JOURNAL OF PROTEOMICS 95 (2013) 3–21

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

An introduction to statistical process control in research proteomics☆,☆☆

*David Bramwell**

Biosignatures Ltd., Keel House, Newcastle Upon Tyne, UK

ARTICLE INFO

Available online 19 June 2013

Keywords:
Quantitative proteomics
Quality control
Statistical process control
Control chart
2-DE
LC–MS

ABSTRACT

Background: Statistical process control is a well-established and respected method which provides a general purpose, and consistent framework for monitoring and improving the quality of a process. It is routinely used in many industries where the quality of final products is critical and is often required in clinical diagnostic laboratories [1,2]. To date, the methodology has been little utilised in research proteomics. It has been shown to be capable of delivering quantitative QC procedures for qualitative clinical assays[3] making it an ideal methodology to apply to this area of biological research.
Objective: To introduce statistical process control as an objective strategy for quality control and show how it could be used to benefit proteomics researchers and enhance the quality of the results they generate.
Results: We demonstrate that rules which provide basic quality control are easy to derive and implement and could have a major impact on data quality for many studies.
Conclusions: Statistical process control is a powerful tool for investigating and improving proteomics research work-flows. The process of characterising measurement systems and defining control rules forces the exploration of key questions that can lead to significant improvements in performance.

Biological significance
This work asserts that QC is essential to proteomics discovery experiments. Every experimenter must know the current capabilities of their measurement system and have an objective means for tracking and ensuring that performance. Proteomic analysis work-flows are complicated and multi-variate. QC is critical for clinical chemistry measurements and huge strides have been made in ensuring the quality and validity of results in clinical biochemistry labs. This work introduces some of these QC concepts and works to bridge their use from single analyte QC to applications in multi-analyte systems.
This article is part of a Special Issue entitled: Standardization and Quality Control in Proteomics.

Abbreviations: SPC, Statistical process control; QC, Quality control; VSN, Variance Stabilisation Normalisation; TP, True positive; FP, False positive; TN, True negative; FN, False negative.

☆☆ This article is part of a Special Issue entitled: Standardization and Quality Control in Proteomics.

* *Tel.*: +44 191 6453645; fax: +44 191 2302131.
E-mail address: David.Bramwell@biosignatures.com.

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2013.06.010

1. Introduction

1.1. QC is ill defined and scenario dependent

In general, there is no universal, practical definition of quality. As technology improves, what was once the pinnacle of performance can seem imprecise and insufficient in a new context. Even in the restricted scenario of a proteomics research lab there is no single definition as to what is a 'high quality experiment'. It depends on what the researcher is attempting to achieve.

The very first stage of implementing a quality control work-flow is to work out 'What measurable definition of quality will deliver the goals of our research programme?'.

Consider the following two scenarios; choose to measure the abundance of an analyte in ten biological samples to an accuracy of six decimal places at a cost of £100 per sample, or measure the same analyte in one hundred biological samples to three decimal places at a cost of £10 per sample. By design, both experiments effectively 'cost' the same but which is of higher 'quality'? In terms of the accuracy of measurement, it is the first. But what if ten samples do not provide sufficient statistical power to reliably detect the change in analyte abundance that is occurring? The experiment is then of low quality in terms of its potential to deliver a meaningful result at all.

It is possible to partially answer 'What does quality mean for this research programme?' by asking the following more specific questions;

- What is the current technical performance of the systems employed?
- Is this the best performance that can be achieved?
- How do I ensure performance is maintained at these defined levels during measurement of the experimental variables?

The impact of variance within the measurement system on measurements obtained from the experimental samples is one of the key aspects that need to be understood in order to design high-quality experiments and generate meaningful results. It is therefore necessary to characterise and track the performance of the measurement system.

This paper will focus on practical statistical process control (SPC) work-flows for answering these three questions, with a worked example using a freely available data set.

1.2. System characterisation — what is the current technical performance of the systems employed?

Even the most complex system can be practically characterised by treating it as a 'black box'. A known input is introduced and the corresponding output measured. It is assumed that if the system is repeatedly given the same input, any deviation from a constant value noted within the outputs must logically have been introduced by some internal factor within the 'black box' process. If this system is to be used to reliably perform measurements then a requirement would be that, over a reasonable number of repeats of the same input, the majority of outputs are centred on a constant value with predictably distributed errors around it. If these assumptions are correct, it is relatively easy to set rules which determine, for subsequent measurements, if the black box is 'in control' i.e. behaving as we would expect, or 'out of control' i.e. producing an output that we believe to be unlikely given the input and the characterisation of the system previously performed.

The simplest form of system characterisation involves measuring the same thing, the same way, a number of times. This simple definition has some significant implications when considering multiple analyte proteomic measurement systems. For example, it assumes that for an identical input sample the system can be reasonably expected to produce the same answer for every analyte on subsequent runs (subject only to noise variation inherent within the system).

Initially, system characterisation is an exploratory process. How many sample replicates are required is system dependent. This is a very common analysis scenario and there are data visualisation tools and techniques available that assist in this process. Several of these will be used in this paper to explore the properties of a 'real world' data set.

1.3. Process improvement — is this the best performance that can be achieved?

In most cases, system characterisation will lead to the exploration of factors that impact upon data measurement. For example, changing reagent batch may be found to shift the operation point of the system. If down-stream processes do not compensate for this it will have an impact on the overall variance of the system and may also introduce inter-batch bias to measurements. In the initial stages of implementing a QC work-flow it can be highly beneficial to explore such factors and look at mechanisms to mitigate them. The variance in a set of measurements has a direct impact on the number of samples a study requires to have sufficient statistical power to detect significant effects, if they are present.

1.4. Ongoing QC — how do I ensure continuing performance at these defined levels during the measurement of the experimental variables?

Once time and effort has been spent characterising a system, subsequent changes in its performance must be detected. Unnoticed drift can make the difference between a study drawing strong conclusions, not drawing any conclusions at all or even mis-reporting a technical issue as a true biological effect. Plans and procedures should be in place to consistently and objectively manage such issues.

Factors within the process can change at any time and in subtle ways. It is important to detect change quickly so it can be investigated and its impact assessed and mitigated. Initial system characterisation can only report on effects present at that time and repeat characterisation may be required at regular intervals if there is a suggestion that parameters may have changed — this is frequently known as 're-calibration'.

1.5. Statistical process control in manufacturing

Historically, almost all man-made objects were custom pieces made by individual craftsmen of varying skill. As technology

advanced, assembly of complex objects from several components became progressively common. For such strategies to succeed, it is necessary to manufacture components according to certain common criteria, such as size and composition to ensure they fit together and function as desired. Failure to do this produces 'waste' in terms of items which do not meet the grade and time spent assessing them. This led to the concept of 'acceptable limits' or tolerances at the point of manufacture of the components. The advantage of this strategy is two-fold as it allows component manufacture to be optimised to 'sufficient' measurable levels, and it also means that the final complex object need not be completely assembled in order to find out a component is defective.

The most efficient strategy for any complex assembly is to detect and fail bad components as fast as possible. This is easy to understand considering complex and valuable final assemblies; you really don't want to build a whole car before identifying a manufacturing fault in a £5 engine gasket. The cost of failing late i.e. having to strip the engine down to replace the defective gasket, will be substantially more than the cost of the part itself. Similarly, in proteomics, despite the fact that the cost of samples and processing them may be high, the cost of inaccurate or false conclusions being used in efforts to translate results into the clinic or being used as the basis for follow on research can have huge cost implications.

Statistical process control (SPC) is a production and process optimisation strategy, utilising a set of tools and techniques which provide a framework for supporting the requirements outlined in the previous sections. It has been in common use for decades and has repeatedly proven to have a potentially major, impact on the quality and performance of a process. Simply performing statistics on measures of a process and acting on them is not however 'statistical process control'. Fig. 1 shows the distinction quite clearly. Panel 1a shows 5000 randomly generated points (simulated from a normal distribution with a *mean* of 0 and *sd* of 1). Panel 1b shows the

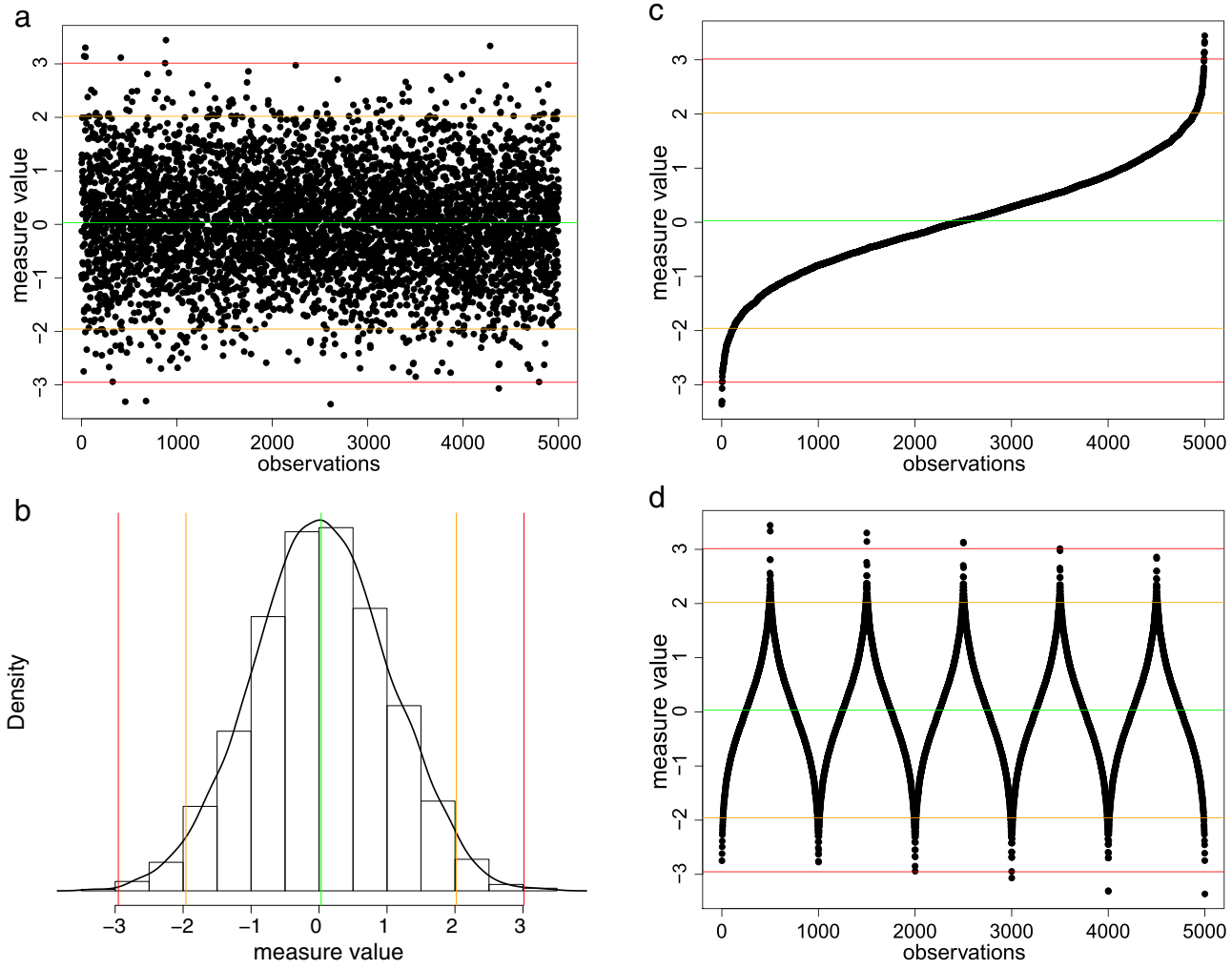

Fig. 1 – A simulation of 5000 measures with 'normally distributed' error variation (generated using *mean* = 0, *sd* = 1). Individual observations are shown as black dots. The green line shows the calculated *mean* of the points shown. The orange lines denote the calculated ±2 *sd* limits. The red lines denote the calculated ±3 *sd* limits. All of the panels refer to the same set of 5000 points. a) The observations in the order that they were generated. b) The histogram and estimated density for the generated points. c) A reordering of the points that has no impact on the *mean*, *sd* or density distribution. d) Another reordering of the points that has no impact on the *mean*, *sd* or density distribution.

density distribution of the points which, as expected, looks to be normal. Panels 1c and 1d show other sets of 5000 points that produce identical distribution functions to the one in panel 1b. Panels 1c and 1d were simply created by re-ordering the points in panel 1a so, by definition, the distributions are mathematically identical. Generally, people look at panel 1a and think the points are ordered randomly, but feel that the data shown in panels 1c and 1d is subject to some form of assignable process as there are clear patterns. As far as standard distributional statistics is concerned, these data sets are identical; nothing in the *mean* and variance calculations would raise any concerns. In pure probability terms the ordering of points shown in panels 1a, 1c and 1d are also identical — the exact sequences of points are equally likely to have occurred at random.

Shewhart [4] realised that, in terms of process control optimisation, distributional statistics were not sufficient in driving improvements. All that matters in improving a process is that there is an assignable cause to the variation observed and whether it can be practically eliminated. He takes a pragmatic standpoint and separates variation into 'chance causes' and 'special causes'. The main difference between the two is that 'special causes' have an assignable cause that can be investigated and potentially improved if the costs and benefits warrant it. 'Chance causes' are simply random variations inherent in the process that are not assignable to a specific cause. Simplistically 'chance causes' are totally random in nature and 'special causes' are expected to have some form of structure over a finite time range.

Considering Fig. 1a once more, it is evident that the distribution of points with time has no visible pattern. We would be hard pressed to find some form of assignable cause, even if it exists. In contrast, Panels 1c and 1d appear to have clear patterns that lead us to believe there may be some form of ongoing process we can identify and isolate, with the intent of mitigating. Only by considering the ordering of the points do we have a way forward in process improvement.

It may be that the point ordering shown in panels 1a, 1c and 1d have equal probability of occurring, but given our experience of the world, panels 1c and 1d have a far higher probability of having an assignable cause. This is a key point in the SPC approach. It is impossible to say if the output of a process is random or not given a finite sample of points — which means that there is no analytical definition of whether a process is subject to only 'chance causes' of variation. A practical solution is to look for 'tell tale' signs that variation is from assignable causes and investigate these as they occur.

If there is no evidence of assignable cause variation then the process may be considered 'in control' and its output can be trusted. A process with evident signs of assignable cause variation may be considered to be 'out of control'. This has major implications when the process under consideration is a measurement system, such as a laboratory test or proteomics multiple analyte assay. If there is assignable cause variation in the measurement, then one of the base assumptions of many statistical tests is actually violated, potentially rendering their conclusions invalid. Considering panel 1c again it is possible to imagine a situation where ill-advised batching of samples (e.g. 2500 cases followed by 2500 controls) could lead to an apparently significant between-group difference where no true difference exists. Shewart, in his decades of consulting and research, reports he never found a process that did not have some form of assignable cause variation when characterisation was performed with the intent of bringing it into 'control' [4].

SPC can therefore be defined simply. SPC involves measuring some aspect of a process and analysing these measurements in a way that allows the elucidation of 'assignable cause' variation and as such assists in its identification and remedy. This statement implies that some form of skilled individual or team is critical in that personnel must be able to interpret, diagnose and act upon what they learn about a process via continual monitoring and improvement. A second implication is that this never ends; a system cannot be characterised once and the assumption then made that all subsequent outputs will be of high quality for ever after.

Fig. 1 has already introduced one of the most powerful tools employed in SPC, the control chart. In its simplest form this is an ordered series of measurements from a process. The panels have statistical limits placed on them (i.e. the lines showing 2 and 3 standard deviations from the *mean*). If the assumption is made that the output shown in panels 1c and 1d is the result of assignable causes, then it is possible to explore 'rules' which can be used to objectively discriminate the type of output in panel 1a from that in panels 1c and 1d. It may be that, although visually the difference is very striking, it is actually non-trivial to derive a metric that works in a generic sense (as this would require the ability to definitively tell random from non-random). A property of random sequences is that individual values cannot be predicted from their predecessors. In most practical situations, distributional properties can be calculated over a 'window' of samples and used to make a prediction about future events. For example, if the *mean* and *sd* of the first 2500 points is calculated, this information can be used to predict the probability of future point values and test these predictions. Such a prediction would be: "Assuming this sample of 2500 points is sufficient to estimate the *mean* and *sd* of the underlying noise processes, the next 2500 points are expected to have a similar *mean* and *sd*." Using the examples in Fig. 1, this prediction holds for panels 1a and 1d but not for panel 1c. Using a window of 2500 samples identifies there is something 'non random' about point ordering in panel 1c. Whilst the 2500 sample window did not pick up the structure in panel 1d, many other window sizes would, e.g. 250 observations per window.

Lag plots are another visualisation tool commonly used in the characterisation phase of SPC to assist in finding cyclical structure in the data (such as that highlighted by the windowing examples). They can be used to systematically explore multiple 'data windows'. Fig. 2a shows an example lag plot. The concept is very simple; ordered points are plotted against each other with a time 'lag' i.e. if the chosen time lag is 1 then *observation 1* is plotted against *observation 2*, observation 2 against *observation 3* and so on. In random data any given observation should not be consistently related to other observations so no clear structure should be visible in the lag plot.

Fig. 2a shows an example of a typical lag plot for random data and corresponds to the data shown in Fig. 1a. Fig. 2b corresponds to the point ordering in Fig. 1d. There is clear visible structure (the lag plot for Fig. 1c is not shown but is essentially identical, except for less 'spread' at the extremes). At this point

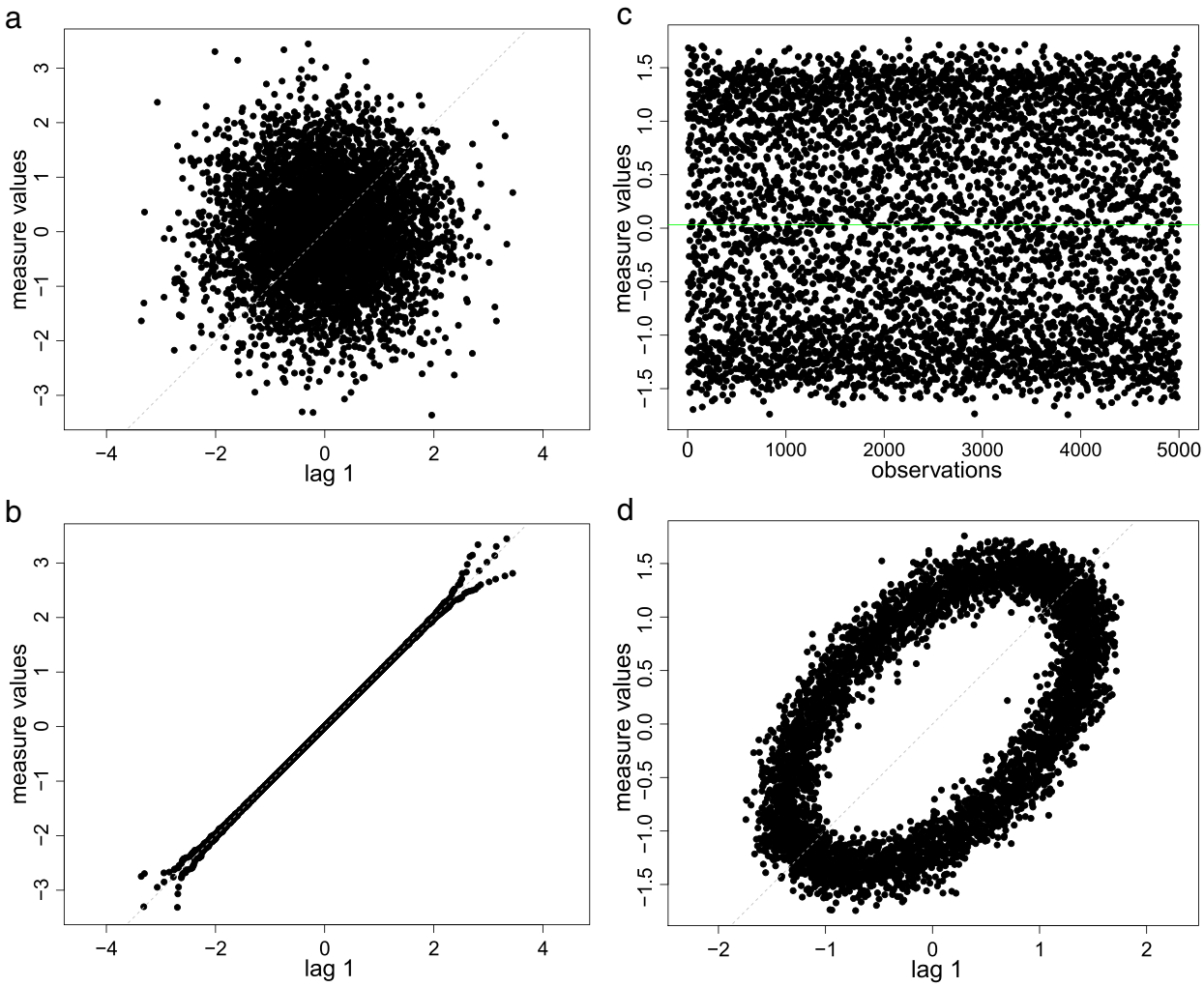

Fig. 2 – Lag plots for exploring structure in ordered observation data. Visible structure in the plots suggests 'non random' components to the observed variation.a) Lag plot for data shown in Fig. 1a. b) Lag plot for data shown in Fig. 1d. c) The observations in Fig. 1a with added sinusoidal offsets. d) Lag plot for data shown in panel c.

the lag plot appears to add little value, as the distinction between the three cases is as obvious in the control chart views as the lag plot views. Fig. 2c and d show a more meaningful example. The distribution of points shown in Fig. 2c was created by adding sinusoidal variation to the points in Fig. 1a. The points do not have the clear structural elements of panels 1c and 1d. Fig. 2d shows the corresponding lag plot for the data shown in Fig. 2c. There is now a clear structure in the lag plot showing an underlying cyclical nature to the points. It is advisable to explore a range of different lags when characterising a system to detect cyclical behaviour with varying cycle times. Results which do not resemble Fig. 2a are cause for concern. In this example, although the points shown in Fig. 2c do not have the same distributional properties as those shown in Fig. 1, they do have the same *mean* and variance.

The 'windowing' of samples introduced in the previous section shows what may be part of a practical solution i.e. it shifts the question from 'is this data subject to only random variation?' to 'have the noise properties changed?' Given the assumption that the window of points is representative, predictions can be made about how likely any given deviation from the *mean* is within a window i.e. assuming a normal distribution, it is expected around 5 in 100 points will be 2 *sd* or more from the *mean*. If variance in the process starts to increase, more points than expected will be beyond these limits. The same will result if the *mean* starts to shift; sensitivity to this can be increased by considering the deviation direction i.e. if a lot more points cross +2 *sd* than expected.

A probability can therefore be assigned to each point subsequently measured and the first stage of an SPC based QC procedure implemented. QC limits can be set according to local circumstances. For instance, it may be that allowing items with 'bad measures' to proceed to later processes is very costly. Given that staff is happy to investigate events that are relatively unlikely, a rule could be to investigate every event which exceeds 2 *sd* from the *mean*. This introduces the concept of a 'run length'; the average of how many measures are investigated as a proportion of those made. In this case 5 in every 100 (i.e. 1 in every 20 measures) are predicted to trigger investigation, giving an average run length of 20. Note that if the distributional assumptions are correct and the process is 'in control' all of the 5 per 100 measures beyond the limits are

expected to be false positives i.e. they are simply expected variation. This is the 'false positive' price paid to enable detection of change in the process if it occurs. This sort of compromise is well accepted. $p < 0.05$ is often used as a threshold for significance of an experimental difference. Simplistically, its meaning translates to 'I am happy that, given certain distributional assumptions, I would only expect to see a difference in means as big or bigger than the one observed (occurring by chance variation alone) once in every 20 repeats of this experiment.'

Experiment design always involves cost and benefit thinking (e.g. increasing the number of samples per experimental group will have an implied cost but it will also improve confidence in the results). A similar thought process is required in SPC where the implications of a rule need to be considered. For example, if investigating a rule failure involves repeating a measurement, and this is cheap and easy to do, then it would make sense to proceed. However, if repeat measurements are very costly (or even impossible) it makes sense to attempt to reduce the false positive rate. A practical approach to achieving this is to consider 'chains of events', i.e. the likelihood of observing consecutive extreme measurements. This approach can produce rules more capable of discriminating between 'non random' behaviour and false positives from chance variation. 'Chains of events' allow probabilities to be combined, producing fewer false positive results whilst retaining operational flexibility. For example, if a single point has a probability of 0.05 of exceeding 2 *sd* then the probability of consecutively observing two such points is $0.05 \times 0.05 = 0.0025$ (one in every 400 pairs of points). The combination is far less likely to occur by chance than each individual event.

The control chart is a powerful tool when used by an experienced operator — humans have a particular natural aptitude for seeing patterns in data. Whilst the experienced practitioner is critical in the QC process, objective rules are required to ensure consistency and transferability between personnel thereby providing the capability to detect problems as rapidly as possible. The control chart display, in combination with a skilled operator, would be sufficient to detect issues in the example observations shown in Fig. 1. However, whilst visually detecting and reporting obvious structure in the data displayed in panels 1c and 1d is trivial, can rules be derived which can do this objectively?

As an illustration, let us assume that 5000 measurements have been made, obtaining the result shown in Fig. 1a. The *mean* and *sd* limits shown are derived from these values. Although this is a synthetic data set, it is clear within 10 observations that both panels 1c and 1d have many more points which exceed the 2 *sd* limits than would be expected from the initial distributional assessment, panel 1a. A 'number of events beyond threshold' rule would pick these up very quickly.

Setting control limits transforms a continuous output metric into a dichotomous system; it is either within the control limits (considered 'in control') or outside them (potentially 'out of control'). This gives four possibilities;

- rule reported out of control and process actually out of control (*true positive* or TP);
- rule reported out of control but process actually in control (*false positive* or FP);
- rule reported in control and process actually in control (*true negative* or TN), and;
- rule reported in control but process was actually out of control (*false negative* or FN).

Rules may also be linked, to improve sensitivity to changes in process state. For example, a measurement falling outside the upper limit of 3 *sd* may trigger an 'alert state' rule which counts how many of the subsequent 10 measurements are outside the 2 *sd* thresholds. This information can be used to assess whether the initial event was a chance occurrence, or an indication of the process entering an out of control state. Generally, widening the thresholds will generate fewer FPs at the cost of increased risk of FNs and conversely narrowing the thresholds reduces FN risk at the cost of more FPs. Improving the process can also reduce the total number of false results which is one of the reasons why continuous process improvement is a major objective of SPC.

There are two possible explanations for an observation falling outside of a defined limit (for repeats of the same input); 1) it is a chance extreme point, or 2) something within the 'black box' has changed. Since all measurements have a level of uncertainty it is usual to consider 'out of control' events as triggering a process or alert state. It may or may not lead to a 'stop the line' (i.e. stop all measurements and investigate) or 'reject observation' outcome depending on the situation. A balance between quality and real-world constraints such as time, cost and impact on downstream processes should be achieved. It is important to consider each case and create an objective set of rules, triggered by fixed criteria. At a minimum, investigation of an 'out of control event' must end with a documented decision about the response made to the event, who decided this and why. QC management requires observation of the number of FPs versus TPs over time for any given rule set, assessment of the process quality benefits realised versus the time spent dealing with false alerts and modification of the rule set according to this information.

The control chart and rule combination is very flexible, intuitive and powerful. Westgard et al. [5] provide far more detail on the rules used in clinical chemistry laboratories and how they can be linked together to achieve quality objectives. Section 1.3.3.15. of NIST/SEMATECH e-Handbook of Statistical Methods [6] reviews system characterisation using these tools whilst chapter 6 of the same on-line resource provides a more in-depth overview of some of the concepts discussed in this text.

Many proteomics discovery research studies are essentially complex data 'manufacturing' pipelines yet it is common to wait until the 'whole car has been assembled' before any of the measurements are questioned. The application of SPC to multiple analyte measurement systems is not entirely obvious, as simple application of one rule per analyte will soon result in every sample failing; if a 'within 2 *sd* of the *mean*' rule was used, on average 1 analysis in every 20 would be expected to fail by chance variation alone. Therefore every sample would have a high probability of failing one of the many QC rules as soon as more than 20 analytes were being measured simultaneously. The rest of this paper and a companion paper [7] show practical ways to implement SPC work-flows whilst avoiding this 'multiple' testing problem.

2. Materials and methods

Many proteomics data sets with 'repeated measures of the same sample' already exist; every data set created using the 2D DIGE strategy [8] has a repeated standard in every gel for example. The DIGE design actually produces three data sets; a set of repeated standards (single stain), a set of samples (single stain) and the DIGE combined and corrected data set itself. DIGE is designed to compensate for per analyte shifts in the measurement platform and as such can mask them. It is important to know about these shifts from the point of view of process characterisation and SPC, even if downstream compensation is in place. The built-in standard provides us with a wealth of information to explore the properties of the measurement system with respect to potential noise factors. This paper uses the 'three data set' strategy to explore practical work-flows enabling statistical process control in research proteomics.

The 2D DIGE data set used is that described in Jackson et al. [9], which is freely available online. The general details of the sample processing protocol are not repeated here but some factors specifically relevant or unique to this paper are described (more processing details are available in the Supplementary data).

Originally, the study was designed to assess population variance amongst healthy subjects. Briefly, 31 male and 30 female participants were sampled to produce the data set. Two blood samples were taken per individual with a 7-day interval between sampling (except for one male individual who only provided a single sample). This produced a set of 2 samples for 60 donors resulting in 120 dual channel gels (totalling 121 with the one visit sample included). There was a known production issue with the first 2D gel batch causing the gels to overrun (i.e. relative to the rest of the gel set some of the spot features have run off the bottom edge of these gels). These six additional gels were omitted from the original analysis performed by [9] but included in the on-line data to allow others to explore QC issues. These six additional sample re-runs bring the total up to 127 gels producing 254 image files (the failing samples were re-run once all the samples had been processed and added to the data set). The gels are a two stain DIGE experiment design where the same pooled internal standard sample was run using Cy3 on every gel along with one of the samples in the Cy5 channel (see Fig. 1 in the Supplementary data for a schematic overview of the sample set).

Detailed meta data was recorded for the whole data set and was also provided on-line. The key meta data factors (using the nomenclature defined in the original publication) we consider in this paper are:

- *2D Gel Batch* — the gel production batch (1 to 14, batch 1 has known issues and batch 14 is a re-run of those samples)
- *IEF Batch for IPG strip* — which IPG strips were run together (1 to 12)
- *Labelling Batch* — which samples were labelled together (1 to 7)
- *Visit* — which visit the sample was from (2 or 3)
- *Position in Tank* — which location the gel was in the tank (1 to 5)
- *Scanner* — two Typhoon 9400 scanners were used across the whole gel set (1 or 2)
- *PMT Voltage for Cy3* — what was the voltage set for the scanner for the Cy3 images (note the scan images were intentionally optimised once per gel batch by pre-scanning one of the images and optimising the PMT voltage to get serotransferrin on the gel to be in a set range)
- *PMT Voltage for Cy5* — what was the voltage set for the scanner for the Cy5 images (optimised in same manner as Cy3 images)

The analysis in this paper uses the same image alignment as the original paper but has different spot feature data. The aligned images were loaded into Progenesis SameSpots version 4.5 [10] and features were detected automatically. This produced 1004 spot features. The feature outlines were applied to the images and 'raw volumes' calculated (by simply summing together the calibrated pixel values within the spot feature boundaries). No background correction, normalisation or further processing is included in these values.

The style of work-flow outlined supports the post analysis of an already completed data set and also an ongoing study where any new samples are aligned to the chosen reference space and the pattern is simply applied and metrics derived. This in no way restricts the downstream analysis where different feature detection could be used.

All subsequent analysis presented in this paper was performed in R [11] (for base functionality and plotting) with custom scripting and supporting packages rms [12] (for *varclus*), VSN [13,14] (for *vsn2*), and car [15] (for *dataEllipse*).

This paper explores these raw values under four distinct normalisation/data transformation strategies:

1. *log_2 raw volumes* — the raw values under a log base 2 transformation
2. *Variance Stabilisation Normalisation (VSN) of the standards* — VSN transform of the standards based on raw volumes using a reference group including standards in 2D gel batches 2–6. These samples are effectively mapped to a log_2 space.
3. *Variance Stabilisation Normalisation (VSN) of the samples* — VSN transform of the samples based on raw volumes using a reference group including standards in 2D gel batches 2–6. These samples are effectively mapped to a log_2 space.
4. *Difference of the VSN samples with respect to the VSN values of their standard channels* (i.e. *per gel VSN(Cy5)–VSN(Cy3)*) — values corrected to their standards based on the previous two VSN calculation criteria.

The normalisation strategy employed is the *vsn2* command from the VSN R package [13]. This version has the capability of allowing a reference set to be specified that defines the reference normalisation space, this feature makes it easy for us to add new observations as they become available.

For the standards set, the normalisation reference was based on '2D Gel Batch' 2 through to 6. The same subset was used for the samples but based purely on sample related measures i.e. standards are referenced to standards and samples are referenced to samples. The VSN algorithm produces normalised values in a space analogous to log_2 transform of the raw data and

so should be approximately comparable in scale terms to our log_2 raw data sets.

The final data set achieves 'standardisation of samples' simply by subtracting the aforementioned standard data from the samples data on a gel by gel basis. No attempt was made to shift the *mean* of the result so these samples will not necessarily have an effective *mean* ratio of 1. The 'difference VSN' samples are equivalent to a ratio space representation of feature expression.

3. Results

3.1. Control charts for proteomics data

As discussed above, the first task to be carried out is system characterisation. log_2 transformed raw feature volume data from all of the standards was used to highlight additional considerations and visualisations before systematically considering multiple transforms and different data types.

Consideration of the meta data in Section 2 indicates many factors which may be expected to affect the bulk of the features simultaneously (e.g. the PMT voltage setting on the scanner). This suggests that an obvious first metric for exploratory system characterisation is some form of global measure. Fig. 6 from [9] shows the design plot based on a subset of the meta data factors and how they group a global measure (the median values of all features on a gel).

In general, there is no requirement to use a global metric. A subset of features could be used and can provide more sensitivity to specific issues [7]. There is also no limit on the number of metrics or rules used except that the portfolio of rules should be considered as a whole in cost benefit terms.

The metric presented in all of the graphs and analysis in this section is the median value of all of the 1004 features from a gel image under the specified transformation or normalisation scheme. All the results have used batches 2–6 as a reference set for calculation of the *mean* and control limits and for the purposes of normalisation when applicable.

The various control chart versions in this section add more and more information to the plots to assist in the exploration of 'assignable cause'. In general this involves layering information from meta data factors onto the chart.

Fig. 3 shows a control chart based on taking the chosen metric for each of the standard gels. This chart is slightly different to those presented in the Introduction. The observations are in '2D Gel Batch' order and the batch information is now shown on the plot. Visually, there is already a suggestion that assignable cause variation is present, with some correlation to '2D Gel Batch'.

Fig. 4 shows the distributional information for the data shown in Fig. 3. Visually there may be some suggestion of non-normality but the distribution would not really suggest an issue with particular batches.

Given that visually several batches appear to be subject to some form of assignable cause, we can explore defining limits that would objectively report this. In the Introduction, the creation of statistical limits and the possibilities of windowing data was discussed, but details of how the *mean* and limits should be calculated was not. Shewart [16] formulated the initial work in this area. The Shewart version of the control chart usually calculates limits from batched observations and utilises running limits. Subsequently, Levey and Jennings [17] developed a version that is more convenient for routine use in clinical chemistry laboratories. Their approach utilises a number of reference samples from which the limits are derived. They suggest that the reference set should remain valid for a certain time period or reagent batch after which the system should be re-assessed and potentially re-calibrated. There is an obvious trade off between generating stable and meaningful limits and the number of standards you need to produce them.

There are no clear specifications on how many samples should be used to derive the population metrics. Fig. 5 shows how the estimate of the *mean* and *sd* cumulatively develops across the data set as increasing numbers of observations are included. Despite the variation present across the whole graph, the estimates are relatively stable after 20–40 observations even with the inclusion of the scanner setting induced shifts in batches 8, 10 and 12. Over the full set it appears that using smaller numbers of observations to assess the *mean* and control limits has a tendency to underestimate the variance, resulting in narrower limits and potentially more false positive 'out of control' events. Consider Fig. 6, this uses the standard gels from 2D gel batches 2 through 6, which results in tighter limits and the 'rejection' of three observations which were not rejected when the limits were set using all observations. Also note in Fig. 6 that stricter 2 *sd* limits (which would flag up all points in the yellow or red areas) would have at least one 'rule fail' for all of the batches that look to have shifts visually.

The first question that occurs once the 'black box' is reported as being 'out of control' is; 'Is there an assignable cause?'. This leads to consideration of batching and meta data. As previously mentioned, the [9] study recorded many meta data factors whilst processing experimental samples. These range across reagent batch, the scanner used and its settings, and which gels were run in the same tank and their position in the tank. Generally experimenters have to consider the factors that they believe could systematically influence their results and at least record the details. Batch meta information can be layered on top of the control chart to allow for a visual inspection of potential factors affecting process performance. Fig. 7 shows such a display.

The original experiment was not designed to allow full separation of the meta factors and as such they are not all independent. This can be explored more formally using variable clustering. Fig. 8 shows the output from the *varclus* command from the rms package [12] in R [11]. In this analysis the meta factors are themselves clustered together (irrespective of the measured feature data in the gels). This shows which meta data factors are related. In this gel set, 'Labelling batch', '2D Gel Batch' and 'IEF batch for IPG strips' are closely related so it is unlikely that one of these factors can be isolated as a lone cause of variation.

The final tool used is the lag plot which allows us to assess cyclical behaviour within sample windows, which was also discussed in the Introduction. Fig. 9 shows a version where additional information is provided. The observations have been coloured by '2D Gel Batch' (and also numbered) which

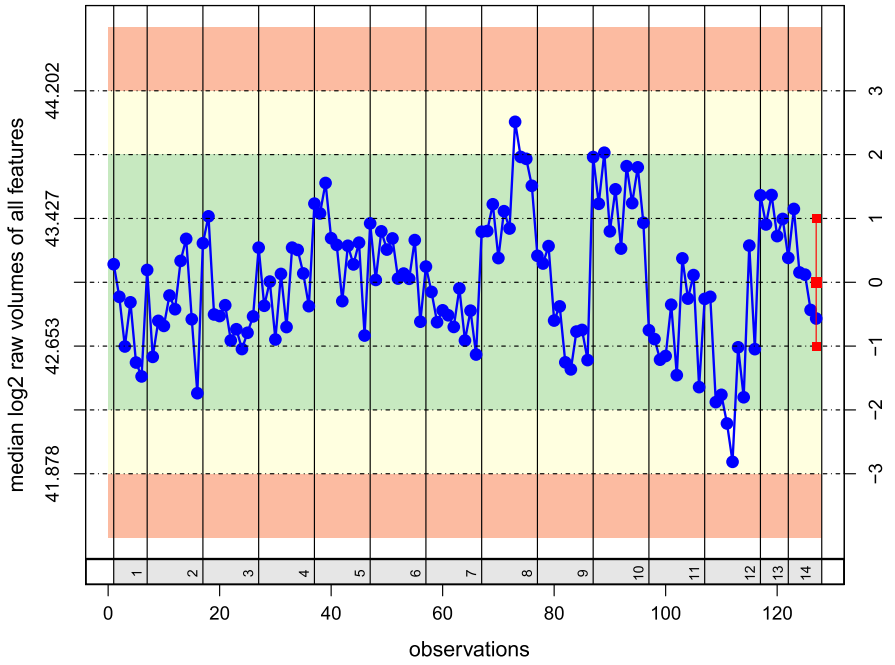

Fig. 3 – An example control chart for the median of standard feature values per gel. The blue points and lines show sequential observations of the chosen 'metric'. The left hand y-axis shows the values in the original metric space and the right hand axis shows the metric points mapped to a standardised space based on an estimate of the mean and standard deviation of the chosen reference set (all observations in this plot). Observations in the reference set are coloured grey in the bar at the bottom of the plot. The green background rectangle highlights the area that points within 2 *sd* of the mean reside. The yellow rectangles show where points between 2 and 3 *sd* of the *mean* reside and the red areas show where points between 3 and 4 *sd* from the *mean* would reside. The vertical lines denote boundaries between subsequent batches, in this case, the 2D gel batch. The number of the batch is given in the corresponding rectangles at the bottom of the plot. The red lines and squares at the right of the plot show the mean and ±1 *sd* points of all of the observations regardless of what the reference set is.

gives a visual indication as to whether outlier behaviour is batch related. A simple set of data bounds have been added which depict where most of the data is expected to reside, given estimates based on the data itself and normality assumptions. The inner ellipse is expected to contain 50% of the data and the outer ellipse 95% (using the *dataEllipse* function from the car [15] package). For a normally distributed random process, a uniform cloud of points is expected to develop. In Fig. 9 we can see that this appears to be the case for most of the batches, but batches 8, 10 and 12 deviate from this distribution.

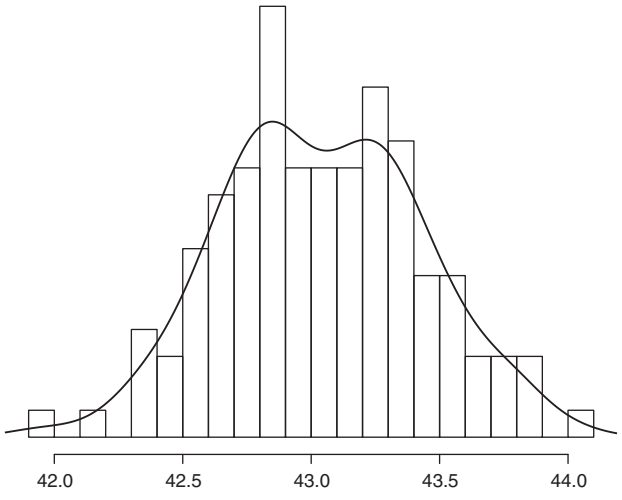

Fig. 4 – Distributional overview for median of all features applied to image log_2 raw volumes.

The visualisations and metrics discussed do not report on the absolute performance of the system. Being 'in control' does not mean a system has high performance; it simply means that it is operating in the manner we would expect it to given some window of past performance. It is obvious that a 'noisy system' will have wider absolute limits than a 'low noise' system. Our focus in this paper is on initially characterising the state of the system, providing a route to improvement and creating rules that objectively warn when system performance may be unexpectedly changing. The practitioners may well wish to benchmark their system against other such systems as part of the characterisation process but this is beyond the scope of this paper.

The tools and visualisations described can now be used to explore properties of the data set and give an indication of the sorts of QC issues that SPC could objectively report on.

3.2. *Standards log_2 transformed raw volume*

Fig. 7 shows the data for all of the log_2 raw volume results from the standards. As noted earlier the chart is unremarkable except for apparent excursions in batches 8, 10 and 12, most likely linked to scanner settings. It is interesting that batch 1

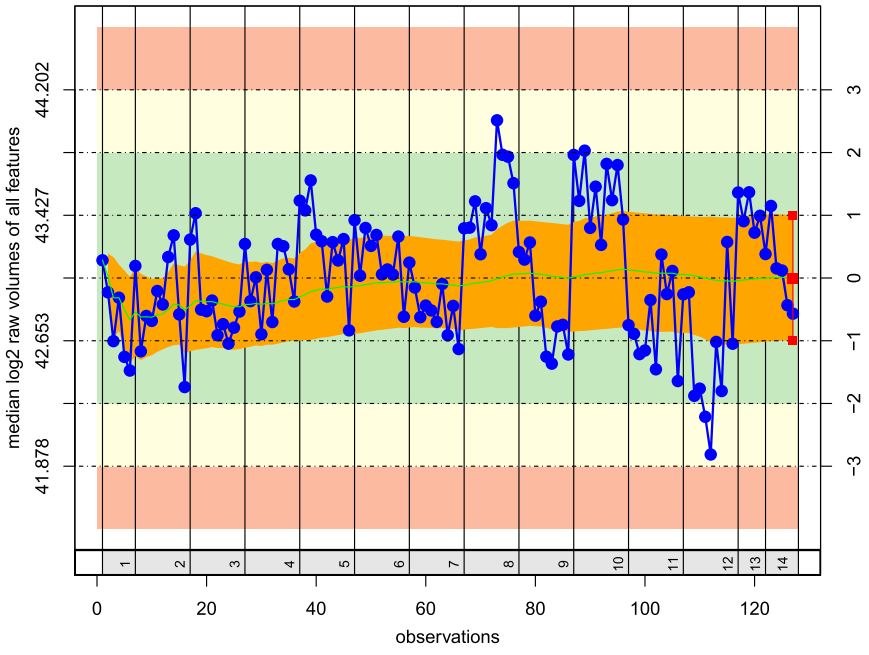

Fig. 5 – The green line shows how the estimate of the *mean* alters as more observations are included. It shows a cumulative estimate i.e. the line value at observation 20 includes all observations up to 20. The orange polygon shows the same estimate for how the *sd* develops as more points are used in its calculation.

(with the known running error that caused around one fifth of the gel features to be missing) looks very well behaved under this scheme and metric. This is not entirely unreasonable given the way the gel scans were optimised for scanning and the use of the median of a large number of features.

Fig. 9 shows the lag plot for the same data. As mentioned in Section 3.1 the points appear to be clustered reasonably tightly with notable excursions for batches 8, 10 and 12.

Fig. 10 shows a design plot similar to the one presented in the Jackson et al. [9] paper. The design plot is another

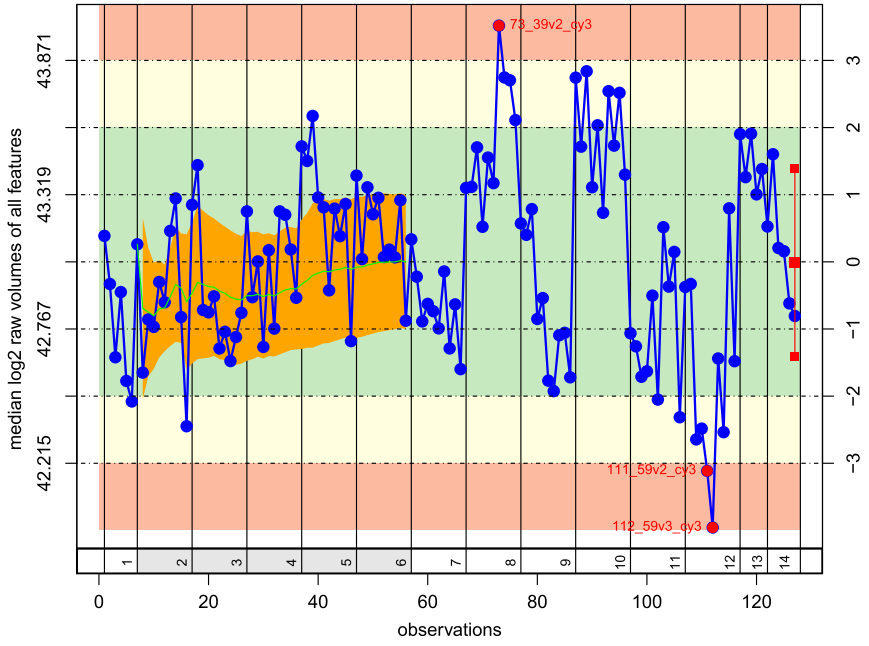

Fig. 6 – This plot shows using a reduced reference set. 2D gel batches 2 through to 6 are used to calculate the estimates for *mean* and *sd* that are used in the control limits. Observations that are coloured red instead of blue denote points outside the defined control limits (*mean* ± 3 *sd* in this example). The observations used in the reference set can be seen coloured grey in the rectangles at the bottom of the plot.

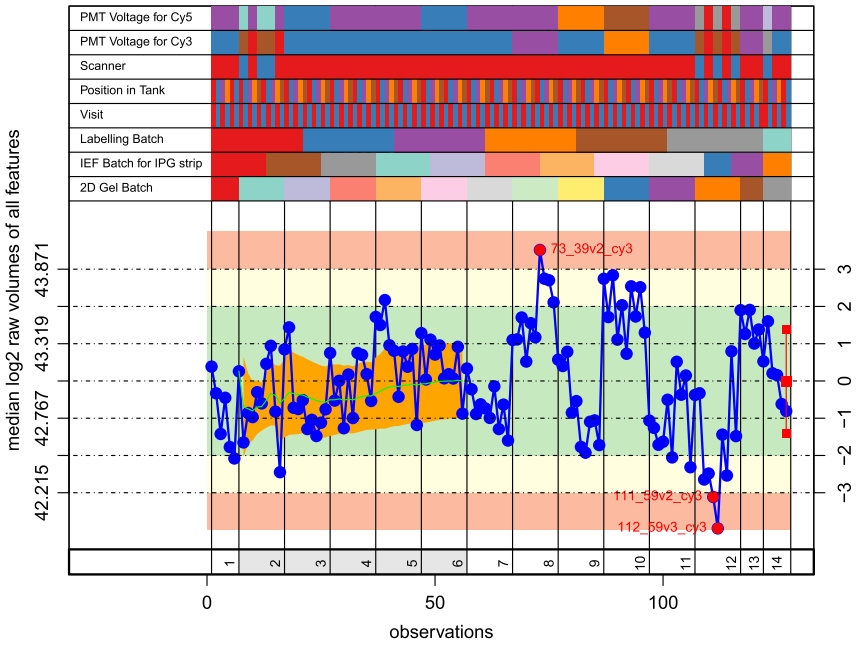

Fig. 7 – Control chart with 'meta ribbons' added. The meta ribbons show how various meta batches correspond with observation values. Each strip is independent, so the colours used in '2D Gel Batch' ribbon are not related specifically to the colours used in say 'IEF Batch for IPG strip' ribbon. A different colour is used for each unique batch factor.

alternative visualisation of the data shown in the control chart. Design plots were discussed in detail by Jackson et al. [9]. Briefly, instead of focussing mainly on a single meta data batch factor, the design plot lists a number of factors and shows how the median of the observations included under different batch schemes would plot on the control chart. The chart has been extended to show the same statistical limits as the control charts. We can see from this data presentation that the most outlying 2D gel batches are 12, 10, 8 and 13 which supports the conclusions of the previous two figures. There is a suggestion that IEF batch for IPG strip 10 is an outlier but as we noted under the variable clustering analysis (Fig. 8) that the meta data does not offer us the capability to really separate out the factors with any confidence. The 'PMT Voltage for Cy3' data seems to suggest that the PMT voltage setting has quite an impact on the metric being used — as we may expect given this is raw scan data.

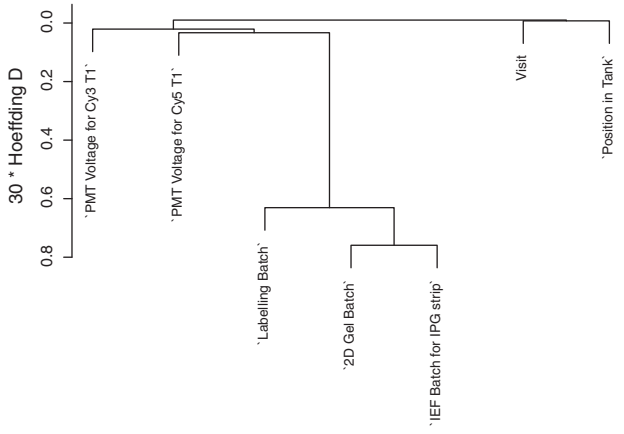

Fig. 8 – Variable clustering for 'all standards' meta data.

3.3. *Standards under Variance Stabilisation Normalisation*

The original study utilised a scanning rule that sought to standardise batch scanning by optimising the value of a few features on the gel (serotransferrin). Normalisation schemes, such as VSN, can utilise information from many features and through this can provide more reliable estimates of the behaviour of the whole population that are more robust to outliers. They also add assumptions about the underlying distributions of the data. It can be useful to explore the impact of these assumptions by adding analysis steps in stages. The previous section showed the raw data, now we explore how further processing and assumptions can change our view of the data. Fig. 11 shows the control chart for the standards under the VSN scheme. It is clear that the normalisation scheme has successfully compensated for the differences in scanner setting with batches 8, 10 and 12 now appearing to fit in with the rest of the batches. Batch 1 (with gel over-run issues) is now clearly showing as an outlier.

Fig. 12 tells a similar story with a tighter cluster of batches showing batch 1 as a clear outlier. There is also some suggestion that batch 14 may have an undiagnosed issue.

Fig. 13 clearly shows how the VSN transformation has reduced the impact of the scanner settings with the ANOVA for PMT Cy3 now reporting no significant difference between the settings based on using the batch medians as a metric. Again, batch 1 is a clear outlier under 2D gel batch grouping and there is a suggestion of problems with batch 14. As meta factors for the IEF and labelling batches are known to correlate with the 2D gel batch they are not considered any further here.

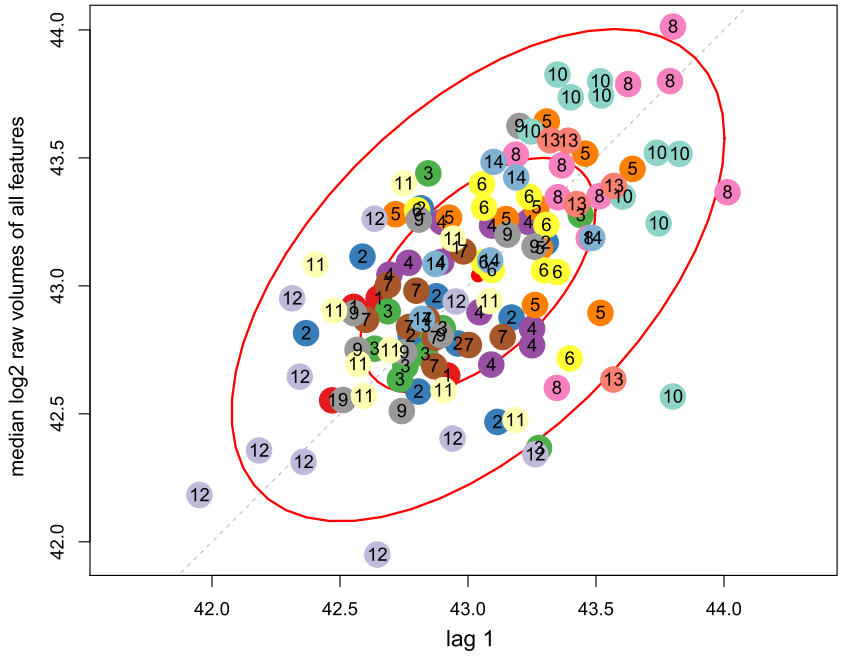

Fig. 9 – Each observation is plotted with respect to its immediate predecessor. In a random process we would expect this to produce a reasonably tightly clustered ellipse of points. Systematic trends developing will tend to move the centre of the clusters of points along the $y = x$ line. The points have been coloured to match the 2D gel batch colours used in the meta data ribbon on the control charts and the numbers are the batch numbers. The red ellipses show data bounds that denote areas in which 50% and 95% of the data points would be expected to lie assuming a 2D normal distribution. We can see that the points from batch 12, when taken as a group, are shifted down from the main group suggesting a systematic shift of operating point for that group.

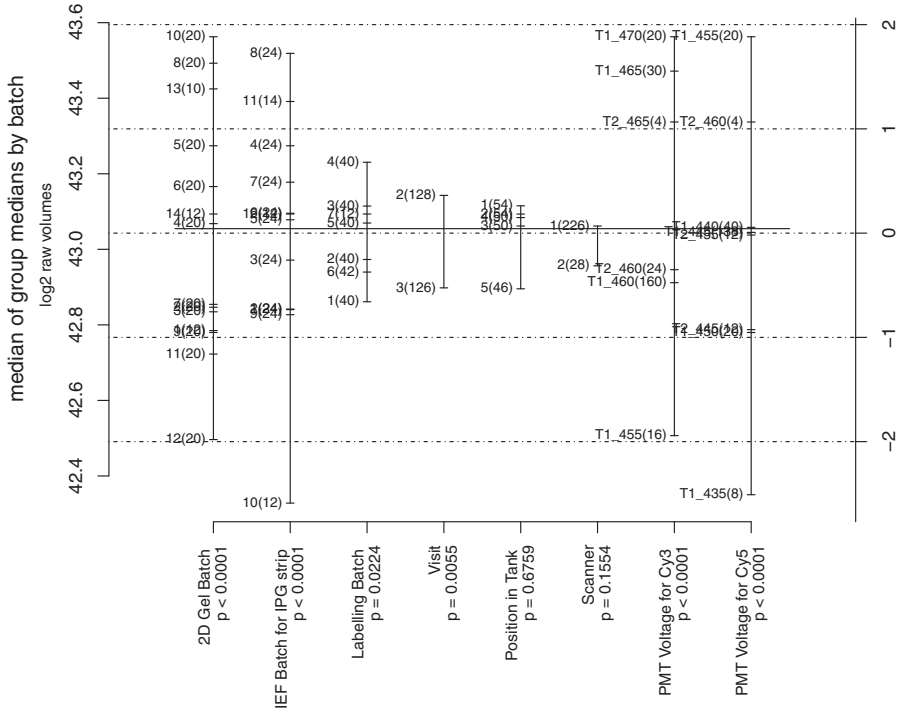

Fig. 10 – Median of group medians under various meta data batching schemes. If we take the leftmost line first, this corresponds to the data that we have plotted in the previous control charts. Each point on the vertical line shows the median of the 'metric' points within that batch. The numbers in brackets are the number of the observations for the given batch. Each of the subsequent vertical lines shows the results for batching the observations by alternative meta data criteria. The *p* values supplied with the vertical line labels are the results of an ANOVA of the metric observation values grouped by the corresponding meta data and they are suggestive of factors that may be significant in producing shifts in the data.

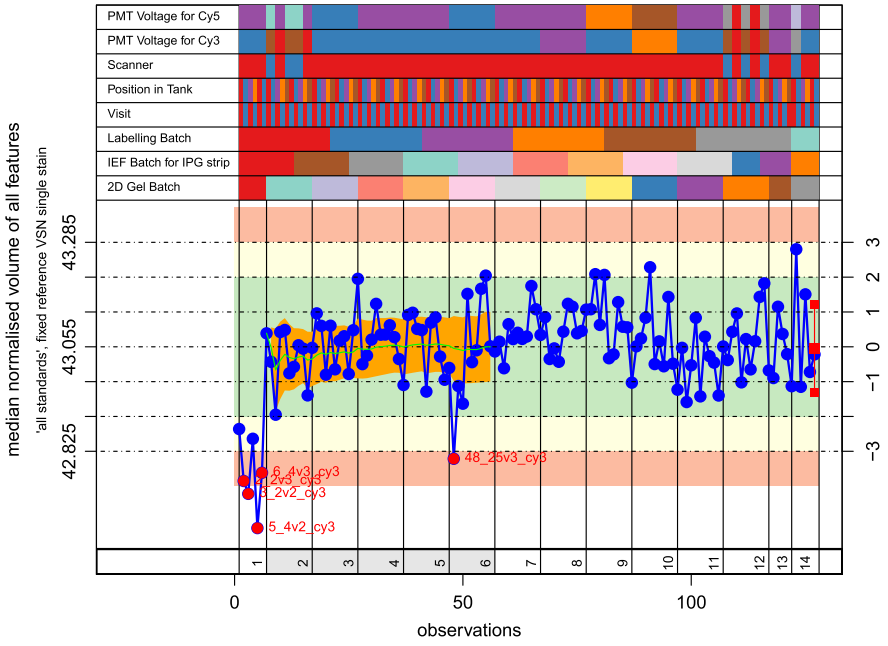

Fig. 11 – Control chart for median of all features applied to image set 'all standards', fixed reference VSN single stain.

The raw and the normalised views of the data are relatively simple to produce and we would recommend the use of multiple views and data processing schemes in parallel when exploring the system characteristics.

3.4. *Samples under Variance Stabilisation Normalisation*

Fig. 14 shows the control chart for the samples under the VSN scheme. The results are essentially the same as for the standards with batch 1 clearly an outlier. This is a very encouraging result as it suggests that, for proteomics data, the statistical process control technique not only works for repeats of the same sample but can also function when biological variation is included not only in the data itself but also in the reference set. This may not be the case for all data sets so it is recommended that this finding should be confirmed for any given experiment design.

It is not entirely surprising that a global metric of samples behaves in a similar manner as a global metric of standards. Proteomics has been using the assumption that 'not all

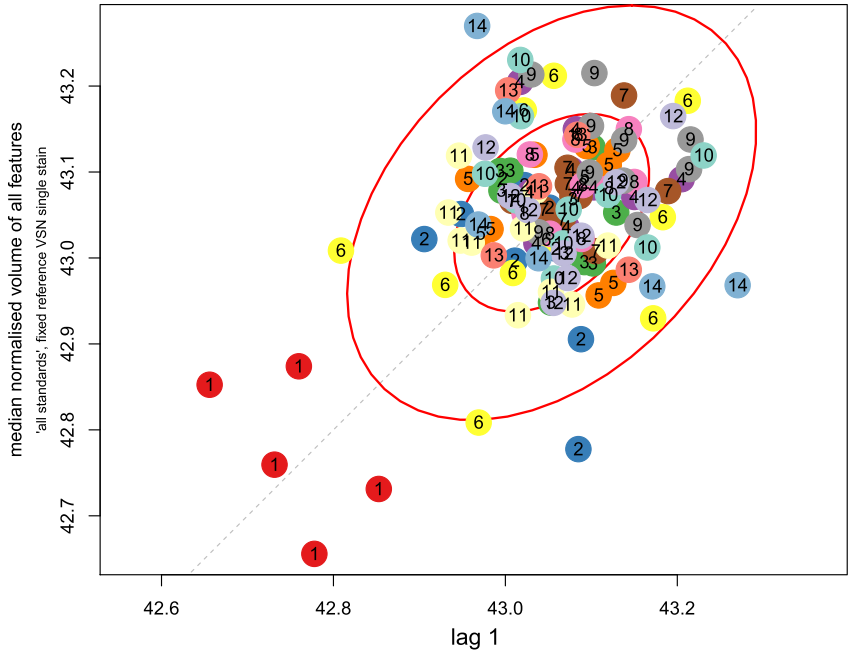

Fig. 12 – Lag plot for median of all features applied to image set 'all standards', fixed reference VSN single stain.

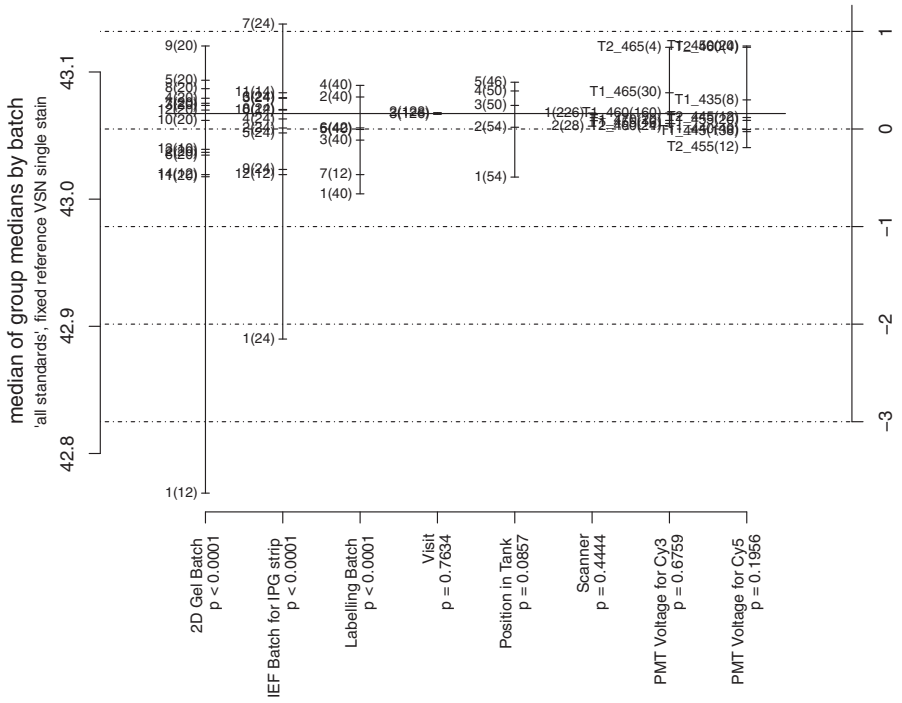

Fig. 13 – Design plot for median of all features applied to image set 'all standards', fixed reference VSN single stain.

analytes are affected under the experimental conditions' for some time; it is assumed by almost all data normalisation schemes. The presentations in this section suggest that the same assumption can be used to consider some sample features as 'built in standards' for QC purposes. The fact that the DIGE gel pair is highly correlated could have contributed to the similarity of the result but both analyses were conducted independently. At least, this demonstrates that the biological variation within the sample set does not prevent technical issue rules being derived.

Fig. 15 shows the corresponding lag plot. It, too, is similar to the standards version (Fig. 9) but with less suggestion of an issue with batch 14. The design plot for the samples (Fig. 16) is also very similar to the standards version and again fails to show an issue for batch 14.

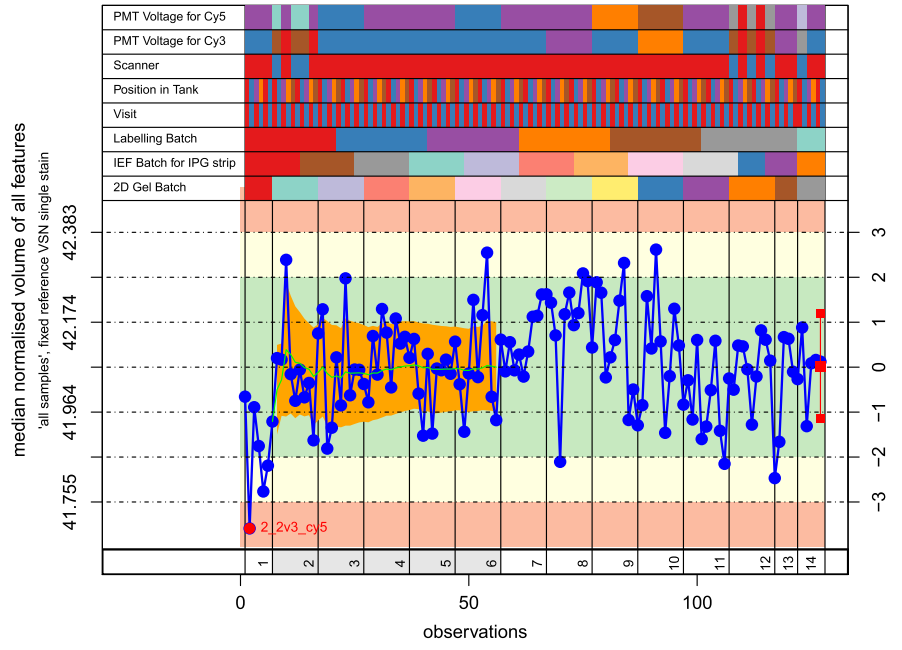

Fig. 14 – Control chart for median of all features applied to image set 'all samples', fixed reference VSN single stain.

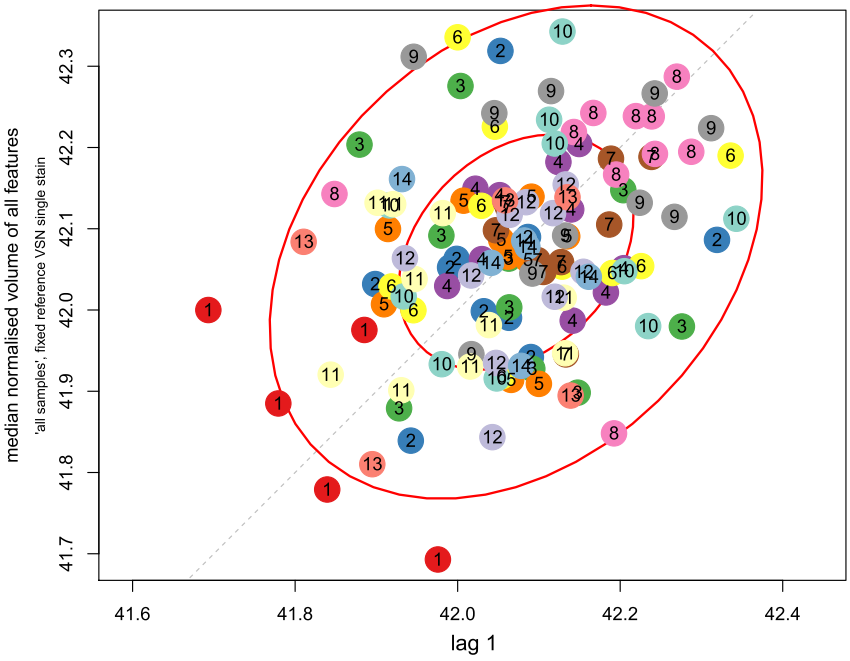

Fig. 15 – Lag plot for median of all features applied to image set 'all samples', fixed reference VSN single stain.

3.5. *Samples under Variance Stabilisation Normalisation difference to standards*

Fig. 17 shows the control chart for the VSN difference case. Batch 1 is not reported as having any issues. This is expected as the over-run affects both stains within the gel equally and hence is masked. In this case, the batch quality issue was easily identified by the experimenters but in other cases the images are altered in a more subtle way which may still matter, for example a technical issue with albumin overloads that causes material to be deposited as streaks in an area of the gel containing other features. In this case the problem may not be detected so easily because the issue equally affects both sample and standard images and can mask the real expression values without anything unusual appearing in the differential data. Jackson and Bramwell [7] explore these issues in more detail.

It is important that an experimenter is informed when such masking may be occurring. The QC procedure can be adapted to utilise not only the whole image but subsets of features that report on the QC of different sections of the data. It may be that the investigation of an outlier highlights such a localised issue and a rule is created to report specifically on this in the future (see [7] for specific examples).

There are indications of an issue with batch 14 with the meta ribbon suggesting that there may be a problem with the 2D gel batch, IEF batch or labelling batch. Batch 14 was the re-run batch of batch 1. This is a post analysis so there is little more that can be done other than be cautious about results from that batch. Additional analyses to localise the origin of a noted difference are also possible by comparing differential results versus other batches. If this analysis had been performed as part of the original experiment it would have been possible to explore the issue more fully and potentially re-run the samples. Batch 14 did not stand out as an outlier batch in the original Jackson et al. analysis. This may be a result of the differing feature sets employed as 4534 features were used in the original paper compared to 1004 in this analysis.

Figs. 18 and 19 show the lag and design plots for the VSN difference data. These support the conclusions drawn from the control chart although it is interesting to note that in the design plot the scanner setting value spread seems to have a larger relative impact than was observed in the VSN single stain experiments. It may be that future experiments should explore whether the scanner setting optimisation strategy needs review. Simply locking the scanner settings to a pre-set value may also cause issues, as this leads to some gels losing dynamic range whilst others saturate out certain feature sets. The control chart approach gives a metric to explore this, but the impact of such procedural changes must be considered for the entire process 'end-to-end'. It is important to take a pragmatic approach and aim to identify the metrics which report on key factors, i.e. those directly impacting upon the desired quality of the experiment. Becoming unduly concerned with optimising all aspects of a process can actually be counter-productive, as this consumes valuable time and resources yet can potentially yield negligible benefits.

4. Discussion

Implementation of even a simple statistical process control work-flow can provide useful and timely information on potential technical issues occurring within complex proteomic analysis processes. Control charts are easy to create and interpret, and allow for objective rules and procedures to be implemented to detect and deal with issues that may impact upon quality. A crucial fact is that these procedures are

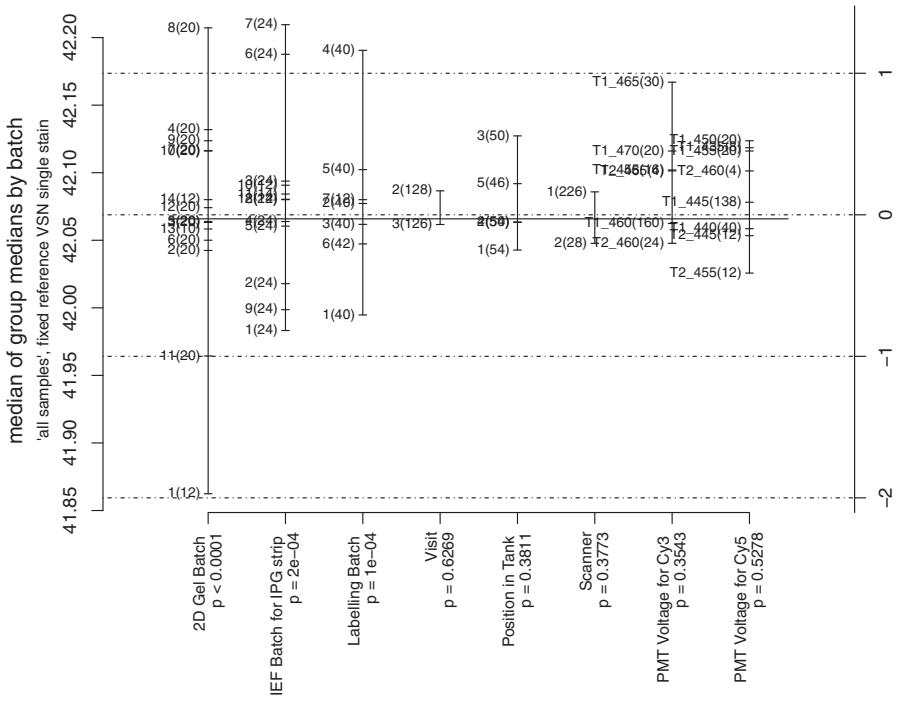

Fig. 16 – Design plot for median of all features applied to image set 'all samples', fixed reference VSN single stain.

objective — if an observation is outside of pre-determined limits a procedure is triggered, which will lead to one of a defined set of outcomes.

Once process metrics have been defined the resourcing overhead of running the procedures is generally much lower than one may initially assume and can easily be scripted using freely available software tools to run, in many cases, automatically.

As discussed above, process characterisation is key to successfully implementing SPC. For this reason, we recommend extensive routine meta data collection during analyses, at least until the main factors impacting upon the performance of a process are regarded as being well-understood. We also recommend that samples are analysed blind with regard to their experimental grouping and that all statistical QC processes are completed before this blind is broken and

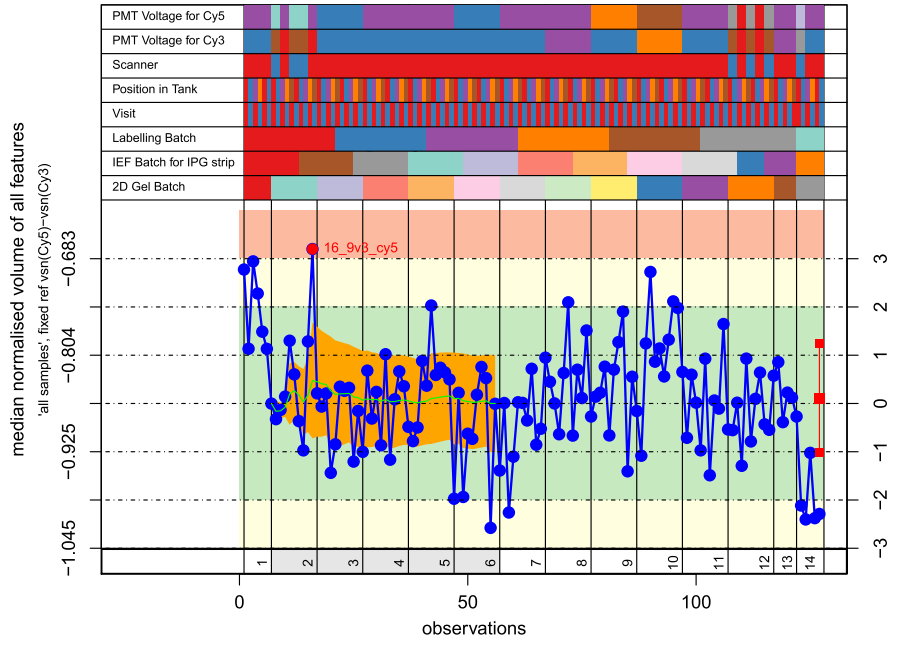

Fig. 17 – Control chart for median of all features applied to image set 'all samples', fixed ref VSN(Cy5)–VSN(Cy3).

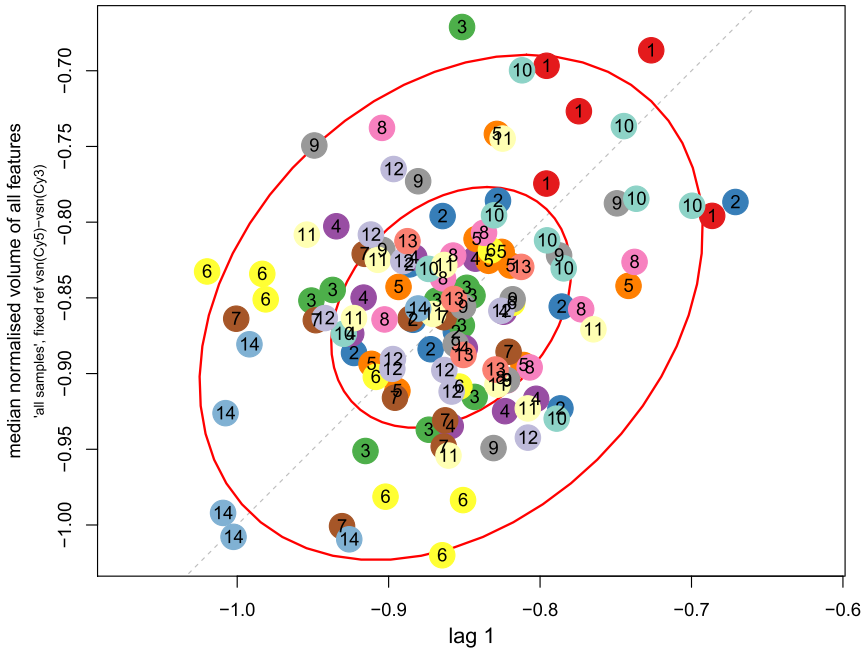

Fig. 18 – Lag plot for median of all features applied to image set 'all samples', fixed ref VSN(Cy5)–VSN(Cy3).

data analysis intended to address experimental hypotheses commences. This ensures that QC decisions on inclusion and rejection of data from the final analysis can be made objectively and consistently across the entire experimental data set, without being influenced by whether data corresponds to 'control' or 'test' samples.

Provided that QC metrics are stable, it does not matter if the final data analysis strategy is different to the QC strategy, so long as the process implemented suits a particular researcher's requirements. It should not matter that the QC procedure does not use all the data used in the final analysis or that it does not use the final intended quantitation strategy,

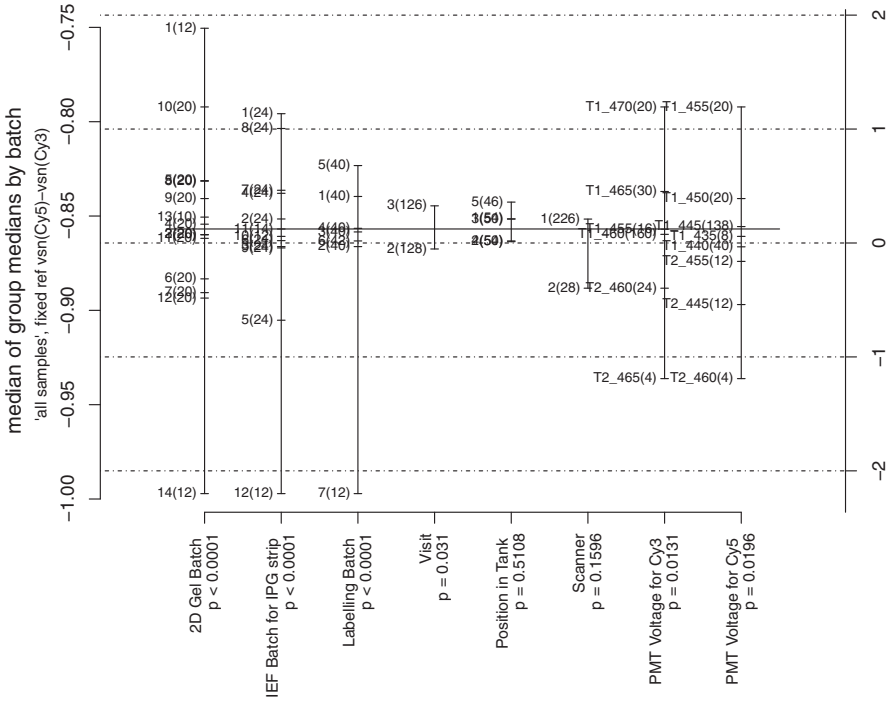

Fig. 19 – Design plot for median of all features applied to image set 'all samples', fixed ref VSN(Cy5)–VSN(Cy3).

although it may be beneficial to implement QC at multiple stages in the processing and analysis pathway as we have shown.

This article was only intended to serve as an introduction to statistical process control and demonstrate the potential of the technique. The analyses discussed outlined procedures capable of; objectively rejecting a known 'bad' gel batch without visual inspection of the actual gels themselves, identifying scanner setting changes and highlighting a potential issue for further investigation in the DIGE data. As every DIGE experiment generates three data sets that can be analysed in this manner, a large amount of data is already in existence which can be used to explore rules and assess performance in many research areas.

More complex and advanced QC schemes could be defined, which would still map back to the simplicity of the control chart methodology but operate not only on single derived global measures, but also on all or just particularly relevant subsets of data. This strategy has the potential for better assessment of locally occurring effects [7].

One of the key concepts of SPC is 'the process of continuous improvement'. In practice, this means making pragmatic choices and operating with a good compromise between overall quality assessment and resource overheads, whilst maintaining vigilance for unexpected events. If a serious quality issue occurs it should be investigated, especially if existing procedures failed to identify it. This then gives an opportunity for procedures to be improved upon, with the result over time being an overall improvement in process quality and hence also in the quality of the results it generates.

This article has shown that the SPC approach provides useful and timely information for 2D gel electrophoresis data. *Is it really applicable and necessary for the more automated and cutting edge MS based work-flows used by most proteomics researchers today?* History seems to suggest that the answer is clearly that attention to quality and reproducibility is crucial for translation from research to practical applications. The research proteomics community would consider clinical chemistry lab analyte measurement as a 'gold standard' to aspire to, but the level of performance routinely available today is the result of many years of effort. Tonks [18], in his 1963 paper, reports, when 170 Canadian clinical chemistry labs analysed pre-prepared samples for, now routine, analytes such as total protein, total cholesterol and glucose, that 'Over 40% of the 3762 values reported fell outside of the allowable limits of errors'. SPC has played an important role in achieving the improvements leading to the systems and processes that clinicians now rely on. Today's fully automated, custom designed, clinical chemistry systems are still frequently used with SPC oversight and are never assumed to just consistently provide accurate results 'because this is the latest and best hardware available'.

In terms of ease of implementation, recent 'IDfree' MS approaches, such as those detailed in the QuaMeter suite of tools [19], are ideal candidates for tracking via control charts and are amenable to essentially automated application and reporting. The 'IDfree' approaches also allow the 'data manufacturing' pipeline to be split down, removing the additional assumptions and complexity of the downstream analysis. Less of the 'car' needs to be built before you find the defective 'gasket'.

The recent ABRF PRG 2012: Quality Control LC–MS/MS study [20] discusses the need for improved QC oversight in LC–MS/MS analysis and highlights how broadly applicable and robust the QuaMeter metrics are across a range of proteomics analysis work-flows. The early reports [21] show metric data plotted in a sequential sample format analogous to the control chart and many exhibit clear structure. The 2013 presentation from that study [22] demonstrates some potentially surprising outcomes. Results were highly variable month-to-month with respect to both the number of peptides and even the parent protein species detected. This being despite contributors analysing the same, relatively simple sample mixture, on the same hardware platforms, using the same standardised work-flow and under the full knowledge that these analyses were part of a longitudinal QC study. That is not to say this would be the case with an experimental work-flow but it is certainly a factor worthy of further exploration and a clear target to assist in guiding process improvement.

5. Conclusions

Statistical process control is a very versatile QC technique that utilises simple but powerful charts and rules. It has been proven time and again to have high value and applicability, including routine use in clinical diagnostic laboratories for the assurance of quantitative and qualitative assays. This article has demonstrated that through the use of some simple techniques and the monitoring of some simple derived metrics we are able to reveal significant facts about a data set. These same metrics could be measured on-line to provide early warning of developing issues with little ongoing effort. Implementation of similar techniques in other industrial sectors has clearly shown that their regular use leads to increasing process and output quality.

The process of defining 'quality' in any particular setting and the consideration of factors which may impact upon this is valuable in its own right. Spending time as a research team simply trying to carry this out and considering rules which objectively report upon 'quality' can improve experimental designs and procedures, even if formal control rules are not implemented. But, it must always be remembered that 'One gets what one measures' — if you do not measure quality, then you will not get it.

Acknowledgements

The author would like to acknowledge Dr. Ben Chaffey, Dr. David Jackson and Will Dracup at Biosignatures for proofreading and the discussion of concepts and principles. Additional acknowledgement is given to Mark Lambert, Bill Hally and David Miller at Biosignatures for algorithm development, technical infrastructure and discussions.

The author would also like to acknowledge Dr. Robert Tonge, now at Waters Corporation, for both supervising the original study from which the data set is derived, and many discussions on approaches to ensuring the quality of proteomics data. The author would also like to acknowledge contributions from the remaining authors of the original paper and thank them for their decision to make the data and details publicly available.

Disclosure statement

The author is an employee of Biosignatures Ltd. and he has commercial interests in Nonlinear Dynamics Ltd.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2013.06.010.

REFERENCES

[1] Allison F. An historical review of quality control in hematology. Am J Med Technol 1983;49(9):625.
[2] Westgard J, Westgard S. The quality of laboratory testing today. Am J Clin Pathol 2006;125(3):343–54.
[3] Green G, Carey R, Westgard J, Carten T, Shablesky L, Achord D, et al. Quality control for qualitative assays: quantitative QC procedure designed to assure analytical quality required for an ELISA of hepatitis b surface antigen. Clin Chem 1997;43(9):1618–21.
[4] Shewhart W. Statistical Method From the Viewpoint of Quality Control. Dover Publications; 1986 .
[5] Westgard J, Barry P, Hunt M, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. Clin Chem 1981;27(3):493–501.
[6] NIST/SEMATECH. e-Handbook of Statistical Methods. URL http://www.itl.nist.gov/div898/handbook/; 2012.
[7] Jackson D, Bramwell D. Application of clinical assay quality control to multivariate proteomics data: a workflow exemplified by 2-DE QC. J Proteomics 2013;95(1):22–37.
[8] Alban A, David S, Bjorkesten L, Andersson C, Sloge E, Lewis S, 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(1):36–44.
[9] Jackson D, Herath A, Swinton J, Bramwell D, Chopra R, Hughes A, et al. Considerations for powering a clinical proteomics study: normal variability in the human plasma proteome. Proteomics Clin Appl 2009;3(3):394–407.
[10] NLD. SameSpots Software v4.5. Accessed 27/09/2012; URL http://www.nonlinear.com/products/progenesis/samespots/overview/; 2012.
[11] R Foundation for Statistical Computing, Vienna, Austria. R: A Language, Environment for Statistical Computing3-900051-07-0; 2012 [accessed 28/12/2012; URL http://www.R-project.org/].
[12] Harrell Jr Frank E. rms: Regression Modeling Strategies. R package version 3.5-0; URL http://CRAN.R-project.org/package=rms; 2012.
[13] Huber W. Variance stabilization and calibration for microarray data. R package version 3.24.0 http://www.bioconductor.org/packages/2.10/bioc/manuals/vsn/man/vsn.pdf; 2012 . [Accessed 28/12/2012; URL].
[14] Huber W, Von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002;18(Suppl. 1):S96–S104.
[15] Fox J, Weisberg S. An R Companion to Applied Regression. URL . 2nd ed.Thousand Oaks CA: Sage; 2011 [URL http://socserv.socsci.mcmaster.ca/jfox/Books/Companion].
[16] Shewart W. Economic control of quality of manufactured product. Van Nostrad; 1931 .
[17] Levey S, Jennings E. The use of control charts in the clinical laboratory. Am J Clin Pathol 1950;20(11):1059–66.
[18] Tonks DB. A study of the accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories. Clin Chem 1963;9(2):217–33.
[19] Ma ZQ, Polzin KO, Dasari S, Chambers MC, Schilling B, Gibson BW, et al. QuaMeter: multivendor performance metrics for LC–MS/MS proteomics instrumentation. Anal Chem 2012;84(14):5845–50.
[20] ABRF Proteomics Research Group. Overview of studies. Accessed 23/05/2013; URL http://www.abrf.org/index.cfm/group.show/Proteomics.34.htm; 2013.
[21] Bunger MK, Andacht TM, Bennett KL, Bystrom C, Dangott L, Leszyk J, et al. The 2012 PRG Study: Assessing Longitudinal Variability in Routine Peptide LC–MS/MS Analysis. Accessed 23/05/2013; URL http://www.abrf.org/ResearchGroups/Proteomics/Studies/prgposterABRFconferencefinal.pdf; 2012.
[22] Bunger MK, Andacht TM, Bennett KL, Bystrom C, Dangott L, Leszyk J, et al. The 2012–13 PRG Study: Assessing Longitudinal Variability in Routine Peptide LC–MS/MS Analysis. Accessed 23/05/2013; URL http://www.abrf.org/ResearchGroups/Proteomics/Studies/2013_ABRF_PalmSprings_Talk.pdf; 2013.