Determining a significant change in protein expression with DeCyder™ during a pair-wise comparison using two-dimensional difference gel electrophoresis.

Natasha A Karp[1], David P Kreil[2], and Kathryn Lilley[1].

1: Department of Biochemistry, University of Cambridge

2: Department of Genetics / Inference Group (Cavendish Laboratory), University of Cambridge

Corresponding Author: Kathryn Lilley

University of Cambridge, Department of Biochemistry, Building O, Downing Site, Cambridge CB2 1QW Tele: 01223 766442 Fax: 01223 333345

Email: ksl23@cam.ac.uk

Shortened version of the title: Determining a significant change in protein expression

Abbreviations:

DIGE: Difference gel electrophoresis

DIA: Differential in-gel analysis

BVA: Biological variance analysis

IPG: Immobilized pH gradient

cDNA: Complementary DNA

Graphical abstract: figure 2B

Keywords: Expression proteomics / Normalization / differential in-gel analysis / Difference gel electrophoresis

Summary

Two-dimensional difference gel electrophoresis (DIGE) is a tool for measuring changes in protein expression between samples involving pre-electrophoretic labeling with cyanine dyes. Here we assess a common method to analyze DIGE data using the DeCyder™ software system. Experimental error was studied by a series of same-sample comparisons. Aliquots of sample were labeled with N-hydroxyl succinimidyl ester-derivatives of Cy2, Cy3, and Cy5 dyes and run together on one gel. This allowed assessment of how experimental error influenced differential expression analysis. Bias in the log volume ratios was observed, which could be explained by differences in dye background. Further complications are caused by significant gel-to-gel variation in the spot volume ratio distributions. Using DeCyder™ alone results in an inability to define ratio thresholds for 90 or 95% confidence. Using more conservative thresholds, confidence in the significance of a change increases, but results in substantially more false negatives. An alternative normalization method was thus applied which resulted in improved data distribution and allowed greater sensitivity in analysis. When combined with a standardizing function, this allowed gel-independent thresholds for 90% confidence. The new approach, detailed here, represents a method to greatly improve the success of DIGE data analysis.

1 Introduction

2-D polyacrylamide gel electrophoresis is an important tool in proteomics as thousands of protein spots can be resolved, resulting in a global view of the state of a proteome [1]. Significant advances in proteomics have been realized through a union of mass spectrometry with 2-D gel analysis. Protein bands or spots can rapidly be identified through in-gel digestion, mass spectrometry, and database searching [2].

Early comparative proteomics relied on using images from different gels. Gel-to-gel variation, however, lead to problematic detection and quantification of differences in protein expression [3]. The high degree of gel-to-gel variation in the spot patterns has resulted in difficulties distinguishing biological from experimental variation [4]. To overcome these issues, Űnlű et al. (1997) developed an approach involving the multiplexing of samples, called 2-D difference gel electrophoresis (DIGE) [5], which has since been commercialized by Amersham Bioscience. DIGE involves labelling samples prior to electrophoresis with the spectrally resolvable fluorescent CyDyes (Cy2, Cy3, and Cy5). The labelled samples are then mixed before isoelectric focusing, and resolved on the same 2-D gel. Variation in spot intensities due to gel-specific experimental factors, for example protein loss during sample entry into the immobilized pH gradient strip, will be the same for each sample within a single DIGE gel. Consequently, the relative amount of a protein in a gel in one sample compared to another will be unaffected.

The first generation of CyDyes, supplied as N-hydroxyl succinimidyl ester-derivatives, reacts with the epsilon amino group of lysine residues in 3% of the proteins in a sample [6]. These 'minimal' dyes are charge and mass matched, which results in the labels having an equivalent impact on migration, allowing accurate comparison of fluorescent intensity [7]. Recently, new dyes that maximally label cysteine residues have become available [8]; however, this study focuses on the minimal dyes which are well established in the field. Minimal dyes have different system gain due to differences in laser intensities, fluorescence, and filter transmittance [4]. The spot volumes measured from DIGE gels must be normalized to compensate for these differences. Typically, the normalization method used is based on the assumption that the majority of protein spots have not changed in expression level.

For DIGE, two analysis methods have been developed and are options within the DeCyder™ software available through Amersham Biosciences. This is currently the only software available that is specifically designed for use in conjunction with DIGE that allows co-detection of proteins spots from multiple samples. Early work focused on difference in-gel analysis (DIA), which compares expression levels of two or three samples within a single gel [9-12]. In a DIA experiment, each sample is labelled with a different dye. The samples are pooled, and run on the same gel. Images are obtained from the gel for each fluorescent label. The software co-detects the spots from these images and, after normalization, compares the volume of a spot from one sample directly to that of another. The difference between two samples is reported as a

volume ratio. Depending on sample availability, this can be repeated many times and the results combined for greater reliability. The DeCyder™ version 4 manual suggests setting a threshold of two times the standard deviation (SD) of the distribution of the log volume ratios from a same-sample comparison [6]. The rationale behind using the 2xSD is that this will provide a measure of experimental variation that will encompass 95% of the spot changes that arise from experimental variation assuming the data follows a normal distribution. Consequently, protein spots with a fold change greater than this threshold are considered significant changes with a 95% confidence that the change is a difference between the two samples not due to random chance [6].

Amersham Biosciences subsequently developed a second approach, which they called biological variance analysis (BVA), where expression levels of samples are compared across gels [4, 10]. In a BVA experiment, the Cy2-labelled internal standard is run on every gel. The internal standard is a pooled sample comprising of equal amounts of each of the samples to be compared. This ensures that all proteins occurring in the samples are represented, allowing both inter- and intra-gel matching. The spot volumes from the labelled samples are compared to the internal standard giving standardized abundances, which allows the variation in spot running success to be taken into consideration. In a DIA approach to comparing multiple samples, each possible pair of samples must be compared directly in a gel; whereas, in a BVA approach, the internal standard allows the standardized abundances to be compared across gels. Given that BVA is based on DIA comparisons between each sample and

the internal standard, issues identified within this paper will have similar impact for BVA.

Whilst the inclusion of an internal standard has been demonstrated to improve the accuracy of relative protein quantification between samples [4], the BVA approach is not always appropriate due to the required cost, labour, time, and the requirement for greater amounts of sample. In many situations, when sample is limited or a preliminary investigation is intended, a pair-wise comparison with DIA can be used with a threshold above which a change is considered significant. Methods to measure experimental variance must be used to determine thresholds above which a change in expression levels can be considered statistically significant at a defined confidence level. Taking this approach, a statistical basis for interpreting protein expression level changes in the absence of replicates can be achieved.

In this article, we describe the assessment of experimental variance to provide guidelines for the interpretation of changes in protein expression in a pair-wise comparison approach. In this study, soluble proteins extracted from the gram-negative bacterium *Erwinia carotovora* were chosen as a simple biological system. To measure the experimental variance, aliquots of the same sample were labelled with different fluorescent dyes prior to multiplexing. In this situation, all proteins should be present in a ratio of one to one. Deviation from this reflects experimental variance.

Initially, all data was used in the analysis and consequently included a variety of artefact spots. This approach was chosen to reduce user intervention and hence increase reproducibility. Results from the same-sample comparison, however, identified artefacts apparently arising from differences in dye signal backgrounds. With these dye-specific artefacts present and the ratio distributions varying extensively from gel-to-gel, gel-independent statistical guidelines could not be determined for DeCyder™ normalized data for 90 or 95% confidence. Use of an alternative normalization method that compensates for differences in CyDye backgrounds is therefore proposed here that together with a scoring system allows generic threshold guidelines to be put forward. Our findings equally apply to multi-gel experiments standardized by a common reference sample.

2 Materials and methods

2.1 Sample preparation

Bacterial samples were grown in liquid broth media (10 g/l Bacto Tryptone, 5 g/l Bacto Yeast Extract, 5 g/l sodium chloride) at 30ºC and agitated at 300 rpm overnight and harvested by centrifugation for 10 min at 4ºC at 5000 rpm. Cells were resuspended in lysis buffer (8M urea, 4% w/w 3-[(3-Cholamidopropyl)dimethylammonio]-1-propane-sulfonate(CHAPS), 5 mM magnesium acetate, 10 mM Tris pH 8.0 and protease inhibitor cocktail set I at 1x concentration (Calbiochem, Germany) and lysed by sonication (3x10 s pulses on ice). Cell debris was removed by discarding the pellet formed after centrifugation for 10 min at 4ºC at 4500 rpm. To harvest the soluble protein fraction the sample was

centrifuged at 13 000 rpm for 10 minutes at 4ºC and the pellet discarded. The protein concentration was determined using the Bio-Rad DC protein assay as described by the manufacturers (Bio-Rad, UK).

2.2 CyDye labelling

Samples were labelled using the fluorescent cyanine dyes developed for 2-D DIGE (Amersham Biosciences, UK) following the manufacturer's recommended protocols. 50 µg of protein were labelled with 400 pmol of amine reactive cyanine dyes, freshly dissolved in anhydrous dimethyl formamide. The labelling reaction was incubated at room temperature in the dark for 30 minutes and the reaction was terminated by addition of 10nmol lysine. Equal volumes of 2x sample buffer (7 M urea, 2 M thiourea, 2% amidosulfobetaine-14, 20 mg/ml DTT and 2% Pharmalytes 3-10) were added to each of the labelled protein samples and the two samples were mixed. Rehydration buffer (7 M urea, 2 M thiourea, 2% amidosulfobetaine-14, 2 mg/ml DTT and 1% Pharmalytes 3-10) was added to make up the volume to 250 µl prior to isoelectric focusing.

2.3 Experimental design

Three 50 µg portions of *Erwinia carotovora* wildtype sample were labelled individually with Cy2, Cy3, and Cy5. The labelled samples were pooled and separated with 2-D gel electrophoresis as detailed below. The labelling experiment was independently repeated to produce six gels for the assessment of experimental variation.

2.4 2-D Protein separation by 2-D gel electrophoresis

13 cm Immobilized linear pH gradient (IPG) strips, pH 3-10 (Amersham Biosciences, UK) were rehydrated with CyDye-labelled samples for 10 hours at 20°C at 20 volts using the IPGphor II apparatus following manufacturer's instructions (Amersham Bioscience, Sweden). Isoelectric focusing was performed for a total of 40 000 Vhours at 20°C at 10 mA. Prior to SDS-PAGE, the strips were each equilibrated for 15 min in 100 mM Tris pH 6.8, 30% glycerol, 8 M urea, 1% SDS, 0.2 mg/ml bromophenol blue on a rocking table. The strips were loaded onto a 12% 13 cm (1 mm thick) acrylamide gel. The strips were overlaid with 1% agarose in SDS running buffer containing a 5 milligrams of bromophenol blue. The gels were run at 20 mV for 15 minutes and then at 40 mV at 20ºC until the bromophenol blue dye front had run off the bottom of the gels. A running buffer of 25 mM Tris pH 8.3, 192 mM glycine, and 0.1% SDS was used.

2.5 Gel imaging

Labelled proteins were visualized using a Typhoon™ 9410 imager (Amersham Biosciences, UK). The Cy3 images were scanned using a 532 nm laser and a 580 nm

band pass (BP) 30 emission filter. Cy5 images were scanned using a 633 nm laser and a 670 nm -BP30 emission filter. Cy2 images were scanned using a 488 nm laser and an emission filter of 520 nm BP40. All gels were scanned at 100 µm resolution. The photomultiplier tube was set to ensure a maximum pixel intensity between 40 - 60 000 pixels. Images were cropped to remove areas extraneous to the gel image using ImageQuant™ V5.2 (Amersham Biosciences, UK) prior to analysis. To ensure scanner performance was typical, comparable scans were obtained from a second Typhoon scanner in another institute.

2.6 DIA image analysis

Gel analysis was performed using DeCyder™ DIA V4.0 (Amersham Biosciences, Sweden), a 2-D gel analysis software package designed specifically to be used for DIGE. The estimated number of spots for each co-detection procedure was set to 2000. As recommended, an exclusion filter was applied to remove spots with a slope greater than one to reject spots that are likely to be contaminated with, or solely are dust particles.

2.7 Normalization

2.7.1 DeCyder™ normalization method

Within DeCyder™, normalization occurs by a scaling factor a to adjust the primary spot volume (equation 1) such that a frequency histogram of the log volume ratio centres on zero.

$V_1' = V_1 * a$ (1)

Here V_1' is the primary spot volume and V_1 is the normalized primary spot volume.

Outliers are excluded from the model curve fitting procedure by discarding data below 10% of the main peak height in the frequency histogram of log volume ratios [6]. To determine the scaling factor, a normal distribution is fitted to the main peak of the histogram using a standard least squares gradient descent [6]. The factor *a* is then chosen to shift the mean of the fitted normal distribution to zero, reflecting the assumption that most proteins are not expected to be differentially expressed.

The DeCyder™ program converts the calculated normalized log volume ratio R (equation 2) to an expression ratio E, by equations (3) and (4), to give a value greater than one or less than minus one, that is reported to the user as the measured 'ratio change'.

$R = \log_{10}(V_2 / V_1')$ (2)

Here V_1' is the normalized volume of the spot in the primary image and V_2 is the volume of the spot in the secondary image.

$E = (V_2/V_1')$ for $(R > 0)$ i.e., $V_2/V_1' > 1$ (3)

$E = -(V_1'/V_2)$ for $(R < 0)$ i.e., $V_1'/V_2 > 1$ (4)

2.7.2 Alternative normalization method

The spot volumes after background subtraction were extracted from the DIA module by the export results table facility in DeCyder™. The raw data before background subtraction were not available for export. To normalize the data, for each dye, the volumes were rescaled using relationship (5) where the scaling factor *a* adjusts for dye-specific system gain, and the additive offset *b* compensates for any constant additive bias present after the subtraction of local background estimates.

$$V' = a*V + b \quad (5)$$

The transform parameters *a* and *b* are determined by an iterative trimmed least squares maximum likelihood estimate assuming that most proteins are not expected to be differentially expressed. For this process, code for the normalization of DNA microarray data by Huber *et al*. (2002) [13] has been adapted for our purpose. Both additive and multiplicative noise are explicitly allowed for in the model employed. It should be noted that, if additive noise is also permitted, it is not a log-transform that decouples the variance from the signal intensity (a property desired for statistical analysis), but an asinh-transform [13-16]. The difference between the two transforms is pronounced for signal values close to zero, but disappears quickly for larger signal levels. For the gels encountered in our study, normalized signal levels were high, so that the asinh transformed data were identical to the log-transformed data within measurement error.

2.8 Validation of empirically determined thresholds

In a ‘leave-one-out approach’ to cross-validation, thresholds based on observed experimental variance were empirically determined from data pooled from five of the six gels. The thresholds were then tested on the remaining, sixth gel to assess whether the false positive rates were compatible with expectations. This was repeated so that each gel was once excluded from the pool and used as a test set.

3 Results

3.1 Analysis of significance with DeCyder™ DIA

Within DeCyder™, normalization occurs by a scaling factor to adjust the primary spot volume for system gain differences between the two dyes (see methods section 2.7.1 and equation (1) for details). From the normalized volumes of a spot, a log volume ratio is calculated, converted to an expression ratio, and reported to the user as indicative of changes in protein expression. The expression ratio, however, forms a disjoint distribution that complicates analysis. This paper subsequently focuses on the log volume ratio.

When labelling the same sample with the three dyes, and normalizing for dye-specific differences, equal amounts of fluorescent signal from the three dyes are expected for any particular resolved spot. Any deviation from the expected log-ratio of zero reflects experimental variance. Whichever dye combinations were compared, the ratio distributions were not normally distributed as they exhibited excessive (heavy) tails and asymmetry (skew) (Table 1 & Fig. 1).

Figure 2 demonstrates how deviations from normality mostly arise from low volume spots. Two effects can be distinguished: (A) the heavy tails of the distribution; and (B) a directional bias at low volumes. Transforming the data with a log function allows better visualization of the data across the several orders of magnitude covered and clearly shows the bias at low volumes. The direction of this bias depends on which dye combinations are being compared. Whilst a log transform will affect the magnitudes of relative errors, the transformation cannot introduce a bias. These observations suggest that the directional bias is an artefact of the system. The possible origin of the bias is discussed further in section 3.2.

Amersham Biosciences recommend using 2x the standard deviation of a same-same comparison as the threshold for the detection of a significant change when using the DIA approach [6]. This is a classic method for the detection of outliers in normally distributed data based on two times the standard deviation giving thresholds for 95% confidence. However, as the normal distribution is a bad approximation to the data, the validity of these thresholds is questionable. The data is clearly skewed and exhibits strong kurtosis (Table 1).

Previously, all spots were included in the analysis; however, some of these will not arise from protein signals. Consequently, an alternative approach might be to only consider ‘confirmed’ spots as it could be argued that the deviations from normality arises from non-real spots. The user confirms spots as real protein spots, after examining a three-dimensional display of the local data around the spot in the

DeCyder™ program for a regularly smooth cone profile (see Fig. 3 for examples). For this study, one gel was chosen, and each spot was examined and either confirmed as 'real' or excluded from the analysis. Confirming spots, however, is time intensive and the user intervention can lead to user-specific errors. User validation lead to approximately 50% of mostly low volume spots being discarded, after which the data set was re-normalized and re-analyzed. The log volume ratio distribution, however, was still distorted. Using only confirmed spots improved the distribution by reducing skew; however, the distribution had increased kurtosis (Table 2). Furthermore, a bias at low volumes remains (Fig. 4).

Using only confirmed spots has the effect of removing low volume spots, as these are harder for a user to distinguish as 'real'. Thus, a comparable filtering method would be to delete low volume spots below a threshold from the data set, especially since it could be argued that these low volume spots would not have sufficient volume for mass spectroscopy. However, the exclusion of low volume spots does limit the sensitivity of DIGE.

Deletion of spots below different volume thresholds does not improve skew or kurtosis of the data (Table 3). The removal of spots with spot volume <40 000 has, however, deleted all spots which triggered the bias originally seen in the log volume ratios at low spot volumes (Fig. 2). Even after removal of the bias, the data are still far from normal, which results in the classic 2x SD approach still being inappropriate. Continued deletion of low volume data does not lead to further improvements of the

data distribution, as shown in the last two columns of table 3. This suggests that the remaining data distortion arises from other systematic effects.

3.2 Normalization

The raw spot volumes obtained by DeCyder™, co-detected and background corrected, can be exported and allow analysis prior to normalization. Examining the raw volume data in Cy3 versus Cy5 analysis shows a good linear relationship on a large scale (Fig. 5A). After log transformation (Fig. 5B) the bias in the distribution for low volume spots can clearly be seen, which indicates that the bias arises prior to normalization. For the comparison of spot volumes from different dyes, systematic differences need to be compensated for. Various normalization methods to address this have been developed.

The normalization system used in DeCyder™ is based on a constant scaling factor to account for differences in system gain (see equation 4 and section 2.7.1). This normalization method has been shown to not correct for the bias seen at low volumes (Fig. 2). A different approach to normalization has been developed for fluorescently labelled cDNA microarrays that includes scaling factors as well as additive offsets (see equation 5 and section 2.7.2) [13, 15] . Processing the DeCyder™ spot volume data with this normalization method instead yields data distributions that do not exhibit any bias for the low volume spots (Fig. 6).

The need for additive offsets in normalization can be explained by systematic differences in the estimation of dye background signal fluroescence. Fluorescent background arises from intrinsic fluorescence in the gel and glass plates. It was assessed by scanning a number of blank gels after completing the second gel dimension with a fixed photo multipler tube setting of 600. Background was highest in Cy2 (4679±717) followed by Cy3 (4243±1615), lowest in Cy5 (93±48), and was independent of user and scanner.

Within DeCyder™, the local background is estimated by the 10th percentile value of all the pixel values observed on the spot boundary. This is subtracted from each of the pixels that form the spot prior to summation in the calculation of the spot volume. Estimation of background is difficult in a 2D protein system as spots can frequently overlap. Any background estimation procedure that has a dye specific bias can lead to the bias in the log ratios seen in figure 2. This is illustrated by the following hypothetical example:

In this hypothetical example, a spot has the same quantity of Cy3 and Cy5 labelled protein and the Cy5 dye has a system gain two times that of the Cy3 dye (see Table 4). In this example, the 10th percentile approach underestimates the background in the Cy3 channel. This could occur if the noise of the background is high such that the 10th percentile value is further from the real background value. The spot signal is calculated by the subtraction of background from each pixel that contributes to the volume and is then normalized. After the DeCyder™ normalization, the Cy3 spot has a greater volume than the Cy5 spot even though the protein amounts are equivalent.

The use of the alternative normalization to correctly compensate for the differences in background through additive offsets leads to equivalent volumes (Table 4).

The error introduced by a bias in the estimation of the background will have an impact across all volume ranges; however, as the signals increase, the relative error will decrease. The low volume spots, however, form a significant proportion of the spots in the data set. Use of an alternative normalization method that compensates for dye-specific background effects improves the accuracy of spot volume estimation. The use of the alternative normalization will, therefore, extend the sensitivity of the analysis method.

Despite successful removal of bias at low volumes, the data is still heavy tailed (see below), however, the normality has improved significantly compared to DeCyder™ normalized data. This can see in a normal quantile plot (QQ plot), which can be used to assess normality. In a QQ plot the percentiles of the observed data values are plotted against the percentiles from a normal distribution. If the observed distribution is normal a straight line will be seen. Deviations from the straight line indicate deviations from normality. A QQ plot of the data normalized by the DeCyder™ approach highlights the extensive deviations from normality (see Fig. 7). For the data normalized with the alternative approach the data is significantly more normal but still exhibits heavy tails (see Fig. 8). These remaining systematic deviations from normality are not pursued further here and require additional research to identify possible causes. The presence of the heavy tails results in the standard deviation

based thresholds for the alternative normalized data being equally inappropriate for determination of significant expression changes.

3.3 Percentile approach for determining thresholds

Initially, the analysis focused on the DeCyder™ normalized log volume ratios of either the whole data set, or the data set where spots with volumes less than 40 000 had been excluded. Cross-validation showed that no gel-independent thresholds could be proposed because the distributions from the gels varied significantly (see supplementary data).

As a new alternative approach, a robust Z score is introduced (see equation (6)).

$$Z = [T-\text{median}(T)]/\text{MAD}(T) \qquad (6)$$

Here MAD is the median absolute deviation and is used as a robust estimator of scale. The median is a robust estimate of location, and T the normalized log volume ratio. This transformation is essential in overcoming the gel-to-gel variation observed, and now allows pooling of data from multiple gels as the distributions of the transformed data all center on zero with a comparable spread.

With the bias varying significantly from gel-to-gel in DeCyder™ normalized data, the Z score approach could not overcome the gel differences to allow generic

thresholds. The alternative normalization, together with the calculation of a robust Z score, however, made the distributions comparable enough to allow gel-independent thresholds to be put forward. On average, the false positive rates observed in cross-validation were compatible with expectations for thresholds for 90% confidence (Table 5). For more stringent thresholds, larger deviations were observed (see Table 5). Asymmetric thresholds are expected, as the distribution is asymmetric even after the alternative normalization method. Robust thresholds of -2.33 and +1.52 were obtained for 90% confidence for the Z score data sets calculated after alternative normalization.

3.4 Range approach for determining thresholds

To allow use of DeCyder™ in its current format, one option is to look for expression changes that occur outside of the range seen in the same-sample analysis. Table 6 demonstrates the variation seen from gel-to-gel, even with low volume data excluded. Due to the high variation, gel-independent thresholds cannot be proposed. However, from the range of expression ratio values observed, thresholds of ± 2.8 would have given no false positives in these sets of same-sample experiments. This suggests a ratio change larger than this is unlikely to occur by chance. Thresholds of ± 2 can be used to assess expression changes with high confidence that the change is significant as less than 0.1% false positives were observed in the six same-sample gels. With these extreme thresholds, however, the number of false positives is low but the number of false negatives will be very high and, consequently, this results in a very low sensitivity of detecting expression changes.

4. Discussion

Multiplexing of samples by the use of spectrally resolvable dyes with co-detection of signals has lead to significant improvements in quantitative expression analysis of 2D gels. A pair-wise comparison of two samples within the same gel is often the method of choice for sample-limited situations and forms the basis of multi-gel comparisons (such as BVA). Inherent with the in-gel analysis process using fluorescent CyDyes is an artifact at low spot volumes that could be explained by differences in dye backgrounds. After removing low volumes spots, the data is still far from following a normal distribution. In particular, the data is highly asymmetrical; consequently, the use of SD based thresholds on the log volume ratio obtained from DeCyder™ is statistically unjustified. Attempts were made with various transformations, scoring processes, and methods that made no assumptions regarding the distribution of data, to find gel-independent thresholds. However, generic thresholds for 95 or 90% confidence that held up to cross-validation could not be obtained for the DeCyder™ normalized data. Thresholds based on the range of values observed when low volume spots were discarded could be used for DeCyder™ normalized data. With these extreme thresholds, the confidence in a ratio change above the threshold is high; however, the number of false negatives will increase accordingly. Consequently, a large number of spots of interest will be missed.

Use of an alternative normalization method that for each dye includes an offset element along with a scaling factor, and hence can account for differences in dye

backgrounds, improves the distributions of the ratios obtained considerably. The distributions are still not normal and vary strongly from gel-to-gel. However, calculating a standardizing robust Z score for the offset & scale normalized volume data lead to distributions where gel-independent thresholds could be calculated that stood up to cross-validation. Using this analysis method for a Cy3/Cy5 comparison, Z score thresholds of –2.33 and +1.52 could be determined, beyond which a change was significant with 90% confidence.

Analysis of DIGE data poses various problems for the determination of statistical guidelines for the detection of significant change as the gel-to-gel variation is high, and the log-ratio distributions are far from normal and exhibit bias. As BVA is based on pair-wise comparisons between each sample and the internal standard, issues identified within this paper will have a similar impact for BVA. A dye swap to balance dye effects will negate the bias if the bias magnitudes are similar; however, the results in this study suggest that the bias is quite variable across different gels. Incorporating the alternative normalization approach, which removes the bias at low volumes, will increase the sensitivity of both DIA and BVA approaches to DIGE analysis in detecting significant changes.

Within this manuscript, various recommendations have been made to improve the accuracy and sensitivity of DIGE analysis; ideally future revision of common analysis tools will incorporate these findings.

In summary, for a DIA approach there are two options:

An expression ratio approach:

1. Use DeCyder™ to co-detect the spots
2. Apply a filter to exclude spots with a volume less than 40 000
3. In DeCyder™ re-normalize the data set
4. Use expression ratio thresholds of ± 2 for significant change. With these thresholds the user has a high confidence but reduced sensitivity.

Z score approach:

1. Use of DeCyder™ DIA to co-detect the spots and generate the spot volume information.
2. Calculate additive offsets and scaling transforming parameters for the normalization as described in section 2.7.2 and Huber et al (2002) [13].
3. Calculate a robust Z score for each spot using equation (8).
4. Use thresholds of –2.32 and + 1.52 on the Z score to determine significant changes with a 90% confidence.

Experimental variation has been assessed within this laboratory to provide guidelines for determining a significant change. These guidelines, however, are based on the experimental variation occurring within this laboratory. As differences can be expected between different laboratories these values can only be taken as typical. This paper, however, provides a method by which individual laboratories can establish reliable thresholds.

Update: DeCyder™ version five software will soon be released. The issues arising within this study are still present with analysis using the new software.

5. References

(1) Lilley, K. S.; Razzaq, A.; Dupree, P. *Curr Opin Chem Biol* 2002**,** *6*, 46-50.

(2) Fey, S. J.; Larsen, P. M. *Curr Opin Chem Biol* 2001**,** *5*, 26-33.

(3) Asirvatham, V. S.; Watson, B. S.; Sumner, L. W. *Proteomics* 2002**,** *2*, 960-968.

(4) Alban, A.; David, S. O.; Bjorkesten, L.; Andersson, C., *et al. Proteomics* 2003**,** *3*, 36-44.

(5) Ünlü, M.; Morgan, M. E.; Minden, J. S. *Electrophoresis* 1997**,** *18*, 2071-2077.

(6) Amersham Biosciences. *Ettan DIGE User manual*, 2002.

(7) Tonge, R.; Shaw, J.; Middleton, B.; Rowlinson, R., *et al. Proteomics* 2001**,** *1*, 377-396.

(8) Shaw, J.; Rowlinson, R.; Nickson, J.; Stone, T., *et al. Proteomics* 2003**,** *3*, 1181-1195.

(9) Gharbi, S.; Gaffney, P.; Yang, A.; Zvelebil, M. J., *et al. Mol Cell Proteomics* 2002**,** *1*, 91-98.

(10) Bergh, G. V. D.; Clerens, S.; Vandesande, F.; Arckens, L. *Electrophoresis* 2003**,** *24*, 1471-1481.

(11) Hu, Y.; Wang, G.; Chen, G. Y.; Fu, X., *et al. Electrophoresis* 2003**,** *24*, 1458-1470.

(12) Yan, J. X.; Devenish, A. T.; Wait, R.; Stone, T., *et al. Proteomics* 2002**,** *2*, 1682-1698.

(13) Huber, W.; Heydebreck, A. v.; Sültmann, H.; Poustka, A., *et al. Bioinformatics* 2002**,** *18*, 96S-104S.

(14) Munson, P. *Gene-Logic Workshop of Low Level Analysis of Affymetrix GeneChip Data* 2001.

(15) Durbin, B.; Hardin, J.; Hawkins, D. M.; Rocke, D. M. *Bioinformatics* 2002**,** *18*, 105S-110S.

(16) Durbin, B.; Rocke, D. *Bioinformatics* 2003**,** *19*, 1360-1367.

(17)

Acknowledgements

We would like to thank Dr. Alan Robinson for critical reading of this paper, Sarah Coulthurst for provision of sample and Dr. Ian Currie from Amersham Bioscience for interesting discussions. We would also like to thank KuDOS Pharmaceuticals for use of their Typhoon Scanner. This work was supported by the BBSRC Investigating gene function initiative. Dr. Kreil acknowledges support by a Medical Research Council Research Fellowship (G81/555) and N. Karp is funded as a BBSRC research associate (8/G18420).

Figure 1: Example of a frequency distribution of the log volume ratios obtained from a gel where the same sample was analyzed with DIA in a Cy3/Cy5 comparison. The deviations from a normal distribution can be seen in a long tail stretch into the negative log volume ratios, and an asymmetry distribution around the apex.

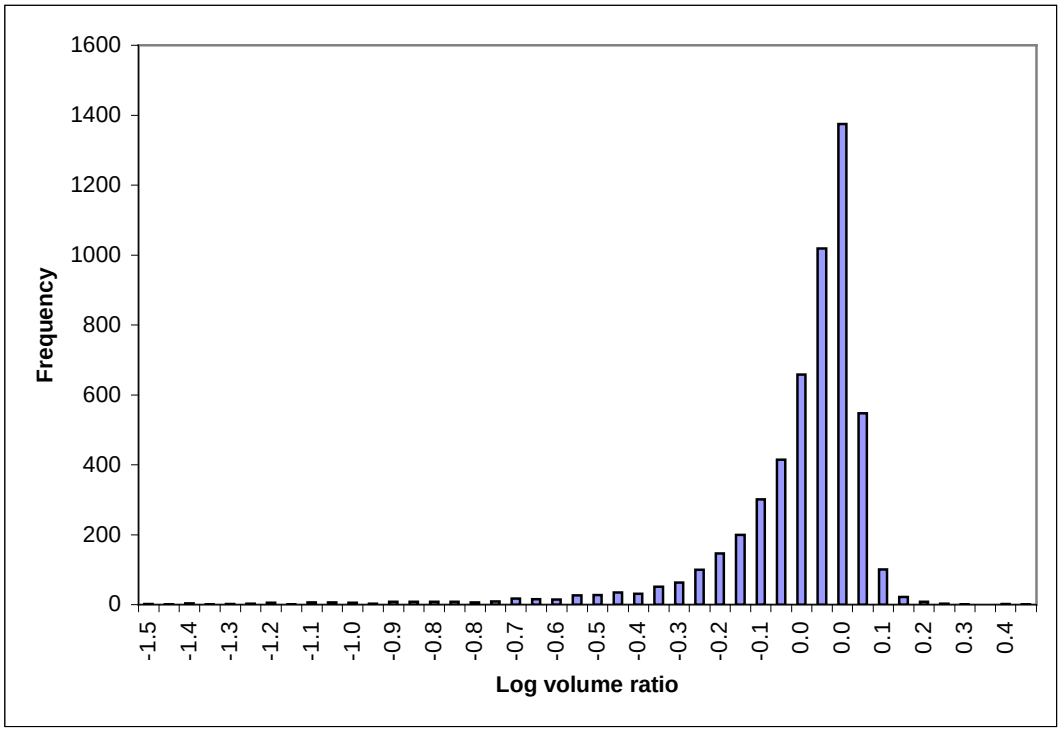


Table 1: Distributional analysis of the log volume ratios obtained from DIA of various dye pair combinations. Data from four gels run with the same-sample labelled with Cy3, Cy5, and Cy2 has been combined. Skew, a measure of asymmetry, for a normal distribution should equal one. Kurtosis, a measure of the pointiness of a distribution, for a normal distribution should equal zero.

Analysis of log volume ratio	Cy3/Cy5	Cy2/Cy5	Cy3/Cy2
Average	-0.072	-0.123	0.052
Median	-0.017	-0.025	0.017
Number of data points	5276	4383	4738
Skew	-3.153	-2.455	0.767
Kurtosis	14.188	7.812	4.061
Standard deviation	0.181	0.307	0.179

Figure 2: Distribution of the log volume ratios versus average spot volumes for a same-sample Cy3/Cy5 DIA comparison. For a same-sample comparison the log volume ratio should centre on zero. Any deviations from this indicate experimental variance. A: On a linear scale, a long tail can be seen at low spot volumes; and B: on a logarithmic scale a clear directional bias is seen at low spot volumes.

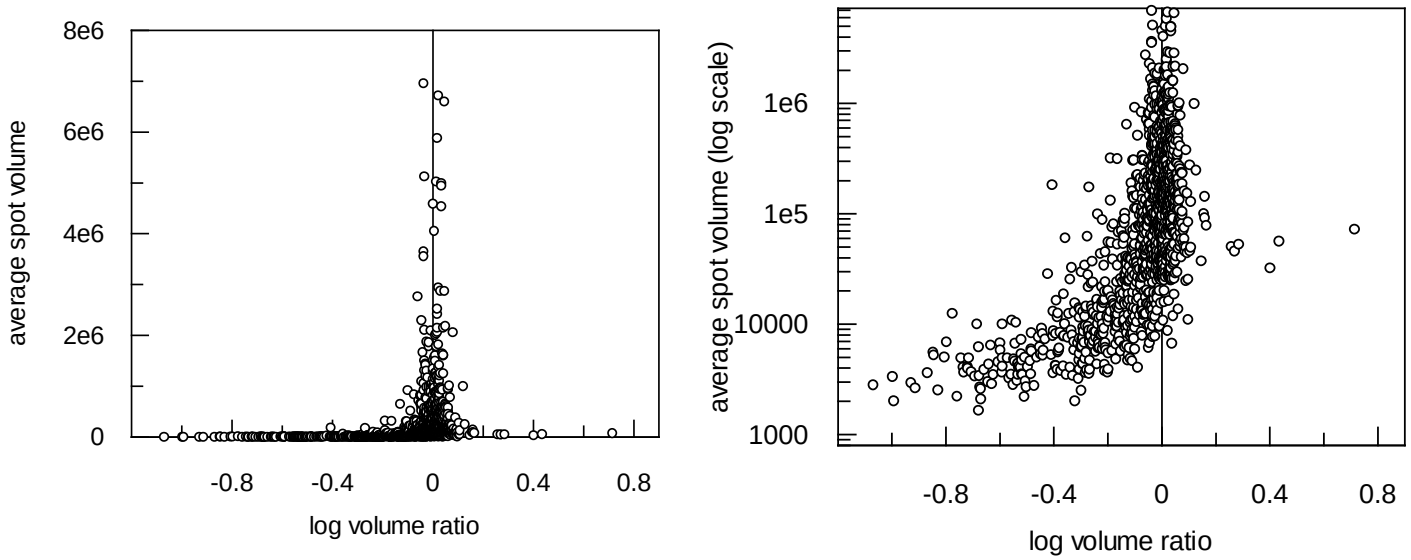
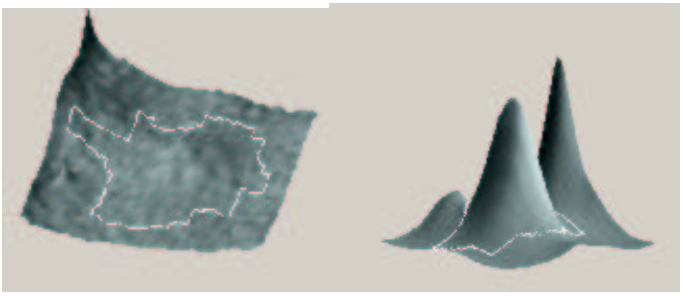


Figure 3: DeCyder™ 3-D display of spot image data used to judge spots validity.

A: example of a low volume spot; and B: example of a clear protein spot.



A B

Figure 4: Distribution of log volume ratios versus average spot volumes for the same-sample Cy3/Cy5 DIA comparison keeping only manually confirmed spots. The bias at low volumes is still apparent.

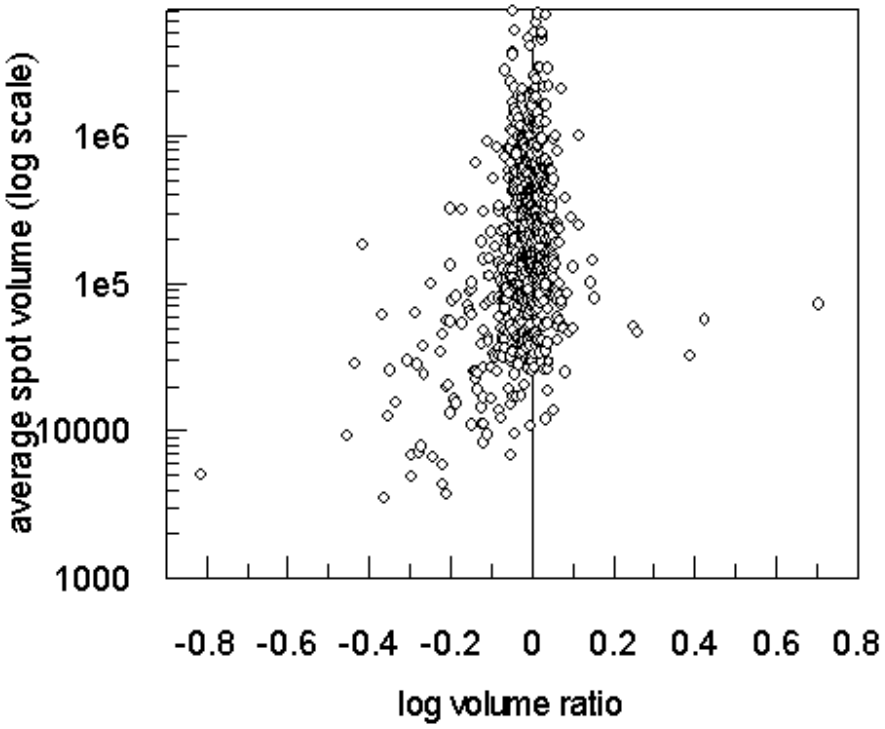


Table 2: Comparison of the distributional analysis of the log volume ratios obtained for all spots versus that of confirmed spots from a DIA comparing same samples labelled with Cy3 and Cy5.

Analysis of log volume ratio	All spots	Confirmed spots
Average	-0.094	-0.020
Median	-0.033	-0.004
Number of data points	1268	684
Skew	19.64	-1.181
Kurtosis	5.315	18.629
Standard deviation	0.181	0.091

Table 3: Comparison of the distributional profiles of the log volume ratio distribution for data sets with different volume ranges for a same-sample Cy3/Cy5 comparison. The column headings indicate the volume cut-off below which spots were deleted from the data set.

Data set	All	< 5000	< 10000	< 20000	< 40,000	< 50000	< 60000
Analysis							
Average	-0.085	-0.039	-0.021	-0.009	-0.004	-0.002	-0.001
Median	-0.021	-0.013	-0.004	0.000	0.004	0.004	0.043
Number of data points	1268	1136	1026	916	752	690	656
Skew	-2.00	-0.85	-0.23	0.44	-1.67	-1.50	-1.64
Kurtosis	5.31	5.22	9.66	16.48	7.73	7.61	8.41
SD	0.181	0.115	0.095	0.086	0.067	0.065	0.062

Figure 5: Analysis of raw spot volumes for a Cy3/Cy5 same-sample comparison. A: comparison of volumes on a linear scale. B: comparison of volumes on a logarithmic scale. On the logarithmic scale the bias at low intensity volumes becomes apparent.

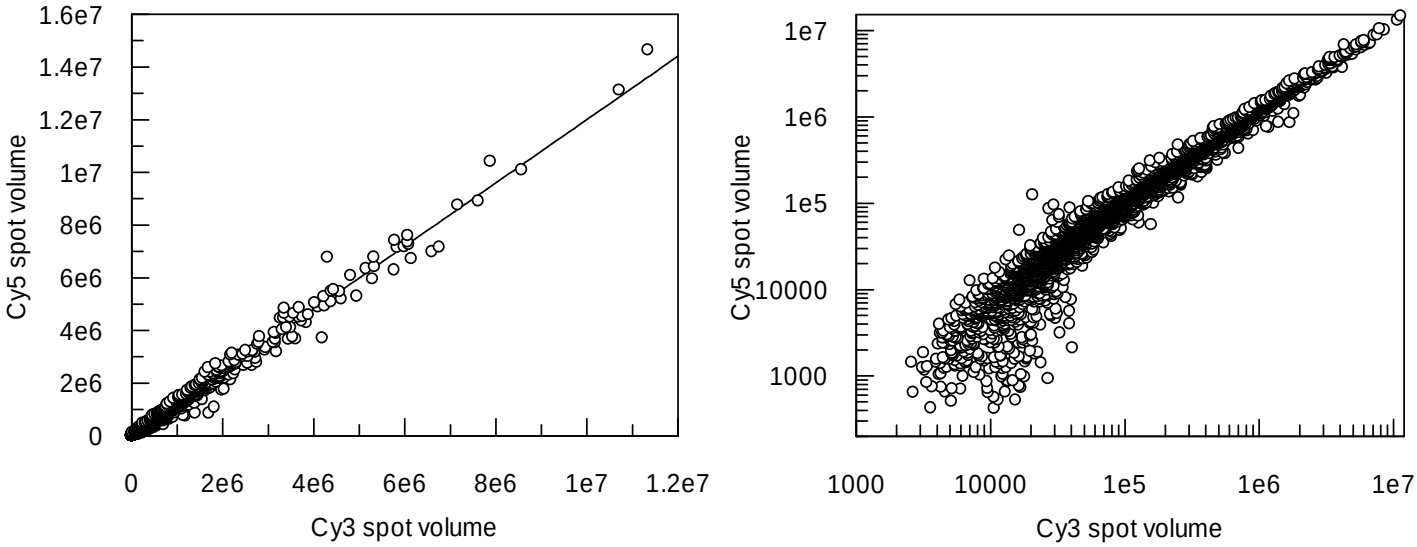


Figure 6: Data distribution when a Cy3/Cy5 same-sample comparison is normalized using the alternative method that includes offsets and scaling factors, showing the log-ratios versus the averages of the spot volumes. Both the x- and y-axis have different scales compared to those of figure 2B due to the normalization employed.

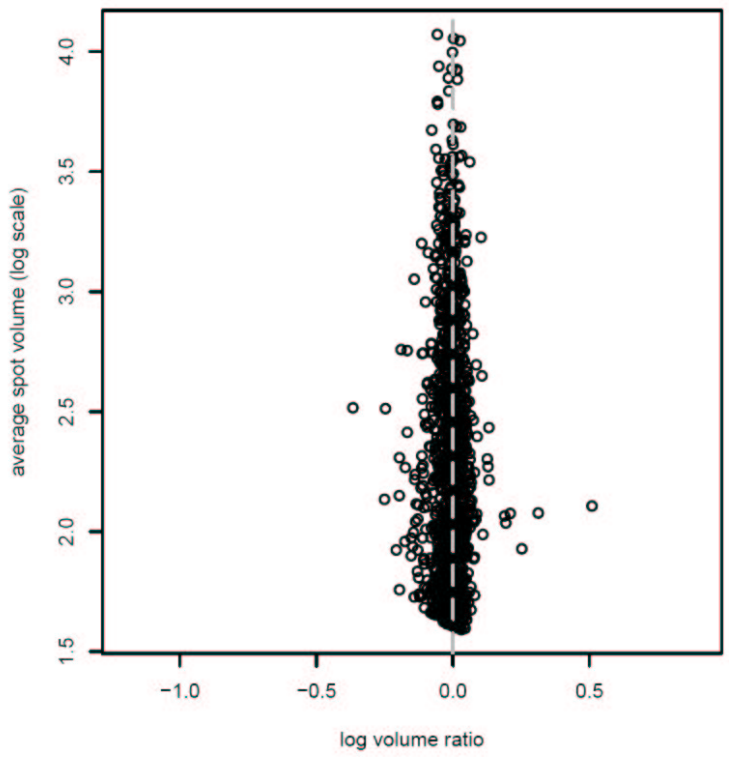


Figure 7: Example of a normal quantile plot for the distribution of log volume ratios obtained from a gel analyzed with DIA in a Cy3/Cy5 same-sample comparison. The data was normalized using the scaling DeCyder™ approach.

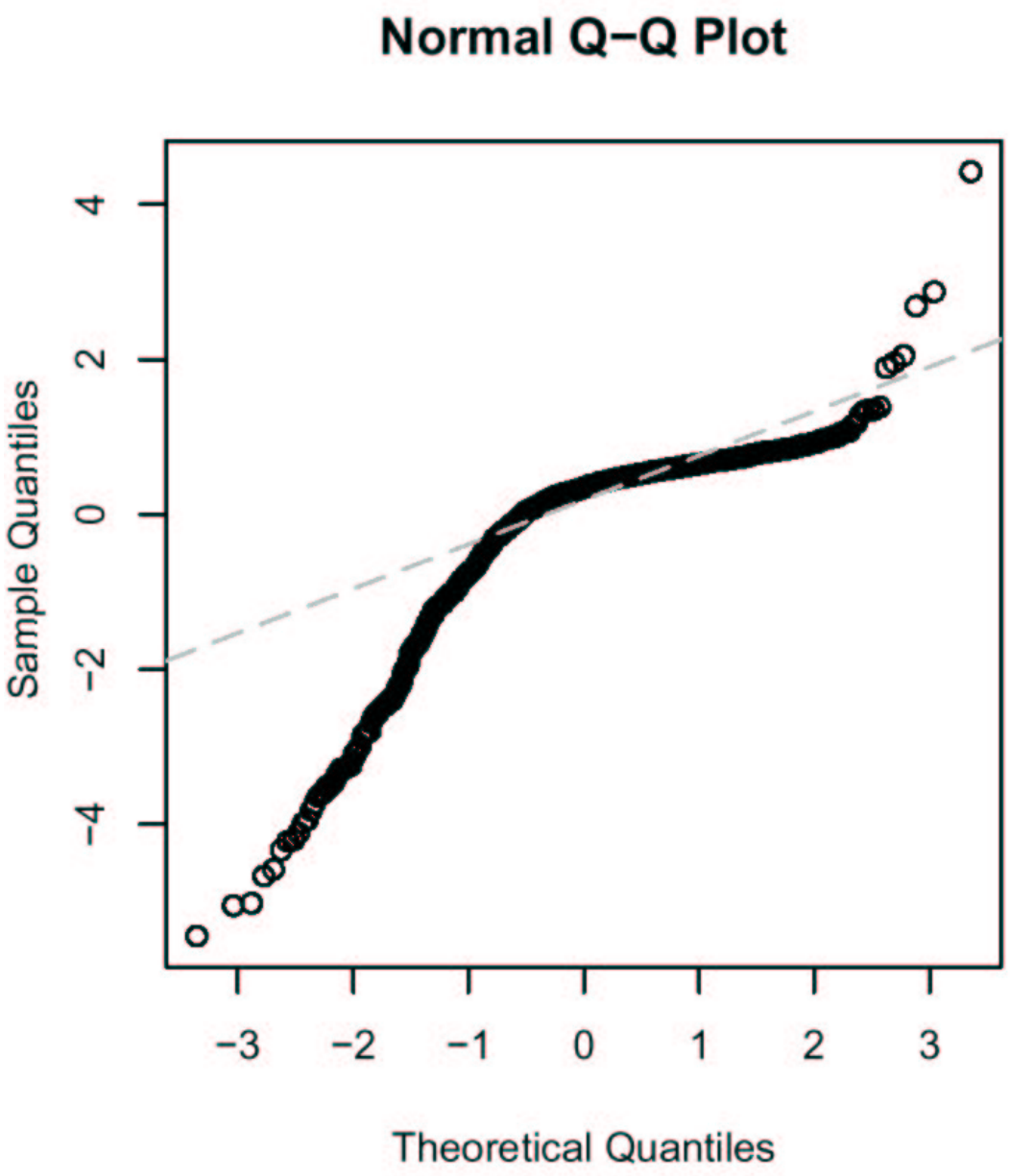


Figure 8: Example of a normal quantile plot for the distribution of the log volume ratios obtained from a gel analyzed with DIA in a Cy3/Cy5 same-sample comparison. The data was normalized using the alternative method.

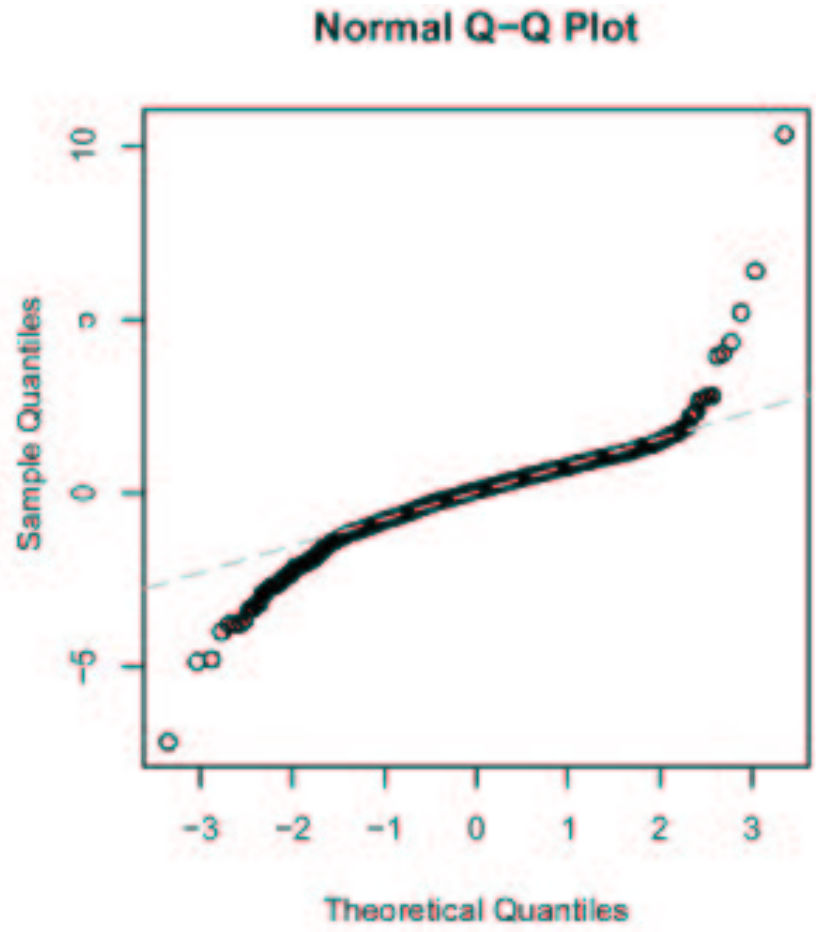


Table 4: Comparison of the impact of bias in the estimation of the background signal for DeCyder™ and the alternative normalization method for a hypothetical low volume spot. Assuming in this hypothetical situation that the Cy5 dye has system gain two times that of the Cy3 dye, Cy3 has a higher background signal, and that the 10th percentile background approach underestimates the background for the Cy3 spot. A: In the scaling normalization the spot volume 1 is multiplied by a scaling correction factor of two to account for differences in dye potency. B: In the alternative normalization method an offset value is added to spot volume 2 after it has been scaled by the difference in potency.

	Spot Cy3 volume 1	Spot Cy5 volume 2
Spot volume from protein (with Cy5 2x potency Cy3)	2000	4000
Background 'real'	4000	100
Observed signal (spot volume + background)	6000	4100
10th percentile calculated background	3500	100
Calculated spot signal (spot volume – background)	2500	4000
After scaling normalization [A]	5000	4000
Signal change from scaling normalization	Spot Cy3 > spot Cy5	
After alternative normalization [B]	5000	5000
Signal change from alternative normalization	Spot Cy3 = spot Cy5	

Table 5: The thresholds for 90 and 95% confidence obtained from the distributions of the robust Z-score calculated on the alternative normalized spot data with the false positive rate measured on the validating gel. The data set was obtained from six Cy3/Cy5 same-sample comparison gels and included all spots volumes. From the various pool combinations an average percentage false positive rate was calculated (shown in bold).

Gel excluded	Results when confidence set for 95%			Results when confidence set 90%		
	Thresholds from pooled gels		Total % false positives	Thresholds from pooled gels		Total % false positives
	Lower	Upper		Lower	Upper	
1	-3.30	1.80	5.95	-2.39	1.49	10.66
2	-3.26	1.89	4.43	-2.28	1.53	10.05
3	-3.32	1.90	2.85	-2.35	1.53	8.56
4	-1.92	1.86	4.02	-2.37	1.52	8.75
5	-2.86	1.93	6.85	-2.20	1.54	11.59
6	-3.34	1.80	5.75	-2.38	1.49	10.9
Average ± SD	**-3.0±0.56**	**1.86±0.05**	**4.98±1.47**	**-2.33±0.07**	**1.52±0.02**	**10.09±1.21**

Table 6: The expression ratio range obtained from Cy3/Cy5 same-sample comparisons for six gels where low volume data (< 40 000) had been excluded.

Gel	Range
1	-1.71 to +2.57
2	-1.97 to +1.57
3	-1.80 to +1.55
4	-1.80 to +1.55
5	-2.75 to +1.37
6	-1.48 to +1.44