DOI 10.1002/prca.200800066 *Proteomics Clin. Appl.* 2009, *3,* 394–407

RESEARCH ARTICLE

Considerations for powering a clinical proteomics study: Normal variability in the human plasma proteome

David Jackson[1], Athula Herath[1], Jonathan Swinton[1], David Bramwell[2], Rajesh Chopra[1], Andrew Hughes[1], Kevin Cheeseman[1] and Robert Tonge[1]**

[1] AstraZeneca Pharmaceuticals, Alderley Park, Macclesfield, UK
[2] Nonlinear Dynamics, Newcastle, UK

Received: March 3, 2008
Revised: September 2, 2008
Accepted: September 3, 2008

Proteomics is increasingly being applied to the human plasma proteome to identify biomarkers of disease for use in non-invasive assays. 2-D DIGE, simultaneously analysing thousands of protein spots quantitatively and maintaining protein isoform information, is one technique adopted. Sufficient numbers of samples must be analysed to achieve statistical power; however, few reported studies have analysed inherent variability in the plasma proteome by 2-D DIGE to allow power calculations. This study analysed plasma from 60 healthy volunteers by 2-D DIGE. Two samples were taken, 7 days apart, allowing estimation of sensitivity of detection of differences in spot intensity between two groups using either a longitudinal (paired) or non-paired design. Parameters for differences were: two-fold normalised volume change, α of 0.05 and power of 0.8. Using groups of 20 samples, alterations in 1742 spots could be detected with longitudinal sampling, and in 1206 between non-paired groups. Interbatch gel variability was small relative to the detection parameters, indicating robustness and reproducibility of 2-D DIGE for analysing large sample sets. In summary, 20 samples can allow detection of a large number of proteomic alterations by 2-D DIGE in human plasma, the sensitivity of detecting differences was greatly improved by longitudinal sampling and the technology was robust across batches.

Keywords:
2-D DIGE / Image analysis / Plasma / Statistical power / Variability

1 Introduction

Biomarkers, as molecular indicators of a particular physiological state, biological process, drug response or disease, are in great demand within the clinical setting. Most existing biomarkers have been identified by conventional biochemical approaches; however, technologies such as genomics, proteomics and metabonomics are being applied in an effort to produce further candidate biomarkers. Proteomics is the study of the protein complement of a cell, tissue or organism, and represents a dynamic entity incorporating changes in expression, transport and sequestration, activity, modification and turnover. Proteomics has recently delivered promising tissue biomarkers, *e.g.* [1–3]. Proteomics has also particularly found application in recent years in the characterisation of the plasma and serum proteome. Plasma and serum enable the collection of multiple longitudinal samples from a single individual, a feat not easily achievable with

Correspondence: Dr. David Jackson, Clinical and Biomedical Proteomics Group, Cancer Research UK Clinical Centre, St. James's University Hospital, Leeds LS9 7TF, UK
E-mail: d.h.jackson@leeds.ac.uk
Fax: +44-113-2429886

Abbreviations: CSF, cerebrospinal fluid; **NLD,** Nonlinear Dynamics; **PG240,** Progenesis discovery version 240; **QC,** quality control; **SS,** SameSpots

* Current address: Clinical and Biomedical Proteomics Group, Cancer Research UK Clinical Centre, St. James's University Hospital, Leeds,UK.

solid tissue biopsies given the invasive nature of collection and potentially large heterogeneity between serial samples. Further, it is our belief that longitudinal sampling, using an individual as its own control and monitoring analyte variations over time, will be a more sensitive experimental design for discovering biomarkers than merely comparing single samples from groups of disparate individuals. Ultimately, plasma or serum biomarkers would also represent a relatively non-invasive test, this being a significant advantage for successful clinical measurement. Some promising biomarker candidates have been produced in plasma and serum, *e.g.* [4–6]. The urine proteome has also been a focus of attention, but is not covered in this study.

A key question often asked by clinical collaborators prior to proteomic studies is a recommendation about the number of samples needed to appropriately statistically power the study, and what power is achievable with a given number of samples. Power is defined as the probability that a statistical significance test will reject the null hypothesis for a specified value of an alternative hypothesis; the ability of a test to detect an existing real effect. To estimate this, it is important to determine the variability inherent in the plasma proteome and the methodology used to analyse it. This was the key issue addressed in this study. Some theoretical considerations relating to this issue are well summarised in Horgan *et al.* [7].

There are a variety of methods suitable for proteomic investigation but this study focuses upon the use of 2-D DIGE. 2-D gel methodology, with the DIGE improvement, has several key strengths; it is excellent for following many thousands of protein forms simultaneously and over 4–5 orders of magnitude of linear dynamic range of abundance, it allows effective resolution of PTMs and their effect upon an entire protein chain and it is highly reproducible with moderate throughput [8–15]. However, one key feature of 2-D DIGE for studying plasma is that the wide range of protein abundance levels precludes effective analysis of low-abundance proteins without some form of fractionation or enrichment. This limitation may be less significant than one might assume given that there is significant precedent for specific forms of the more abundant proteins in plasma to provide useful biomarkers, particularly when used in multiplex assays [16–21]. Typically, the LOD of proteins resolvable in plasma would be between 1 and 10–100 ng/mL without and with enrichment, respectively [16]. A number of strategies for enrichment are in use and can certainly improve the range of proteins covered by 2-D or other technologies in plasma and serum [22–24]. However, detecting very low abundance proteins after the most abundant are removed can still prove problematic given the extreme dynamic range of the milieu [16, 25]. Other issues with sample depletion include increased steps in analysis which inevitably lead to increased sample losses and variability, increased time and cost, potential sample carryover and many methodologies enabling the depletion of target proteins only in their native state, conditions under which proteins bound to the depletion targets are also removed [26]. Increased variability in samples introduced by processing is often overlooked but minimising these effects is key and central to finding reliable differences between sample sets. Bearing these factors in mind, there is certainly still value in examining the unfractionated, undepleted proteome of plasma as one approach.

Several investigations have already been carried out into the variability inherent in the plasma proteome itself, using multiple technology types. For example, plasma has been studied by gel-based proteomics in a moderate number of samples with prior processing by depletion of abundant components but without 2-D DIGE methodology. This study analysed statistical power for up to 20 samples in each of two groups and suggested limits of fold changes detectable in processed plasma as well as determining intersample variability [27]. Cerebrospinal fluid (CSF), which is similar to plasma in its bulk composition, has also been studied by 2-D DIGE, after depletion processing, in a moderate number of samples with interesting results [28]. Nedelkov *et al.* [29] used an affinity-based mass spectroscopic assay of 25 targeted proteins to examine inter-individual variability in 96 individuals [29]. This detected a number of variations in protein modifications present by individual. A further study on five proteins in 1000 plasma samples extended this work [30]. Despite these studies, there has not yet been a large-scale investigation reported in the literature outlining the variability of plasma using 2-D DIGE. The aim of this study was to produce sample size plots for the detection of differences in spot normalised volume from the analysis of a large number of plasma samples. This would extend sample number beyond previous 2-D PAGE studies to test the assumptions of extending previous models and introduce the benefits of 2-D DIGE.

Samples were obtained from a population of postmenopausal women and equivalently aged males. These populations should provide suitable normal control sample sets for studies involving diseases such as prostate, breast and ovarian cancer. Samples of plasma were taken twice at an interval of 1 wk to allow both intra- and intersubject comparisons (longitudinal/paired, or cross-sectional/non-paired designs). Protein levels were quantified across all the samples and statistical sample size plots obtained from quality-controlled data for both such intra- and intersubject comparisons. This allowed the key questions to be addressed: inter- and intrasample variability for the construction of sample size plots allowing the estimation of suitable group sizes for future clinical investigations, interbatch 2-D DIGE methodology variability and evaluation of the benefits of image preregistration on data analysis. In addition to answering these questions, the methodology provides a tool for future examination of the effects of variations in sample handling (including fractionation and depletion strategies), methodology and image analysis upon the sample size plots. Any conclusions drawn must apply to unfractionated plasma, but they may also provide a starting guide for studies focussing upon enriched plasma or other sample types such as serum, ascites

and CSF, and complement previous studies on processed plasma such as the one mentioned above by Hunt *et al.* [27]. The conclusions are also limited to a site-standard collection tube type (lithium heparin), but as detailed in the discussion, effects of tube type were not under investigation in this study.

This study did not focus upon a detailed characterisation of spot variation within technical replicates, which is the variation of the method, rather using biological replicates only in constructing the sample size plots. This is because such variation has already been extensively studied and reported. Earlier work by this group has established variability inherent to the various dye-labelling relationships and quantified DIGE ratio gel-to-gel variation as being minor compared to the difference between biological samples [8]. Other subsequent studies have looked in more detail at spot variation. For example, a single (multiply aliquotted) plasma sample depleted of its major components was studied using DIGE determining spot variability [31]. CSF has also been shown to be robustly analysed by the technology [32]. Several papers by Karp *et al.* [12–15] have also established parameters for 2-D DIGE analyses and provided useful information upon the types of replicates most effective in gel-based plasma proteomics, favouring biological replicates. The paper by Hunt *et al.* [27] also contributes usefully to this topic (although not a DIGE investigation) showing that the number of biological replicates is key in 2-D work, as increasing technical replicates achieves little change in the level of effect detectable. There is also an interesting discussion of powering and replicate types in Horgan *et al.* [7]. These papers taken together strongly argue for the use of primarily biological over technical replicates using the DIGE methodology and characterise the apparent variability in the 2-D and Cy-dye labelling methods. A theme in several of these studies is that the technically most reproducible comparison among the three dyes is Cy3–Cy5 [8, 14]. Analytical distortion may be introduced using a three-colour DIGE analysis and thus, only the Cy3–Cy5 combination was used here [15]. In data analysis, we examined the effectiveness of a new gel alignment tool from Nonlinear Dynamics (NLD; Newcastle, UK). This tool was intended to improve intergel registration and thus, should reduce artefactual variability in the experiment.

2 Materials and methods

2.1 Materials

Plasma collection tubes (Monovette lithium heparin tubes) were from Sarstedt (Leicester, UK). Axygen microcentrifuge tubes were from Thistle Scientific (Glasgow, UK). Protein Assay Reagent and bromophenol blue were from BioRad (Hemel Hempstead, UK). Cy2, Cy3, Cy5, Pharmalytes pH 3–10, DryStrip Cover Fluid (mineral oil), SDS and IPG Immobiline DryStrips (pH 3–10 NL 24 cm) were from GE Healthcare (Amersham, UK). a2DE Optimizer reagent kits were from Nextgen Sciences (Cambridge, UK). CHAPS, lysine, CBB G-250, Tris base, DTT and IAA were from Sigma (Dorset, UK). GelBond PAG film was from Lonza (Wokingham, UK). Urea (ultra pure) was from ICN Biomedicals (Aurora, USA). Glycerol and thiourea were from Fisher (Loughborough, UK).

2.2 Sample collection and processing

Plasma samples were collected from volunteers in the AstraZeneca CPU facility (Alderley Park, UK). Volunteers attended a screening visit (visit 1), including a prestudy medical questionnaire, in the 21 days prior to the first study sample collection. There were two sample collection visits (visits 2 and 3), separated by 7 days. At each of these visits samples were taken for proteomic analysis between 9 am and 11 am from volunteers that had fasted for at least 14 h previous, with the exception of water *ad libitum*.

At each of visits 2 and 3, prior to the day of collection, volunteers were asked to record their food and fluid intake in patient diaries for 24 h before attending the CPU. The volunteers were also asked to ensure that their dietary intake was similar, or preferably replicated, in the 24 h preceding visits 2 and 3. At visits 2 and 3 volunteers donated 162 mL of blood, which was split into three aliquots of 54 mL to provide plasma, serum and platelet-stripped plasma. Volunteers also donated 150 mL of mid-stream urine. The plasma is of relevance in this report and subsequent processing details refer to this. Venous blood (54 mL) was collected into 6×9 mL tubes containing lithium heparin anticoagulant. These samples were centrifuged within 30 min of collection at $1500 \times g$ for 10 min at 4°C. The supernatants were transferred using a plastic pipette to individually labelled sample tubes and shipped on wet ice to the proteomics human tissue laboratory within 1 h of leaving the volunteer. They were then processed for storage by aliquotting into polypropylene Axygen Maxymum Recovery MC-175C microcentrifuge tubes and stored at −80°C. The mean time from leaving the volunteer in the CPU to −80°C storage of aliquots in the laboratory was 52 min and the median 51 min (36–84 min range, 10th percentile 43 min, 90th percentile 63 min).

Thirty-one male subjects were sampled, and 30 female. The extra male recruited was to reach 30 paired samples, as one of the earlier donors only provided the first, visit 2, sample. The average age of the male donors was 57.35 years and the average age of the female donors was 59.13 years. For inclusion and exclusion criteria for the study, and full study restrictions, see Appendix 1 in the Supporting Information. The study was performed in accordance with the ethical principles in the Declaration of Helsinki, Good Clinical Practice (GCP), and applicable regulatory requirements, with written informed consent obtained from all volunteers.

2.3 Sample assay

A modified BioRad Bradford protein assay format was employed for protein concentration (Joanne Hartley, AstraZeneca, personal communication). Samples were assayed

against a BSA standard curve (Pierce). The average protein concentration was 69.3 mg/mL.

2.4 Labelling for 2-D DIGE gels

It should be noted that throughout the 2-D DIGE process, because of practical issues surrounding the maximum number of samples that could reasonably be handled at the same time, several steps required the subdivision of samples into batches. In all cases, batch details were recorded for meta-analysis. The same batch of reagents was used throughout the study for each step of the analysis to minimise variation.

Cy3 and Cy5 labelling of proteins using the lysine minimal-labelling 2-D DIGE chemistry (GE Healthcare) was carried out for plasma essentially as described previously [8]. The protocol used here differed in the lysis buffer used, which was 7 M urea, 2 M thiourea and 4% w/v CHAPS. 1 M Tris-HCl solution (pH 8.5) was used to buffer to a final pH of 8.5. All individual samples were labelled with Cy5. Samples were all also pooled by equal protein concentration and Cy3 labelled to produce an internal standard for DIGE. Protein (75 μg) was labelled *per* dye *per* gel. This corresponded to an average of 1.08 μL of starting plasma taken and loaded *per* channel *per* gel after labelling. The final buffer composition for IEF was adjusted after labelling to 7 M urea, 2 M thiourea, 4% w/v CHAPS, 65 mM DTT (1% w/v) and 0.8% v/v Pharmalytes pH 3–10.

2.5 IEF

For 2-D DIGE gels, Cy dye-labelled protein was applied to 24 cm 3–10 NL IPG strips and strips were allowed to rehydrate overnight under a mineral oil overlay, followed by IEF using the a2DE Optimizer system (Nextgen Sciences) for a total of 110 kV·h in a gradient rising-voltage protocol. IPG strips were run in singlicate, with each visit of each individual on a unique strip. This gave 121 IPG strips in total; the odd strip being visit 2 of sample 49, the non-paired sample.

For preparatory gels, 2 mg of unlabelled protein was prepared by mixing an equal amount of every sample to a final concentration of 10 mg/mL in the IEF buffer described above. This corresponded to 28.9 μL of plasma. This mixture was then used to rehydrate strips and focussed as for DIGE strips on the a2DE Optimizer. Four preparatory IPG strips were prepared, two each for silver and Coomassie stained 2-D gels.

2.6 Second dimension SDS-PAGE

IPG strips were equilibrated in a two-step protocol in pH 6.8 equilibration buffer (100 mM Tris, 6 M urea, 30% v/v glycerol, 1% w/v SDS) containing 1% w/v DTT for 15 min, followed by 4% w/v IAA for 15 min and applied to vertical 10% acrylamide Laemmli SDS-PAGE gels (10.27%T, 2.6%C) for the second dimension separation using a modified ESA Investigator gel system. A bromophenol blue dyefront was used to monitor electrophoresis and gels removed from the tanks upon migration of the dyefront from the gel. For preparatory gels only, the gels were cast against GelBond PAG film.

2.7 Image scanning and processing for 2-D DIGE gels

2-D DIGE gels were scanned using two Typhoon 9400 scanners (GE Healthcare) at 100 μm resolution and saved as .gel files for analysis. Scan settings were first optimised on each machine using 1000 μm prescans to obtain a maximum signal of 80 000–95 000 counts (of 100 004 maximum possible) *per* channel, as assessed using ImageQuant v5.2 (GE Healthcare). The two most abundant spots of albumin were allowed to saturate, to increase visualisation of less abundant species. Therefore, the PMT value was set on the next most intense protein chain, which was serotransferrin (Fig. 1). For each batch of gels, the optimal values from the initial batch were tested again on the first gel to ensure correction for any variation in machine performance. If necessary, the PMT voltage was adjusted slightly to compensate and retain the dynamic linear range at the above levels. The PMT voltage used was recorded for analysis of any effects of the variation. The default values were Cy3 460 V and Cy5 445 V on the first Typhoon, and Cy3 460 V and Cy5 455 V on the second. Images were cropped using ImageQuant Tools v3.0 (GE Healthcare) to exclude any region outside the fluorescent spot pattern profile, and taken for analysis.

2.8 Staining and scanning of preparatory gels

Protein visualisation was essentially either by colloidal Coomassie or silver staining [33–35]. Circular 5 mm diameter adhesive visible reference markers (GE Healthcare) were

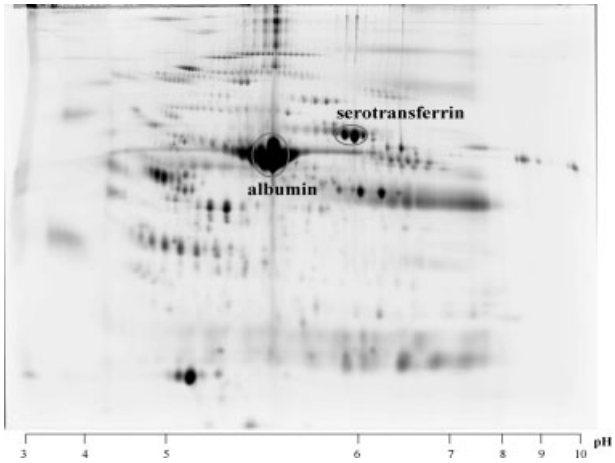

Figure 1. Spots used for PMT optimisation during gel scanning. The reference gel image for the experiment, from sample 49 visit 2, is shown. The albumin forms allowed to saturate during scanning are marked along with the serotransferrin chain used to optimise the image signal.

attached to the gel support film in each case prior to scanning for later triangulation purposes. Gels were then scanned using the ImageScanner (GE Healthcare) to produce 8-bit.tif files.

2.9 Image analysis of 2-D DIGE gels

2-D DIGE gel analysis was carried out using software from NLD. This was done by two methods, A and B:

(A) Progenesis 240 analysis*:* Image analysis was carried out with Progenesis discovery version 240 (PG240; 2006) with a cross-stain analysis approach.

(A-i) Spot detection parameters*:* The spots in the first gel analysed were initially auto-detected and assessed using an area filter to determine a cut-off for dust or other particulates. An area of 50 pixels was determined as the threshold to remove dust in the full analysis.

(A-ii) Selection of base gel for reference gel construction*:* The most suitable base gel to initiate the reference gel was then selected on the grounds of having the largest number of genuine spots and a well-defined profile. The visually very best resolved gels were all auto-detected, being user selected for this step by virtue of having the apparent tightest focussing and least distortion of the 2-D pattern; this is not to imply that any gels had any significant amount of such interference but inevitably, some were clearer than others. Eight gels were selected for assessment. These were found to be spread throughout the set, and not focussed in any particular batch. The spot patterns of those gels were edited to remove any extraneous material and ensure accurate spot detection and the gel with the highest number of remaining spots was selected as the base gel for the reference. This was the gel corresponding to sample 49 visit 2, with 2057 spots.

It was appreciated that perhaps the gel requiring the least amount of alignment of other gels to it, rather than the most populous one, would make the best starting point for generating a reference gel. However, the Progenesis software did not enable such a selection, and the gel having the most number of genuine spots was adjudged to be the best candidate for the base gel in this situation.

(A-iii) Image set analysis*:* A full cross-stain experiment was then run in accordance with the parameters established in (A-i) and (A-ii) above. Only spots with an area of 50 U or greater were detected, and the gel corresponding to sample 49 visit 2 was defined as the base gel for the reference. Other auto-detection parameters were as default, and no average groups were defined. The background subtraction was therefore the 'Progenesis' method and the normalisation was 'Ratiometric'. Editing was then carried out as follows. The two most abundant forms of albumin, that had been allowed to saturate when scanning with the Typhoons, were removed from the analysis. Any extraneous material at the edge of the gel was removed. Finally, detected spots were checked and corrected if necessary using the editing tools. Although it is reasonable to assume that this introduces user intervention, the only edits carried out resulted from clear detection errors and were applied consistently and therefore represented clear improvements in the analysis. The gels were then renumbered, renormalised and background subtracted as appropriate after editing. The reference gel was then regenerated in light of the editing, leading to a master map with 4534 spots incorporating contributions from all gels. The average number of spots *per* individual gel was 1871.

(A-iv) Data export*:* The comparisons window was used to display the normalised volume for every spot in every gel (the Cy5 normalised volume characteristic) against the reference master spot match number. This table was exported as a text file for statistical analysis.

(B) TT900S2S-Progenesis 240-SameSpots analysis (TT900S2S/PG240/SS)*:* This was carried out as part of a collaboration whereby NLD analysed the image set data using software systems in development and at the time unavailable commercially. This was carried out as preliminary data suggested that the improved warp algorithm of TT900S2S, coupled with standardised spot outlines across the whole image set made possible by the SS functionality, would improve the quality of the data analysis, reducing the variability in the dataset. Images were first aligned using TT900S2S. Within this gel set, the base gel for the reference warp image was set to be the same gel as for the previous analysis (A). The aligned images were saved as.tif files and made available to AstraZeneca for in-house analysis with PG240 v2006, which had the SS capability for ready-aligned images. The analysis was set up in exactly the same manner as for the PG240-only analysis using the aligned images and the same parameters. However, the final editing was much simpler owing to the use of SS, which would standardise all spot outlines from the reference gel. Therefore, the only editing was on the final reference gel directly. The two forms of albumin that had been allowed to saturate were removed, and extraneous material was removed. No other edits were performed. SS was then applied, involving the renormalisation of the gels as part of the process. The resulting reference gel had 4615 spots. Since SS had been applied, all gels had the same final number of spots.

(B-i) Data export*:* A comparison-window based table of normalised volume in every individual Cy5 channel was exported exactly as for the PG240-only experiment. A second table incorporating raw spot volumes in both channels of every gel was also produced.

2.10 Statistical analysis: Quality control (QC)

Postregistration data tables of normalised volumes were analysed using the R statistical package version 2.3.1 to 2.5.0 [36]. A range of data plots were produced to highlight spots for verification and to assess the variability of the data. Batch-to-batch variability was assessed by methodological step (*e.g.* labelling, IEF and gel batch) as described in the legends of Fig. 6 and Table 2. This included all 121 gels. The postregistration volume data, together with the QC and explora-

tory data analysis scripts used, are available as Supporting Information.

2.11 Statistical analysis: Sample size plots

After exclusion of any spots failing QC, the remaining spots were analysed. Only samples present at both time points were included in the analysis (omitting the single sample from subject 49, whose second sample was missing). Both data sets were analysed for intra-individual variability. However, after determining from the results that TT900S2S/PG240/SS was superior to PG240 alone, only the TT900S2S/PG240/SS data set was analysed for inter-individual variability. For the original data set without SS treatment (A), only spots present and matched in at least two-thirds of the pairs were included, that is in $\geq$80 gels. For the TT900S2S/PG240/SS (B) data analysis, such filtering was not necessary, as all spots were technically ubiquitous. Males and females were not separated for these analyses.

For each gel spot, the ratio values of the normalised volume intensity measures were log transformed before being subjected to analysis for variability. For each gel spot, a separate ANOVA model was fitted. The model included subject as a factor and the log transformed intensity measure as the response variable. The residual mean squares gave intra subject variance, and hence the square root of the mean squares gave the SD. This SD was used in power calculations to calculate sample sizes for each analytical method (A) and (B) above (see results). The sample sizes were presented as plots giving the summary of the number of samples required to measure a two-fold difference, with 0.05 level of significance and 0.8 (or 80%) power, applying the Bonferroni correction for multiple tests. The SSIZE module of the bioconductor package was used in calculating the sample sizes and in generating sample size plots [37, 38]. It should also be noted that the Bonferroni correction is overstringent given that protein features on a gel are not all truly independent, and may be coregulated members of a chain of post-translationally modified forms; however, it was assumed that overstringency would be more appropriate for determining the number of samples required for clinical studies (and potentially overestimating) than risking under-stringency. Possible alternatives are discussed in Karp *et al.* [15]. The characteristics were selected to be representative of typical proteomic experiments; it is recognised by the authors that this only represents one set of parameters, but each parameter is commonly used in the field and this invariant parameter set still facilitated direct comparison of inter- and intra-individual variability and methodologies.

To elaborate on the rationale for sample size plots, suppose that subjects for some future study will be recruited from two groups, which might differ by a demographic characteristic, clinical phenotype or pharmacological intervention (to suggest just three examples). Suppose that between these two groups there is a two-fold difference in the population mean spot volume, for some protein spot measured by 2-D DIGE, while within each group the subject-to-subject variability for this spot remains the same as that seen in the present study. Whether this two-fold difference will reach formal statistical significance in the future study will depend on its observed sampling variability within each group and on the two sample sizes.

If the two-fold difference occurs in a 2-D gel spot which has low sampling variability, then only a small sample size will be needed to reach formal statistical significance, while if it occurs in a spot with large sampling variability, a much larger sample size might be required. Sample size plots use the range of *per* spot sampling variabilities observed in the current experiment to calculate the number and fraction of spots in which a hypothetical two-fold difference would achieve statistical significance ($\alpha = 0.05$) with reasonable power (0.8), should it occur in a future experiment, as a function of the future experiment's sample size.

All sample size statistical calculations were performed using R statistical system (version 2.5.0) [36].

The spot parameters output from PG240 for each analysis and QC analysis script are available in the Supporting Information. The gel scans analysed are hosted at: http://www.fixingproteomics.org/datasets/.

3 Results

3.1 Power analysis and variability: Sample size plots

Figure 2 was constructed for PG240 only (analysis A) and Fig. 3 for TT900S2S/PG240/SS (analysis B). Figures 2 and 3A refer to the within subject sampling variation observed from two blood draws 1 wk apart, *i.e.* a paired study, with Fig. 2 being calculated for the PG240-only method and Fig. 3A being calculated for the TT900S2S/PG240/SS method. Figure 3B refers to sampling variation between subjects, *i.e.* a non-paired study, for the TT900S2S/PG240/SS method. The vertical axis of each plot in Figs. 2, 3A or 3B shows the number and fraction of all spots in the reference gel image which have small enough sampling variability so that a two-fold change in spot volume will likely show statistical significance in a future study using two groups of the specified size. The data are summarised numerically in Table 1. Considering Fig. 2, if 20 samples at visit 2 were compared to a paired set of samples from visit 3 by PG240 only, a two-fold mean abundance change would be detectable in any of 731 spots with the requisite power should it occur. If only 10 paired samples were available, then the spot set in which one could detect such an effect would be reduced to 589. With TT900S2S/PG240/SS, considering Fig. 3A with paired intra-individual studies, detection of differences in a large number of spots (1742) is possible with 20 samples, and the gradient of the sample size plot is noticeably beginning to reduce, indicating lessening of the benefit of adding more samples to the study. For non-paired independent groups in

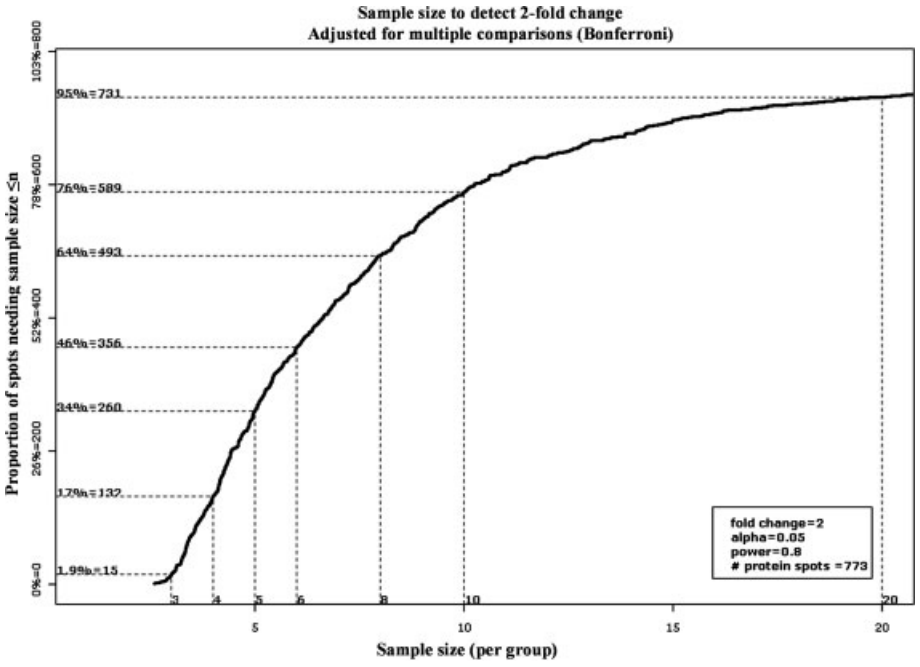

Figure 2. Sample size plot for the PG240-only approach. The number of spots in which a difference of the specified magnitude and α value would be detectable with sufficient power, should it occur, is plotted against the sample size in a paired or longitudinal study. Parameters are a fold change of 2, an α of 0.05 and a power of 0.8.

Fig. 3B, differences in 1206 spots could be detected with 20 samples *per* group (506 fewer spots) and the curve has not yet begun to plateau out.

It is evident from comparing the data in Figs. 3A and B that inter-individual spot variability is greater than intra-individual spot variability, as expected. This has an impact on study design. To illustrate this, in a paired comparison, that is in an experiment comparing a particular time-point (say baseline) of a patient with a second time point (say after treatment) of the same patient or a cross-over design experiment where two different treatment regimes are compared by patient by patient basis when a patient will cross-over to the second treatment after an initial treatment, the number of subjects required will be less. This is mainly because in such an experimental design, the sample size calculations are done using intrasubject variation. On the other hand in an experimental design involving two independent groups (say treatment A group *vs.* treatment B group) the number of subjects will be greater as the sample size calculations are done using inter subject variation.

A visual examination of the spots on the 2-D gel in which differences can be detected by up to 20 samples using the TT900S2S/PG240/SS approach (Fig. 4) indicates that the majority of spots easily visible to the naked eye are included. Low raw-volume spots or those on the threshold of detection are more variable (data not shown).

The individual spot variability data used to generate the sample size plots was also plotted and can be seen in Fig. 5 as a frequency distribution. This was done for the TT900S2S/PG240/SS data set for all 3935 spots included in the sample size analysis: spot normalised volumes were taken (a log 2 transformation had been applied as part of the overall process), and the SD noted on a spot-by-spot basis. Histograms of the distribution of spot SDs are shown. For comparison purposes, the mean SD of the log 2 normalised volumes for intra-individual variability over the entire data set of 3935 included spots was 0.72. The mean within this for the 1742 spots in which differences could be detected between two paired groups by 20 samples was 0.23, and the mean for the remainder unable to be differentiated was 1.11. The mean SD for inter-individual variability over the entire data set of 3935 included spots was 0.95. The mean within this for the 1206 spots in which differences could be detected between two paired groups by 20 samples was 0.34, and the average for the remainder unable to be differentiated was 1.22.

3.2 Power analysis and variability: Batch variation and metadata

Variation across batches in the TT900S2S/PG240/SS workflow was plotted in Fig. 6. The median normalised spot volume was calculated for every gel. For each experimental variable, gels were grouped into their batches. The median value of each batch was plotted (the median of the median normalised spot volumes). The figure shows that there is a shift in median normalised volume, by group, from batch to batch but that such a shift is not systematic for any factor. This suggests the variation may be stochastic.

An ANOVA covariate analysis was performed to calculate the significance of these changes. The input data were the log gel median normalised volumes, grouped by batch. It is evident that some of the factors show significant but minimal batch-to-batch variation, as shown in Table 2. This table shows the ANOVA result, and the largest difference between

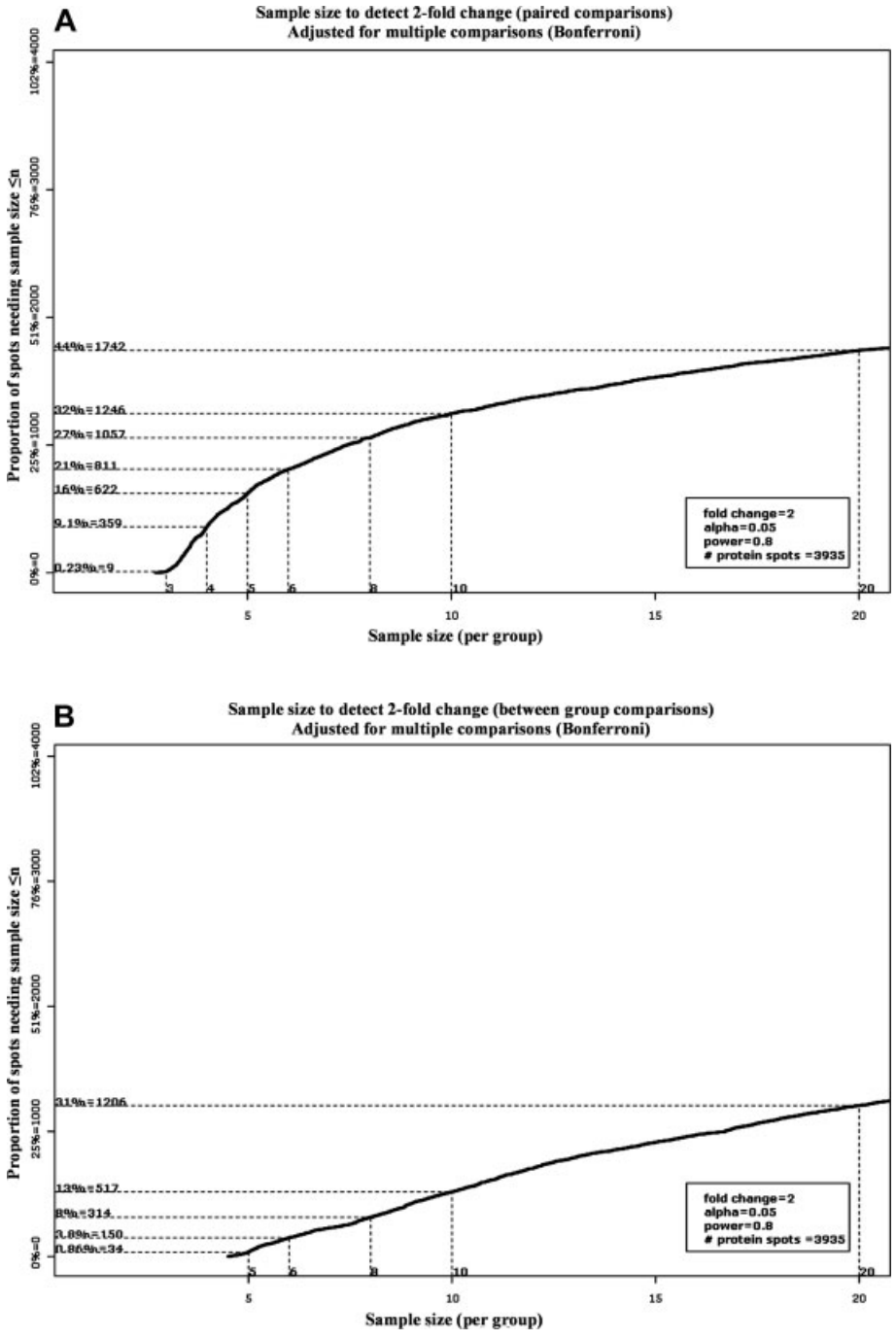

Figure 3. (A) Sample size plot for the TT900S2S/PG240/SS approach (intra-individual variability). The number of spots in which a difference of the specified magnitude and α value would be detectable with sufficient power, should it occur, is plotted against the sample size in a paired/longitudinal study. Parameters are a fold change of 2, an α of 0.05 and a power of 0.8. (B) Sample size plot for the TT900S2S/PG240/SS approach (inter-individual variability). The number of spots in which a difference of the specified magnitude and α value would be detectable with sufficient power, should it occur, is plotted against the sample size in a non-paired study with groups of equal sizes. Parameters are a fold change of 2, an α of 0.05 and a power of 0.8.

the means by batch group of the non log-transformed gel median normalised volumes. The magnitude of the range of the means of the batch medians (max difference) is minimal compared to the two-fold change under investigation. For all 4615 spots on the reference gel, as shown in the table, the largest range was for 2-D gel batch and Cy5 PMT value, with a change corresponding to a 0.089-fold alteration (8.9% of a two-fold change). This was for the widest range possible in the most variable factor, a worst-case scenario.

3.3 Power analysis and variability: Software comparison

The largest number of spots detectable as differences between two groups was achieved by the TT900S2S/PG240/SS analysis approach (Table 1). Intra-individual variability was compared for both data analysis approaches (A) and (B), and at a sample number of 20, the PG240-only approach (A) would detect differences in 731 spots of 773 analysed (shown

Table 1. The number of spots in which a difference in normalised volume could be detected by each analytical approach at a range of sample sizes

Analytical approach	Sample size							Total spots
	3	4	5	6	8	10	20	
PG240 intra-individual	15 (1.9)	132 (17.1)	260 (33.6)	356 (46.1)	493 (63.8)	589 (76.2)	731 (94.6)	773
TT900/PG240/SS intra-individual	9 (0.2)	359 (9.1)	622 (15.8)	811 (20.6)	1057 (26.9)	1246 (31.7)	1742 (44.3)	3935
TT900/PG240/SS inter-individual	0	0	34 (0.9)	150 (3.8)	314 (8.0)	517 (13.1)	1206 (30.6)	3935

The sample size calculation results for each analytical method are summarised by the type of variation studied. Absolute spot numbers are given along with the total number included in the power analysis. The percentage of the total number of included spots is denoted for each value in parentheses.

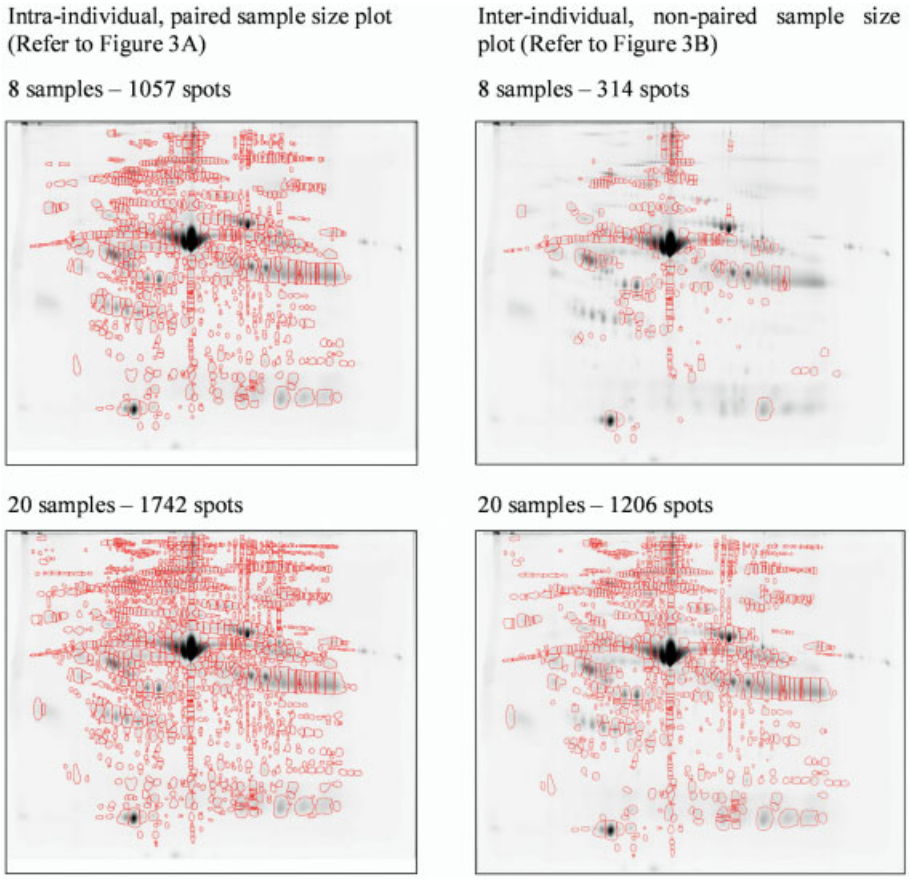

Figure 4. Protein spots in which a difference could be detected between two groups using the TT900S2S/PG240/SS method by sample sizes 8 and 20 (chosen for illustration). Spots are shown on the reference gel image.

on Fig. 2), whilst this number improved to 1742 spots of 3935 analysed with the TT900S2S/PG240/SS approach (B) (shown on Fig. 3A). The TT900S2S/PG240/SS approach shows a 2.4-fold improvement over the 'PG240-only' analysis, at a sample size of 20, in the absolute numbers of spots for which statistical significance might be achieved in a future study, even though this represents a smaller fraction of the total number of spots on the 'TT900S2S/PG240/SS' reference gel.

A visual assessment in PG240 was made of the 359 least variable spots across all gels, in which differences could be detected by 4 samples *per* group. Over 99% were genuine spots and detected correctly in all gels. This suggested that of spots in which differences could be detected by the analysis, at least the least variable end of the spectrum were genuine features, and correctly detected and aligned.

4 Discussion

In this study, we have built on previous work to show that 2-D DIGE provides a robust platform for biomarker dis-

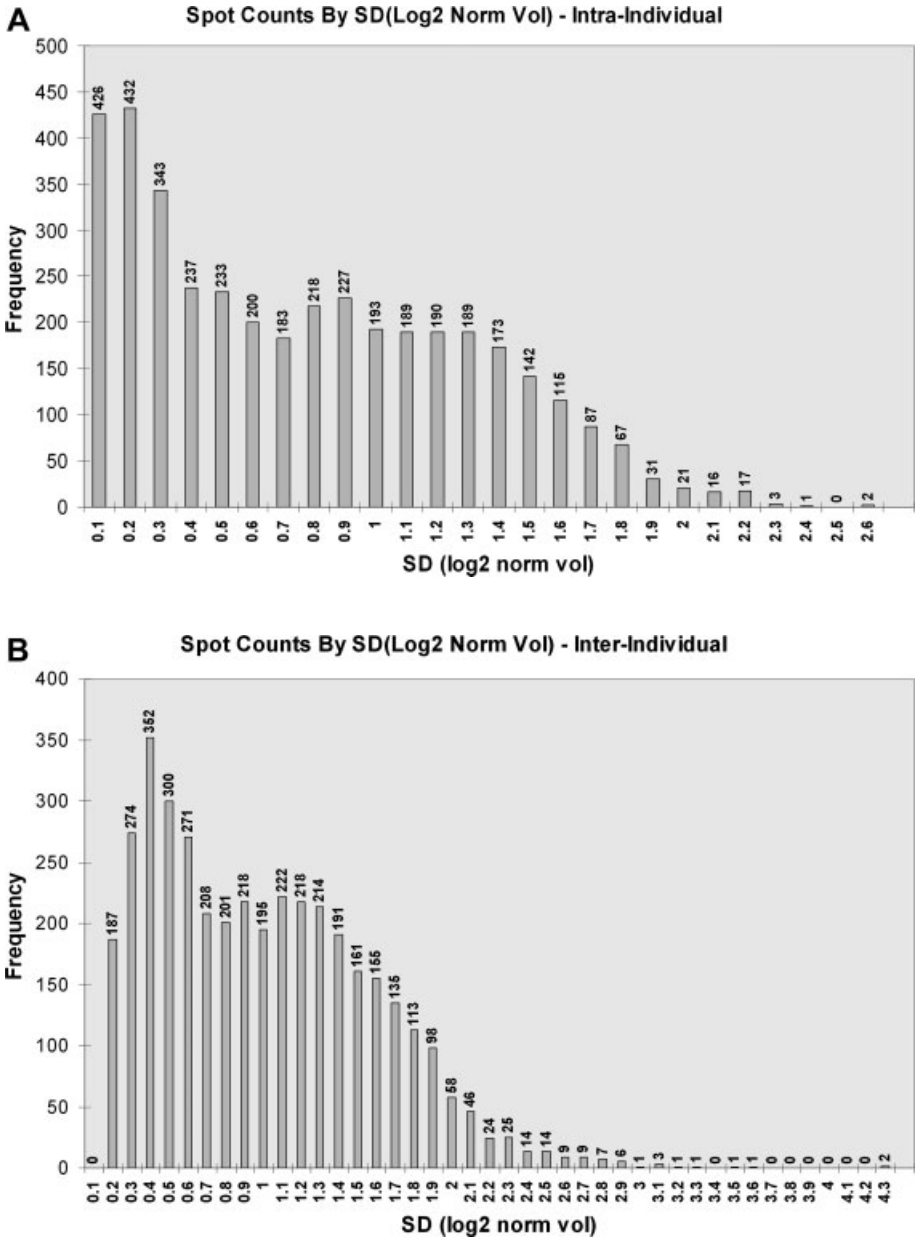

Figure 5. (A) Frequency distributions of spot variability for the TT900S2S/PG240/SS data set. The SD of spot normalised volumes after a log 2 transformation was plotted for all 3935 spots included in the sample size analysis, in this case for intra-individual variability across the data set. All paired samples (120 gels) were included. The bins are upper limits for each class. The counts are added above each bar. (B) Frequency distributions of spot variability for the TT900S2S/PG240/SS data set. The SD of spot normalised volumes after a log 2 transformation was plotted for all 3935 spots included in the sample size analysis, in this case for inter-individual variability across the data set. All paired samples (120 gels) were included. The bins are upper limits for each class. The counts are added above each bar.

covery in human plasma with a limited number of samples. The number of samples needed to discover proteomic alterations with appropriate power is not great ($\leq$20 *per* group) and we would be confident of detecting a two-fold change in abundance, should it occur, in over 1200 spots with this many samples *per* group, even in non-paired studies. In paired or translational studies, where longitudinal sampling is performed, the number of spot differences detectable rises to over 1700. This represents a simultaneous assay and analysis of a great many protein spots showing the power of 2-D DIGE. This is also despite the application of the very stringent Bonferroni statistical correction for multiple testing in large data sets. The amount of each sample needed when dealing with unfractionated plasma is not great and indeed on the order of only a few microlitres.

The authors stress the evident gain in power associated with an intra-individual, longitudinal or paired, design as described in the results. Naturally such a method requires as rigorous control of sample collection between sampling points as between samples. In this study, presampling behaviour and collection protocol for both visits were standardised to minimise any dietary, diurnal or other confounding factors as much as is possible. It is clear that under such conditions a marked improvement in study power is achieved, to such an extent that the authors would recommend such a design wherever applicable and possible.

Table 2. Variation by metadata batch, and significance of that variation, for the TT900S2S/PG240/SS dataset

Cofactor	*p*-Value	DF	Max. difference
Dilution batch	0.462	6	0.028
Labelling batch	0.462	6	0.028
IEF batch for IPG strip	0.040	11	0.062
X2D gel batch	0.009	12	0.089
Overlay buffer	0.035	2	0.030
Equilibration buffer	0.051	3	0.035
Butanol	0.110	4	0.030
DTT	0.221	1	0.018
IAA	0.415	1	0.009
X2D gel tank	0.691	1	0.003
Position in tank	0.217	4	0.029
Typhoon used (LHS/RHS)	0.038	1	0.028
Number of scan pair	0.375	1	0.008
PMT V Cy5	$<$0.001	5	0.089
PMT V Cy3	0.410	3	0.034

Effect of experimental cofactors for the TT900S2S/PG240/SS analysis. For each cofactor, a linear model was computed to predict the median normalised signal by gel as a function of the cofactor. The *p*-value is the *p*-value for the *F* statistic of the related ANOVA, the DF represents the degrees of freedom in the variable, and the max difference is the largest absolute difference in coefficients between any two levels of the cofactor, assessed by the means of the gel medians by batch. Although some cofactors show a strongly significant difference in gel medians, the magnitude of these differences is small ($<$10%) relative to all cut-offs used for individual spot fold changes. This was calculated on all 4615 ubiquitous spots prior to any exclusion of spots by QC.

It should be noted that it would also be possible to use average spot variability to calculate general sample sizes required for a particular fold change of alteration at known power and α value. This was not done here as we desired instead to dissect the proteome in variability terms using a particular threshold as a referent. In clinical studies, a smaller threshold change might well be employed; the values used here are merely typical proteomic parameters for illustrative purposes.

Batch-to-batch variability is not a significant issue with 2-D DIGE proteomics, at least within the same experimental system and site. Combined with the studies cited in the introduction, this work shows that the technology platform is robust for determination of proteomic changes in plasma. The worst-case variation was far lower than a typical fold-change cut-off employed in proteomic analyses, being less than 10% of it, and that was calculated over a number of batches that would be disproportionately large for most studies given the sample size plot results. It would therefore be reasonable to suppose that the worst-case variation in 20-sample-per-group studies would be much lower.

The data analysis method could be further improved by applying a filter to remove spots of low absolute volume after QC. It was observed in preliminary investigations in this and other data sets in-house that such spots are among the most variable and that simple volume filtering can improve power (data not shown). This has also been shown in other studies [12]. However, such a filter would naturally require optimisation by sample and gel chemistry type. The results presented show that the method is robust even without such a filter and that low absolute abundance spots in which differences could not be detected are normally not part of the notable spot profile (Fig. 4). If the low absolute volume spots were targeted to drill down and extract any useful information present, on the other hand, larger numbers of samples *per* group than recommended here might be required given their variability.

The power calculations in this study are tailored towards the rapid analysis of unfractionated plasma which requires very little sample and is a fast, cheap and robust method of proteomic analysis where the analytical variability is minimised. It is appreciated that 2-D DIGE analysis of such a sample is likely to only observe the more abundant proteins in plasma and thus, is more likely to produce biomarkers through the appearance of epiphenomena, or host response to insult, rather than shed tissue biomarkers given the lack of enrichment. This is because the lowest abundance proteins likely to be resolved in plasma based upon practical detection limits determined in Shaw *et al.* [9] and the loadings used upon our DIGE gels are just below 1 μg/mL, which covers most of the classical plasma proteins and only just encroaches upon the region of the proteome including tissue leakage products [9, 16]. There is currently great interest in pushing the lower detection limit of proteins in plasma but it must be remembered that such activities are likely to increase the variability of the sample, increase costs greatly and thus, may be counter productive. It is entirely possible that there are important biomarkers of disease to be discovered among the more abundant proteins providing the investigations are performed with sufficient rigour and examples of abundant biomarker candidates are noted [16–21].

Improved gel registration with TT900S2S and the use of an SS design improves the quality of data obtained, with more spots being present and included in the analysis and more differences detectable between two groups with sufficient power. That more spots are included is expected, since the entire reference spot set is passed out to every sample gel, and so no filtering by occurrence rate is applied. In the non-TT900S2S-aligned set, the requirement that a spot must have been present in two-thirds of sample gels, to reduce the contribution of spurious spots and mismatches, results in many spots being excluded and leaves only the better-matched spots for analysis. Also, it may be expected that the improved prewarping and alignment algorithm of TT900S2S would reduce variability through improving correct matching, and hence cause more spots to be correctly matched and detected as differences, due to a reduction in spurious matching-induced variability. Finally, standardisation in spot outlines may also contribute to this effect. However, it was

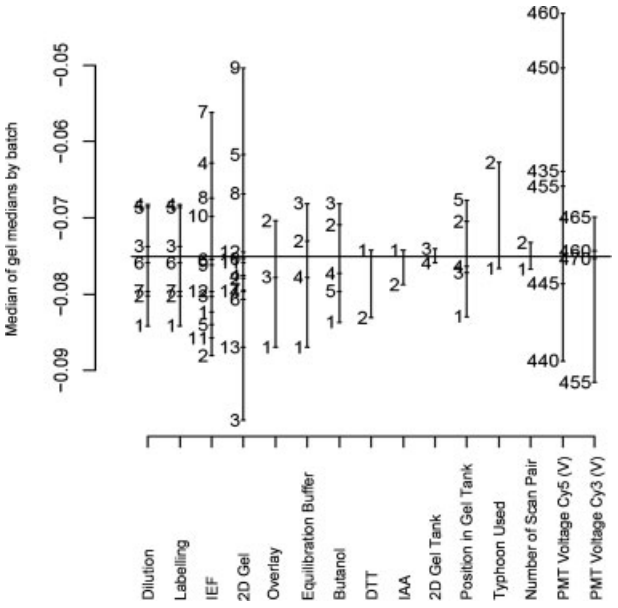

Figure 6. Gel metadata by batch. Median normalised spot volume was calculated for each gel analysed in the TT900S2S/PG240/SS data set. For each batch factor applicable to the 2-D DIGE process, the gel medians were grouped by factor and batch, and the median of each batch was plotted. Solution batches were all made up with the same reagents; they were different preparation batches. The voltages are given directly for PMT settings. The median axis is plotted relative to a line representing the median of all gels' median normalised spot volumes. This was carried out for all 4615 ubiquitous spots prior to any exclusion of spots by QC.

not possible in this experiment to resolve the various levels of contributions of said effects. These points, taken together, explain why a lower proportion of spots remaining in the analysis would have detectable differences between two groups when using TT900S2S/PG240/SS, but a much higher absolute number could be differentiated. The far greater absolute number of spots that could be differentiated, coupled with the data suggesting that most such spots are genuine, leads the authors to commend the TT900S2S/PG240/SS approach.

It is important to note that the SS algorithm applied in this investigation is that provided as a final-step option within the current build of PG240, Progenesis Discovery. As yet, the latest 'full' version of SS, which has a number of steps incorporated into a single workflow ending with a PG240 output, has not been utilised in this investigation. This would allow codetection of spot boundaries using an updated algorithm before the PG240 step, and may improve matters further, but is beyond the scope of this work for the present. Another development under evaluation is the use of an automatic alignment algorithm developed at the University of Manchester by Daniel Allen and Jim Graham. This program aligns gels completely objectively without any user input and is showing promising results, which will form the basis of a future publication.

It would also be interesting to determine the effect of varying the reference gel on the data. This could not be carried out for reasons of time, but others may wish to examine this and other effects using the images we will make available from this work. The automatic alignment algorithm from Allen and Graham could obviate the need for such an imposition of a reference at the warping/alignment stage, although one would be required for the PG240 analysis.

It would be highly desirable to identify all the proteins present. This work is underway and will allow the powering of studies according to protein class. These data could inform clinical decisions on study size by considering the biology under investigation. This was beyond the scope of this study, but the methods of analysis employed allow for this to be added later. Interestingly, Fig. 5 appears to suggest there may be two overlaying curves in terms of spot variability, with an initial peak of low-variability spots being followed by a trough and possible further peak, perhaps representing a shift in the variability distribution between more and less abundant species. Further analysis may focus on this and on biological identification of the spots apparently belonging to each distribution.

The methodology described herein provides a tool with which to investigate all steps in the analytical procedure such as abundant protein depletion, 2-D gel run conditions, image processing algorithms and statistical routines in terms of their effect on overall sample variability and the shape of the sample size plots. It is important to standardise handling and assess the contribution of different preparatory methods to the 2-D DIGE analysis of plasma, and this has been initiated notably by the HUPO consortium amongst others [39, 40]. However, this topic was not under investigation in this study. Instead, collection conditions, materials and timescales were rigidly controlled and a single collection centre used to minimise any such variation, as would be expected for a study providing a template for further studies in the pharmaceutical environment. It is accepted that this study must be regarded as a guide for further studies given that the factors above will vary in different environments. However it certainly allows an evaluation of the technology and its applicability to the analysis of clinical biofluid samples, and can provide a starting point for the appropriate sizing of studies of such samples in general.

In summary, gel-based proteomic experiments need not be large-scale to achieve high statistical power if samples are collected in a controlled fashion ideally using a longitudinal sampling design. Technical reproducibility is good between batches, and preregistration methods improve the data obtained from such analyses. Future work will evaluate alternative software workflows, produce a reference protein map with variability data from all identified proteins, and potentially look at prefractionation of plasma prior to gel-based analysis. Finally, the data (at postregistration PG240 output level) are available for others in the Supporting Infor-

mation, and the gels analysed are available for others to analyse the set as desired and interrogate the unfractionated plasma proteome at http://www.fixingproteomics.org/datasets/.

The authors would like to acknowledge Helen Redding, Wendy Webb and their team in the CPU unit, for clinical logistics, sample collection and processing. Joanne Hartley, Graeme Smith, John Williams, Ian Lee, Sally Price and Hannah Rendell assisted with sample handling, aliquotting and storage in the laboratory.

Conflict of interest statement: David Jackson, Athula Herath, Jonathan Swinton, Raj Chopra, Andrew Hughes, Kevin Cheeseman and Robert Tonge have no financial conflicts of interest around the publication of this manuscript. David Bramwell is the Technical Director of Nonlinear Dynamics Ltd.

5 References

[1] Chen, G., Gharib, T. G., Wang, H., Huang, C. C. *et al.*, Protein profiles associated with survival in lung adenocarcinoma. *Proc. Natl. Acad. Sci. USA* 2003, *100*, 13537–13542.

[2] Cho, W. C., Contribution of oncoproteomics to cancer biomarker discovery. *Mol. Cancer* 2007, *6*, 25.

[3] Okano, T., Kondo, T., Fujii, K., Nishimura, T. *et al.*, Proteomic signature corresponding to the response to gefitinib (Iressa, ZD1839), an epidermal growth factor receptor tyrosine kinase inhibitor in lung adenocarcinoma. *Clin. Cancer Res.* 2007, *13*, 799–805.

[4] Cho, W. C., Yip, T. T., Yip, C., Yip, V. *et al.*, Identification of serum amyloid A protein as a potentially useful biomarker to monitor relapse of nasopharyngeal cancer by serum proteomic profiling. *Clin. Cancer Res.* 2004, *10*, 43–52.

[5] Papadopoulos, M. C., Abel, P. M., Agranoff, D., Stich, A. *et al.*, A novel and accurate diagnostic test for human African trypanosomiasis. *Lancet* 2004, *363*, 1358–1363.

[6] Agranoff, D., Fernandez-Reyes, D., Papadopoulos, M. C., Rojas, S. A. *et al.*, Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet* 2006, *368*, 1012–1021.

[7] Horgan, G. W., Sample size and replication in 2D gel electrophoresis studies. *J. Proteome Res.* 2007, *6*, 2884–2887.

[8] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R. *et al.*, Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* 2001, *1*, 377–396.

[9] Shaw, J., Rowlinson, R., Nickson, J., Stone, T. *et al.*, Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. *Proteomics* 2003, *3*, 1181–1195.

[10] Alban, A., David, S. O., Bjorkesten, L., Andersson, C. *et al.*, A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 2003, *3*, 36–44.

[11] Görg, A., Weiss, W., Dunn, M. J., Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 2004, *4*, 3665–3685.

[12] Karp, N. A., Kreil, D. P., Lilley, K. S., Determining a significant change in protein expression with DeCyder during a pairwise comparison using two-dimensional difference gel electrophoresis. *Proteomics* 2004, *4*, 1421–1432.

[13] Karp, N. A., Spencer, M., Lindsay, H., O'Dell, K., Lilley, K. S., Impact of replicate types on proteomic expression analysis. *J. Proteome Res.* 2005, *4*, 1867–1871.

[14] Karp, N. A., Lilley, K. S., Maximising sensitivity for detecting changes in protein expression: Experimental design using minimal CyDyes. *Proteomics* 2005, *5*, 3105–3115.

[15] Karp, N. A., McCormick, P. S., Russell, M. R., Lilley, K. S., Experimental and statistical considerations to avoid false conclusions in proteomic studies using differential in-gel electrophoresis. *Mol. Cell. Proteomics* 2007, *6*, 1354–1364.

[16] Anderson, N. L., Anderson, N. G., The human plasma proteome: History, character, and diagnostic prospects. *Mol. Cell. Proteomics* 2002, *1*, 845–867.

[17] Miller, J. C., Zhou, H., Kwekel, J., Cavallo, R. *et al.*, Antibody microarray profiling of human prostate cancer sera: Antibody screening and identification of potential biomarkers. *Proteomics* 2003, *3*, 56–63.

[18] Juan, H. F., Chen, J. H., Hsu, W. T., Huang, S. C. *et al.*, Identification of tumor-associated plasma biomarkers using proteomic techniques: From mouse to human. *Proteomics* 2004, *4*, 2766–2775.

[19] Ahmed, N., Oliva, K. T., Barker, G., Hoffmann, P. *et al.*, Proteomic tracking of serum protein isoforms as screening biomarkers of ovarian cancer. *Proteomics* 2005, *5*, 4625–4636.

[20] Ang, I. L., Poon, T. C., Lai, P. B., Chan, A. T. *et al.*, Study of serum haptoglobin and its glycoforms in the diagnosis of hepatocellular carcinoma: A glycoproteomic approach. *J. Proteome Res.* 2006, *5*, 2691–2700.

[21] Jackson, D., Craven, R. A., Hutson, R. C., Graze, I. *et al.*, Proteomic profiling identifies afamin as a potential biomarker for ovarian cancer. *Clin. Cancer Res.* 2007, *13*, 7370–7379.

[22] Li, X., Gong, Y., Wang, Y., Wu, S. *et al.*, Comparison of alternative analytical techniques for the characterisation of the human serum proteome in HUPO Plasma Proteome Project. *Proteomics* 2005, *5*, 3423–3441.

[23] Gong, Y., Li, X., Yang, B., Ying, W. *et al.*, Different immunoaffinity fractionation strategies to characterize the human plasma proteome. *J. Proteome Res.* 2006, *5*, 1379–1387.

[24] Liu, T., Qian, W. J., Mottaz, H. M., Gritsenko, M. A. *et al.*, Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol. Cell. Proteomics* 2006, *5*, 2167–2174.

[25] Echan, L. A., Tang, H. Y., Ali-Khan, N., Lee, K., Speicher, D. W., Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* 2005, *5*, 3292–3303.

[26] Gundry, R. L., Fu, Q., Jelinek, C. A., Van Eyk, J. E., Cotter, R. J., Investigation of an albumin-enriched fraction of human serum and its albuminome. *Proteomics Clin. Appl.* 2007, *1*, 73–88.

[27] Hunt, S. M., Thomas, M. R., Sebastian, L. T., Pedersen, S. K. *et al.*, Optimal replication and the importance of experi-

mental design for gel-based quantitative proteomics. *J. Proteome Res.* 2005, *4*, 809–819.

[28] Hu, Y., Malone, J. P., Fagan, A. M., Townsend, R. R., Holtzman, D. M., Comparative proteomic analysis of intra- and interindividual variation in human cerebrospinal fluid. *Mol. Cell. Proteomics* 2005, *4*, 2000–2009.

[29] Nedelkov, D., Kiernan, U. A., Niederkofler, E. E., Tubbs, K. A., Nelson, R. W., Investigating diversity in human plasma proteins. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 10852–10857.

[30] Nedelkov, D., Phillips, D. A., Tubbs, K. A., Nelson, R. W., Investigation of human protein variants and their frequency in the general population. *Mol. Cell. Proteomics* 2007, *6*, 1183–1187.

[31] Corzett, T. H., Fodor, I. K., Choi, M. W., Walsworth, V. L. *et al.*, Statistical analysis of the experimental variation in the proteomic characterization of human plasma by two-dimensional difference gel electrophoresis. *J. Proteome Res.* 2006, *5*, 2611–2619.

[32] Terry, D. E., Desiderio, D. M., Between-gel reproducibility of the human cerebrospinal fluid proteome. *Proteomics* 2003, *3*, 1962–1979.

[33] Neuhoff, V., Stamm, R., Eibl, H., Clear background and highly sensitive protein staining with Coomassie blue dyes in polyacrylamide gels: A systematic analysis. *Electrophoresis* 1985, *6*, 427–448.

[34] Blum, H., Beier, H., Gross, H. J., Improved silver staining of plant proteins, RNA, and DNA in polyacrylamide gels. *Electrophoresis* 1987, *8*, 93–99.

[35] Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie brilliant blue G-250 and R-250. *Electrophoresis* 1988, *9*, 255–262.

[36] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2007. http://www.R-project.org.

[37] Warnes, G. R., Liu, P., *Sample Size Estimation for Microarray Experiments. Technical Report 06/06.* Department of Biostatistics and Computational Biology, University of Rochester, Rochester 2006.

[38] Warnes, G. R., Lie, P., Li, F., *Estimate Microarray Sample Size*, SSIZE (Version 1.6.0). 2007 (http://bioconductor.org/packages/1.9/bioc/html/ssize.html).

[39] Adkins, J. N., Monroe, M. E., Auberry, K. J., Shen, Y. *et al.*, A proteomic study of the HUPO Plasma Proteome Project's pilot samples using an accurate mass and time tag strategy. *Proteomics* 2005, *5*, 3454–3466.

[40] Rai, A. J., Gelfand, C. A., Haywood, B. C., Warunek, D. J. *et al.*, HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples. *Proteomics* 2005, *5*, 3262–3277.

