were missed by the algorithm or to remove unwanted artifacts. Quantified spot data were then further analyzed in MATLAB R2009a (The Mathworks).

Data Preprocessing. Values that were originally subtracted from the measured volumes by Delta2D were added again to the unnormalized intensity volumes to obtain the original, full spot volumes (i.e., no background correction was performed). This was done because background correction can introduce a bias,[20] which was clearly seen in our data sets (Supplementary File 2, Supporting Information). The data was log-transformed to stabilize multiplicative error variance.[23,24]

To counter dye biases and the loading of unequal sample quantities, all samples were normalized against an artificial reference with a nonlinear intensity-dependent rescaling (NLIDR). This is different from a linear rescaling as done by common software packages, such as, for example, DeCyder. In that case, the log-ratio distributions of the samples (Cy3 and Cy5) over the internal standard (Cy2) are rescaled with a constant factor (i.e., linear). Such a linear rescaling assumes that the dye responses differ by a constant factor and that the expression of the majority of the proteins remains unchanged. But as cyanine dyes have intensity-dependent (nonlinear) responses, normalization techniques such as nonlinear local regression or moving average-based methods yield better results as they remove intensity dependent bias due to the differential labeling rather than a biological difference. This is especially the case when working with the Cy2 dye, which shows a weaker response compared to the Cy3 and the Cy5 dyes.[25] NLIDR was performed as follows: an artificial reference was generated for each spot using the median over all samples. To counter intensity-dependent systematic variation, every sample was rescaled in a nonlinear way with respect to the artificial reference: $M = \log((x_{i,c})/(\mathrm{ref}_i))$ with $x_{i,c}$ being the intensity volume for the ith spot and cth treatment condition and $\mathrm{ref}_i = \mathrm{median}(x_i)$ the median over all measurements of that spot, was fitted against $A = \log(\mathrm{ref}_i)$. This was done with the smooth function in MATLAB using a robust loess nonlinear local regression algorithm. A span that comprised 60% of the spots was chosen as to not overfit the data. Corrected values ($x'_{i,c}$) were obtained by adding the artificial reference values to the residuals of the fit ($\hat{x}_{i,c}$): $\log(x'_{i,c}) = \log(\hat{x}_{i,c}) + \log(\mathrm{ref}_i)$. When appropriate, correction with an internal standard was calculated as follows: $\log(r_{i,c}) = \log((x'_{i,c})/(x'_{i,s} - \bar{x}'_{i,s}))$, $r_{i,c}$ being the standard-corrected intensity volume for the ith spot and the cth treatment condition, $x'_{i,c}$being the NLIDR intensity volume for that spot, $x'_{i,s}$ the normalized volume of the standard for that spot on the same gel, $\bar{x}'_{i,s}$ the mean of the NLIDR volumes over all measurements of the internal standard for spot i. This correction with the internal standard is analogous to the common-practice normalization as done by popular software packages such as DeCyder. In that case the log-ratios of Cy3/Cy2 and Cy5/Cy2 are linearly rescaled with a constant factor to have a mean of zero, a procedure sometimes referred to as "ratio metric normalization". The only difference with our approach is that, instead of a linear rescaling, we use a nonlinear correction method to account for nonlinear intensity-dependent dye differences, which in theory can only improve the quality of the results (Supplementary File 3, Supporting Information). To verify that NLIDR was indeed performing equally well or better than a linear rescaling approach we compared the difference in variance within replicates. Although we could not see a very pronounced improvement by using a nonlinear model we could also see no signs of decline of performance in any of our data sets (see Figure 3.2 in Supplementary File 3, Supporting Information). We included the NLIDR approach in our software package as intensity dependent dye effects might be more pronounced in other data sets.

Spatial Intensity Bias Removal. To account for regional trends, caused by heterogeneities in the gels, we developed a spatial intensity bias removal (SIBR) procedure. For every intensity volume measurement of each biological replicate the following difference $\log(d'_{i,c}) = \log((x'_{i,c})/(\bar{x}_{i,c}))$ was calculated, $x'_{i,c}$ being the normalized volume for the ith spot of the cth treatment condition and $\bar{x}'_{i,c}$ being the mean of the same values over all samples for the same treatment condition. This difference was then fitted using a bivariate robust local regression function as implemented in the curve fitting toolbox of MATLAB. Robustness against outliers was achieved using the bisquare weight method. For the predictor variables, the x- and y-coordinates were used as exported from Delta2D. The span, which controls the degree of smoothing, was set to 10% of all data points. A large enough span assures that a minimal amount of spot-dependent biological variation is removed and that mostly biologically independent spots are considered. Randomizations were made by randomly assigning existing coordinate pairs to normalized intensity volumes and running the algorithm with the same parameters. In this way, no changes were made to the local distribution of spots on the gel.

Statistical Power. To quantify the acquired statistical power, the standard deviation per condition for each spot was calculated after normalization. Then the 75% percentile of these standard deviations was taken as a global measure of variance present in the replicates after normalization. Using this variance, the statistical power was calculated using the SCal4Poolings tool available online.[26,27] We calculated the power for a 2-fold change with $\alpha = 0.001$ and the appropriate number of replicates. A 1.5-fold change with $\alpha = 0.01$ was also calculated.

Results

Gels Exhibit Spatial Intensity Biases. To gain insight in the spatial nature of intensity variation within a gel, we first evaluated internal standard measurements (Cy2 channel). These can be viewed as technical replicates and should only differ from each other because of experimental sources of variation.

We first removed systematic variation caused by differences in dye response or unequal loading on the gel by means of a nonlinear intensity dependent rescaling (NLIDR; see Materials and Methods for details). This NLIDR is based on a loess normalization technique,[28] which is commonly used in microarray experiments and has been used before in DIGE data analysis.[11] After this normalization step, taking into account dye-related sources of variation, considerable differences in the measurements of the internal standard can still be observed. Although this remaining variation can be caused by random

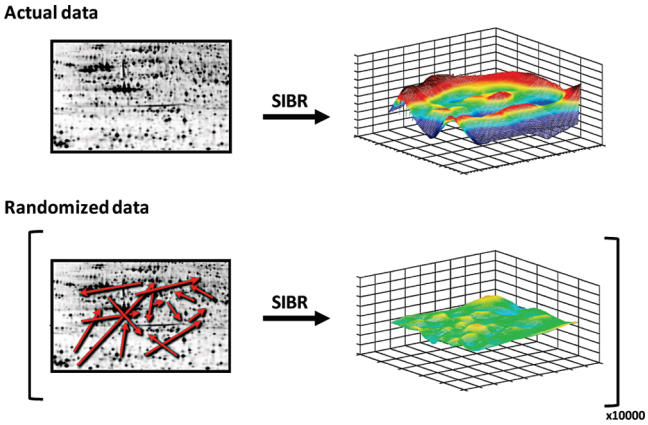

Figure 2. Confirming the existence of spatial intensity trends. Left hand plots represent the gel images, right-hand plots represent the spatial intensity biases as estimated by the spatial intensity bias removal (SIBR). By permutating spot coordinates of actual gel data multiple times and evaluating the magnitude of estimated spatial trends in these randomized data, we rule out the possibility that the SIBR estimated spatial intensity bias observed in the actual data is due to random sampling.

measurement errors purely dependent on the spot that is being observed, the presence of a systematic trend due to an unaccounted experimental factor cannot be excluded. To find out whether this remaining variation is due to random measurement errors or whether a remaining systematic trend was still present in the data, we applied a spatial intensity bias removal (SIBR) procedure to the data.

The SIBR method fits a smoothed surface using a robust two-dimensional nonlinear regression algorithm on the differences between the measurements on a given gel for the internal standard with the mean values over all standard measurements for each spot. The resulting SIBR fits on the measurements of the internal standard show to what extent the NLIDR intensity volumes in a given region of the gel deviate from their respective means. If for some reason a region of the gel exhibits consistently higher or lower values, the fitted surface will deviate from 0. The SIBR planes clearly illustrate that such spatial trends are present in the internal standard data. The lack of persistent patterns that can be distinguished in the spatial trends across the different gels suggests that these trends are related to heterogeneities in the gel matrix, rather than random measurement errors.

To ensure that these results indeed reflect real spatial patterns, we compared fits obtained on the original internal standard data with those obtained on randomly generated data sets (obtained by permuting the xy-coordinates of the spot locations in the original data set 10 000 times). The magnitude of the fits obtained by using SIBR on the permutations is represented in Figure 2. By evaluating the distribution of the standard deviations of the fitted surfaces for the permutations, we estimate the chance that the original data belongs to the same distribution. We can clearly see that the standard deviations of the fits obtained for the original data are substantially larger than the ones obtained for the permutated data (data not shown), in fact all standard deviations of the original data sets show a p-value smaller than 10^{-16} (calculated in MATLAB 2009a and reported as 0, that is, below its computational precision). This confirms that the trends seen here are not caused by random chance, but can be attributed to the existence of gel-dependent spatial trends that manifest themselves at least in internal standard measurements.

Assuming that these spatial trends are not only present in the Cy2 channel, but also in the Cy3 and Cy5 channels, that is, the biological sample measurements, we applied the SIBR algorithm on the measurements of the biological replicates. We assume that neighboring spots of a biological sample have equal chance of having lower or higher values compared to their mean, so that if no spatial bias exists these differences are averaged out due to the symmetrical distribution compared to their respective means. Under this assumption, neighboring spots showing consistently lower or higher values compared to their mean then point toward a spatial trend. Taking into account additional biological variance, we defined the size of the gel regions sufficiently large so that mostly biologically independent spots were considered within a region. By doing this we minimize the possible bias that could arise by incorporating spot chains into our calculation of the spatial effects. Spot chains are mostly made up of isoforms of a protein species and are thus spots likely to be correlated in protein abundance. This correlation between spots in a region would lead to a false estimate of a spatial effect. By looking at a large mass of data points we dilute such correlations in the estimate of the spatial effect (for detailed representations, see Supplementary File 4, Supporting Information).

If the spatial trends are inherent to the gel we can also safely assume that the same trends should occur on the three channels. Using multiple channels further averages out the biological differences and should result in a clearer picture of the spatial trends caused by technical variation in the gels.

To confirm that there was a spatial bias in the measurements of the biological samples, we compared the fits obtained by using only the biological measurements (2 measurements per spot from the Cy3 and Cy5 channel in the normal data sets, 3 measurements from the Cy3, Cy5 and Cy2 channel in the validation experiment) to 10 000 permutations using permutated coordinates. As with the internal standard measurements, the original data sets resulted in spatial trends that were nowhere near the ones observed in the randomized data sets (Table 1).

Internal Standard Is Not Necessary to Control Spatial Intensity Bias. To evaluate whether these spatial trends reflected gel-specific effects that were observed in all channels of a gel we tested whether the special biases observed in the channels containing biological measurements were correlated to those observed in the channel with the internal standard (Table 2). Results showed a relatively high degree of correlation, confirming that the spatial trends are indeed observable in all channels, and that they can be equally well approximated by the measurements obtained from the technical replicates of the internal standard than by those obtained from biological replicates.

As normalization procedures are meant to remove consistent sources of variation from the data, we assessed the efficiency of our normalization schemes by evaluating to what extent consistency among replicates increased compared to the raw unprocessed data. The quality of these improvements was measured by the differences in standard deviations ($\sigma_{raw} - \sigma_{normalized}$) across replicates within treatment conditions for several independent studies (see Material and Methods). Figure 3 shows some of these results for only NLIDR normalization (representing a strategy were only dye discrepancies were normalized for, ignoring any spatial intensity bias), NLIDR + SIBR (normalization for both dye discrepancies and spatial intensity bias, but without using any of the internal standard

Table 1. Spatial Intensity Bias Removal (SIBR) Using Biological Replicates: Spatial Intensity Bias Determined on Real Data vs XY-Coordinate Permutations[a]

Experiments	Permutations μ	Permutations σ	SD $100*(1-10^{-16})$ th percentile	Real SIBR surface
Apis mellifera (CCD)	0.0594	0.0034	0.0875	0.2038
Felis catus (CCD)	0.0473	0.0018	0.0622	0.1217
Mus musculus 1 (CCD)	0.0300	0.0015	0.0424	0.1335
Mus musculus 2 (CCD)	0.0362	0.0026	0.0577	0.1475
Mus musculus 3 (Typhoon)	0.0222	0.0012	0.0321	0.0802
Mus musculus 3 (CCD)	0.0396	0.0020	0.0561	0.1113
Salmonella typhimurium (Typhoon)	0.0346	0.0014	0.0462	0.1439
Validation Experiment: Traditional Design (CCD)	0.0328	0.0025	0.0535	0.1565
Validation Experiment: Traditional Design (Typhoon)	0.0187	0.0014	0.0303	0.1010

[a] μ represents the mean of the standard deviations of the SIBR surfaces of 10 000 permutations for that experiment. σ represents the standard deviation on μ. The third column gives the $(100*(1-10^{-16})$ th percentile of the distribution of standard deviations of all the permutations. The fourth column is the standard deviation of the SIBR surfaces for the real data.

Table 2. Correlation Analysis of Spatial Biases Estimated by the Spatial Intensity Bias Removal Normalization (SIBR) Using Internal Standard Measurements and SIBR Using Only Biological Replicate Measurements[a]

experiments	μ_{corr}	σ_{corr}
Apis mellifera	0.677	0.138
Felis catus	0.676	0.122
Mus musculus 1	0.730	0.375
Mus musculus 2	0.462	0.403
Mus musculus 3 (CCD)	0.377	0.204
Mus musculus 3 (Typhoon)	0.775	0.094
Salmonella typhimurium	0.874	0.095
Validation Experiment: Traditional Design (CCD)	0.808	0.051
Validation Experiment: Traditional Design (Typhoon)	0.777	0.192

[a] μ_{corr} is the mean over all gels of the correlation coefficients between the two fitted surfaces, σ_{corr} is the standard deviation on that mean.

measurements) and NLIDR log-ratios (normalization for both dye discrepancies and spatial intensity bias, the latter by use of the internal standard; this is analogue to classical ratio metric normalization as done in popular software packages such as DeCyder). In this figure, values smaller than zero represent replicate measurements that showed lower error variance for the normalized than the raw data, whereas values bigger than zero represent an improvement of normalized compared to raw data.

Results on all data sets confirmed the reproducibility of the approach. All normalization schemes improved consistency among replicates compared to the consistency observed for the raw data and normalization with the internal standard showed the most prominent improvement (e.g., the *Salmonella* data set). Our results indicate that differences that are measured between the internal standards are most optimal to estimate gel-specific trends. Internal standards function as a control on a spot-specific level, whereas the SIBR algorithm estimates spatial bias for a whole region and thus has a lower resolution of the real underlying spatial deformities. Some data sets (e.g., *Mus musculus* and *Felis catus*) show little difference in between the normalization with the internal standard (NLIDR normalized log-ratios) and the SIBR normalized volumes (NLIDR + SIBR). This indicates that in these data sets the internal standard does not seem to be handling additional sources of technical variance aside from the spatial effects that can also be successfully corrected for with SIBR, resulting in a similar degree of improvement for both techniques.

Increased Statistical Power by SIBR and Internal Standard Exclusion. So while the proposed SIBR normalization does not outperform the use of an internal standard for normalizing spatial intensity biases, it does sport a different important advantage. As SIBR negates the requirement of including internal standard measurements, additional biological samples can be profiled with the same number of gels. More replicates yield an increased statistical power for the detection of differential expression. To be able to detect a differential effect of a certain size, sufficient statistical power needs to be obtained. The acquired power is dependent on the effect size, the statistical significance level (the degree of false positives one expects to observe), the total (biological and technical) variance of the experiment, and the number of (biological and technical) replicates of each observed sample. A higher power means that smaller effects in differential protein abundances can be determined with the same degree of false negatives or that the same effect size can be observed with a higher confidence. If insufficient power is acquired, biologically relevant changes in protein expression might be missed. When choosing whether to use biological or technical replicates in the experimental design, the general consensus goes to biological replicates, except in cases where the technical variation is exceptionally high.[29–31] Reducing the biological variance can also be achieved by pooling the samples[32] but a large number of specimens are needed to observe a substantial increase in power.

In this section we investigate whether the increase in statistical power associated with additional biological replicates, possible with the same number of gels, justifies the omission of the internal standard. To this end, we compare results from two experiments, both consisting of 4 gels and both comparing two biological conditions (see Supplementary Figure 1, Supporting Information), but one contained an internal standard (traditional design) while the other contained extra biological replicates (alternative design).

Upon comparison of the variances obtained with the 4 gels with internal standards compared to the 4 gels without (using respectively the internal standard measurements or the SIBR normalization algorithm to compensate for spatial intensity bias), we observed a mild increase in variance in the latter. Due to the higher replicate number ($n = 4$ vs $n = 6$) this increase in variance did not lead to lower statistical power. A 25% increase in statistical power (2-fold change, $\alpha = 0.001$) was even

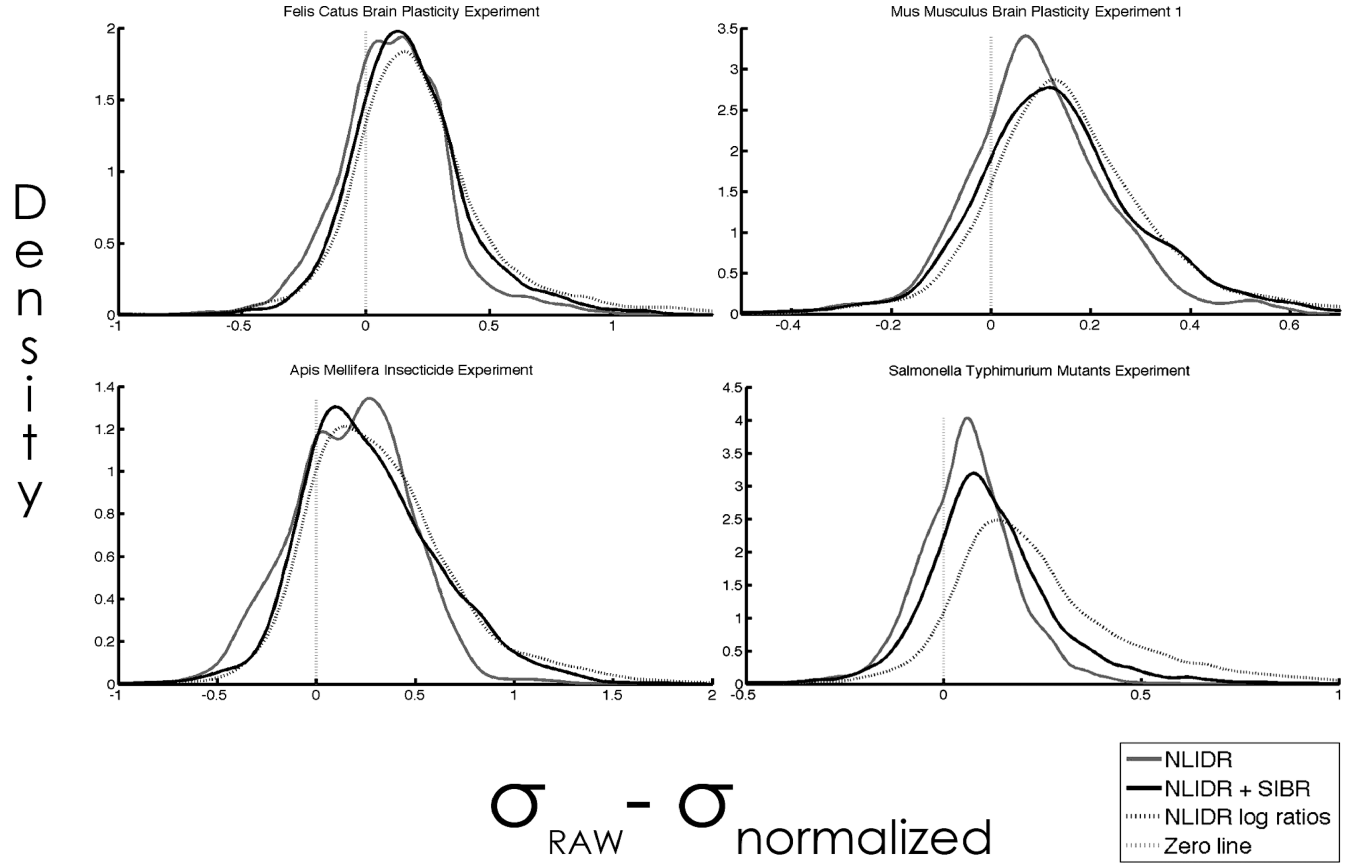

Figure 3. Standard deviation improvements after normalization: These density plots show the improvements in within-treatment-group standard deviations as compared to raw data. The different lines indicate different normalization schemes: nonlinear intensity dependent rescaling only (NLIDR), NLIDR plus spatial intensity bias removal (NLIDR + SIBR), and normalization with the internal standard after NLIDR (NLIDR log-ratios) The more the distribution is "shifted to the right" (bigger surfaces under the curve for standard deviation differences larger than zero), the more improvement can be seen from the normalization, which translates to a lower technical variance resulting in a higher statistical power. Additional plots for the remaining data sets can be found in the Supporting Information (Supplementary File 5).

Table 3. Power Comparison Using Traditional vs Alternative Design[a]

	traditional design (CCD)	alternative design (CCD)	% increase
Variance (75% percentile)	0.11	0.134	
Power (2-fold, $\alpha = 0.001$)	0.341	0.56	64%
Power (1.5-fold, $\alpha = 0.01$)	0.131	0.222	69%
	traditional design (Typhoon)	alternative design (Typhoon)	% increase
Variance (75% percentile)	0.021	0.037	
Power (2-fold, $\alpha = 0.001$)	0.709	0.892	25%
Power (1.5-fold, $\alpha = 0.01$)	0.735	0.801	8%

[a] Power was calculated for different fold-changes and significance levels. We used a measure of variance which encompassed 75% of the spots and calculated 1.5- and 2-fold change with a confidence of respectively 0.01 and 0.001.

observed compared to the traditional experimental design (Table 3). We note that one of the gels of the alternative design set contained several heavily distorted gel regions contributing to a larger variance of this data set. If all gels had been equally undistorted an even more marked increase in power would have been achieved.

Discussion

Gel-based proteomics has come a long way since inception of two-dimensional poly acrylamide gel electrophoresis in the 1970s. Critical improvements in labeling techniques allowed sample multiplexing and paved the way for difference gel electrophoresis (DIGE) of gel-based proteomics. With new developments came new experimental designs and the internal standard design as proposed by Alban et al.[7] seemed a straightforward choice. "Warp first" workflows are emerging in commercial image analysis software nowadays, allowing for full spot matching and thus providing the user with complete data sets (i.e., without missing values). Although these techniques are prone to spot amalgamation in regions of variable spot resolution and can lead to erroneous measurements for the amalgamated species of spots,[33] the resulting complete data sets have some important advantages as well. They are perfectly suited for multivariate statistical techniques, such as, for example, principal component analysis, that can offer a high level view on the data and result in novel interpretations, but they also open doors for new data preprocessing techniques. Data preprocessing is essential in complex high-throughput techniques as it attempts to discern differences caused by experimental factors and real biological differences between the samples, the latter of course being the one of interest. Better

techniques mean less experimental artifacts and more reliable data. With these new developments, experimental design needs to be rethought to maximize the quality of the results.

By using the SIBR method on data sets with permutated spot coordinates for six independent DIGE studies, we demonstrate the existence of spatial trends in the intensity measurements of the internal standard. Using only the biological samples in the entire experiment, we were able to model these trends with SIBR. The fact that they are highly correlated to the spatial trends obtained from the internal standard indicates a dye-independent effect. We compared the use of internal standard measurements with the SIBR normalization algorithm for their potential to remove spatial intensity biases. It is important to remark that both approaches might suffer from correlation issues that translate to a higher similarity between samples on the same gel than exist in reality. When taking ratios, the Cy3 and the Cy5 channel become correlated through the Cy2 channel, undermining the statistical assumption of independent sampling.[34] When using SIBR, measurements can become correlated because the same spatial biases are subtracted from all three samples on a gel. As a consequence, while the observed increases in sensitivity are still valid, calculated power values are not correct as such because the student *t* test, which is by far the most used, does not take such correlations into account. More complex statistical models, however, were beyond the scope of this article and are thus not discussed here.

Not removing the systematic bias that was occurring on the spatial level would in either case have resulted in an even larger correlation of the samples within a gel, and while the use of the internal standard is equally good (and sometimes better) for coping with spatial intensity bias than SIBR, its inclusion in the experimental design comes at a cost of losing one-third of the available sample space. This implies that, by relying on SIBR instead of an internal standard, an equal amount of samples can be run on less gels, resulting in lower costs, shorter sample preparation and image analysis steps, easier handling, and ultimately in less intergel variability. Alternatively, and more importantly, one could choose to run more biological samples on the same number of gels, leading to more measurements of what is really of interest. We demonstrate that by applying such an alternative experimental design (omitting the internal standard, but instead allocating an additional biological sample per gel), in combination with the SIBR normalization scheme, we are able to achieve a significant boost in statistical power for detecting differential expression. A boost in statistical power can reduce the false positive rate for identifying differential protein abundances (less erroneous identifications) without compromising sensitivity, a useful improvement as downstream validation analysis of these results can be very time and cost consuming. Such a boost in statistical power could also be used to obtain a higher sensitivity for the same false positive rate, that is, the ability to discriminate smaller differences in protein abundances.

Acknowledgment. This work is supported by (1) KULeuven Research Council: GOA/08/011, CoE EF/05/007, SymBioSys; CREA/08/023, OT 05-33, OT09/022; (2) IWT: SBO-BioFrame; (3) IUAP P6/25 (BioMaGNet); (4) FWO IOK-B9725-G.0329.09; and (5) HFSP-RGY0079/2007C.

Supporting Information Available: Supplementary files. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kubota, K.; Wakabayashi, K.; Matsuoka, T. Proteome analysis of secreted proteins during osteoclast differentiation using two different methods: two-dimensional electrophoresis and isotope-coded affinity tags analysis with two-dimensional chromatography. *Proteomics* **2003**, *3*, 5.

(2) Schmidt, F.; Donahoe, S.; Hagens, K.; Mattow, J.; Schaible, U.; Kaufmann, S.; Aebersold, R.; Jungblut, P. Complementary analysis of the Mycobacterium tuberculosis proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology. *Mol. Cell. Proteomics* **2004**, *3*, 1.

(3) Wolff, S.; Antelmann, H.; Albrecht, D.; Becher, D.; Bernhardt, J.; Bron, S.; Buttner, K.; van Dijl, J.; Eymann, C.; Otto, A.; Tam le, T.; Hecker, M. Towards the entire proteome of the model bacterium Bacillus subtilis by gel-based and gel-free approaches. *J. Chromatogr., B: Analyt. Technol. Biomed. Life Sci.* **2007**, *849*, 1–2.

(4) Wolff, S.; Otto, A.; Albrecht, D.; Zeng, J.; Buttner, K.; Gluckmann, M.; Hecker, M.; Becher, D. Gel-free and gel-based proteomics in Bacillus subtilis: a comparative study. *Mol. Cell. Proteomics* **2006**, *5*, 7.

(5) Unlu, M.; Morgan, M.; Minden, J. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **1997**, *18*, 11.

(6) Van den Bergh, G.; Arckens, L. Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel-based proteomics. *Curr. Opin. Biotechnol.* **2004**, *15*, 1.

(7) Alban, A.; David, S.; Bjorkesten, L.; Andersson, C.; Sloge, E.; Lewis, S.; Currie, I. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* **2003**, *3*, 1.

(8) Van den Bergh, G.; Arckens, L. Recent advances in 2D electrophoresis: an array of possibilities. *Expert Rev. Proteomics* **2005**, *2*, 2.

(9) Kreil, D.; Karp, N.; Lilley, K. DNA microarray normalization methods can remove bias from differential protein expression analysis of 2D difference gel electrophoresis results. *Bioinformatics* **2004**, *20*, 13.

(10) Fodor, I.; Nelson, D.; Alegria-Hartman, M.; Robbins, K.; Langlois, R.; Turteltaub, K.; Corzett, T.; McCutchen-Maloney, S. Statistical challenges in the analysis of two-dimensional difference gel electrophoresis experiments using DeCyder. *Bioinformatics* **2005**, *21*, 19.

(11) Kultima, K.; Scholz, B.; Alm, H.; Skold, K.; Svensson, M.; Crossman, A.; Bezard, E.; Andren, P.; Lonnstedt, I. Normalization and expression changes in predefined sets of proteins using 2D gel electrophoresis: a proteomic study of L-DOPA induced dyskinesia in an animal model of Parkinson's disease using DIGE. *BMC Bioinform.* **2006**, *7*, 475.

(12) Van de Plas, B. *Age-Dependent versus lesion-induced differential protein expression in mouse neocortex: similarities and differences*; Katholieke Universiteit Leuven, Faculty of Exact Sciences, Department of Biology, Neuroplasticity and Neuroproteomics group:: Leuven, Belgium, 2008; ISBN: 978−90−8649−190−2 ed.

(13) van Brussel, L.; Gerits, A.; Arckens, L. Identification and localization of functional subdivisions in the visual cortex of the adult mouse. *J. Comp. Neurol.* **2009**, *514*, 1.

(14) Sonck, K.; Kint, G.; Schoofs, G.; Vander Wauven, C.; Vanderleyden, J.; De Keersmaecker, S. The proteome of Salmonella Typhimurium grown under in vivo-mimicking conditions. *Proteomics* **2009**, *9*, 3.

(15) Sellers, K.; Miecznikowski, J.; Viswanathan, S.; Minden, J.; Eddy, W. Lights, Camera, Action! systematic variation in 2-D difference gel electrophoresis images. *Electrophoresis* **2007**, *28*, 18.

(16) Aittokallio, T.; Salmi, J.; Nyman, T.; Nevalainen, O. Geometrical distortions in two-dimensional gels: applicable correction methods. *J. Chromatogr., B: Analyt. Technol. Biomed. Life Sci.* **2005**, *815*, 1–2.

(17) Wheelock, A.; Buckpitt, A. Software-induced variance in two-dimensional gel electrophoresis image analysis. *Electrophoresis* **2005**, *26*, 23.

(18) Stessl, M.; Noe, C.; Lachmann, B. Influence of image-analysis software on quantitation of two-dimensional gel electrophoresis data. *Electrophoresis* **2009**, *30*, 2.

(19) Rosengren, A.; Salmi, J.; Aittokallio, T.; Westerholm, J.; Lahesmaa, R.; Nyman, T.; Nevalainen, O. Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels. *Proteomics* **2003**, *3*, 10.

(20) Raman, B.; Cheung, A.; Marten, M. Quantitative comparison and evaluation of two commercially available, two-dimensional elec-

trophoresis image analysis software packages, Z3 and Melanie. *Electrophoresis* **2002**, *23*, 14.
(21) Luhn, S.; Berth, M.; Hecker, M.; Bernhardt, J. Using standard positions and image fusion to create proteome maps from collections of two-dimensional gel electrophoresis images. *Proteomics* **2003**, *3*, 7.
(22) Berth, M.; Moser, F.; Kolbe, M.; Bernhardt, J. The state of the art in the analysis of two-dimensional gel electrophoresis images. *Appl. Microbiol. Biotechnol.* **2007**, *76*, 6.
(23) Huber, W.; von Heydebreck, A.; Sultmann, H.; Poustka, A.; Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **2002**, *18* (Suppl 1), 96.
(24) Karp, N.; Kreil, D.; Lilley, K. Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis. *Proteomics* **2004**, *4*, 5.
(25) Karp, N.; Lilley, K. Maximising sensitivity for detecting changes in protein expression: experimental design using minimal CyDyes. *Proteomics* **2005**, *5*, 12.
(26) Zhang, S.; Gant, T. A statistical framework for the design of microarray experiments and effective detection of differential gene expression. *Bioinformatics* **2004**, *20*, 16.
(27) Zhang, S.; Gant, T. Effect of pooling samples on the efficiency of comparative studies using microarrays. *Bioinformatics* **2005**, *21*, 24.
(28) Yang, Y.; Dudoit, S.; Luu, P.; Lin, D.; Peng, V.; Ngai, J.; Speed, T. Normalization for cDNA microarray data; a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **2002**, *30*, 4.
(29) Asirvatham, V.; Watson, B.; Sumner, L. Analytical and biological variances associated with proteomic studies of Medicago truncatula by two-dimensional polyacrylamide gel electrophoresis. *Proteomics* **2002**, *2*, 8.
(30) Karp, N.; Lilley, K. Design and analysis issues in quantitative proteomics studies. *Proteomics* **2007**, *7* (Suppl 1), 42.
(31) Rowell, C.; Carpenter, M.; Lamartiniere, C. Modeling biological variability in 2-D gel proteomic carcinogenesis experiments. *J. Proteome Res.* **2005**, *4*, 5.
(32) Karp, N.; Lilley, K. Investigating sample pooling strategies for DIGE experiments to address biological variability. *Proteomics* **2009**, *9*, 2.
(33) Karp, N.; Feret, R.; Rubtsov, D.; Lilley, K. Comparison of DIGE and post-stained gel electrophoresis with both traditional and SameSpots analysis for quantitative proteomics. *Proteomics* **2008**, *8*, 5.
(34) Karp, N.; McCormick, P.; Russell, M.; Lilley, K. Experimental and statistical considerations to avoid false conclusions in proteomics studies using differential in-gel electrophoresis. *Mol. Cell. Proteomics* **2007**, *6*, 8.

PR100010U