Journal of Chromatography B, 877 (2009) 1299–1305

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Real-time evaluation of experimental variation in large-scale LC–MS/MS-based quantitative proteomics of complex samples☆

Yishai Levin[a,*], Lan Wang[a], Erin Ingudomnukul[b], Emanuel Schwarz[a], Simon Baron-Cohen[b], András Palotás[c], Sabine Bahn[a]

[a] Institute of Biotechnology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QT, UK
[b] Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge CB2 8AH, UK
[c] Asklepios-Med Bt (Private Practice and Research Centre), H-6722 Szeged, Kossuth Lajos sgt. 23, Hungary

ARTICLE INFO

Article history:
Received 19 June 2008
Received in revised form 29 October 2008
Accepted 7 November 2008
Available online 13 November 2008

Keywords:
Proteomics
LC–MS
MS^E
NanoUPLC
Serum
Fibroblasts
QC
Serum
Biomarker discovery

ABSTRACT

Quantitative proteomic profiling is becoming a widely used approach in systems biology and biomarker discovery. There is a growing realization that quantitative studies require high numbers of unpooled samples for increased statistical power. Large-scale quantitative analyses require an added degree of stringency due to the lengthy study periods and reliance on stability of analytical instrumentation. We present the inclusion of quality control samples alongside clinical samples in the preparation and nanoLC–MS analysis pipelines. These serve the purpose of monitoring, evaluating and reporting experimental variation measured in real-time. This concept is shown for two types of complex biological samples: serum samples and fibroblast samples. In both studies QC samples were added among dozens of clinical ones and analyzed using a label-free quantitative proteomic platform.

1. Introduction

In the past few years, reliable identification of proteins in complex biological samples has become routine practice in many laboratories across the world. Such qualitative information provides researchers with insights into the intricate structures of biological systems. However, many applications, such as disease biomarker discovery, also require the measurement of relative abundance of proteins. Recent advances in analytical instrumentation and bioinformatics now enable the relative quantitation of hundreds of proteins across dozens of samples in a single experiment [1–5]. Thus it is possible to obtain both qualitative and quantitative information providing deeper and deeper insights into the origin and structure of biological systems as well as allowing global proteomic profiling for identification of disease specific biomarkers. Regardless of the method of choice, quantitative analysis requires an added degree of stringency, especially when analyzing complex samples. In order for proteomic profiling experiments to be statistically powered, it is necessary to investigate long sample cohorts. This means that run times are longer, sometimes lasting several consecutive days and even weeks [3,4]. Such complicated and lengthy analyses require close monitoring of experimental variation to avoid the introduction of systematic bias which may result in the identification of artefactual changes. In order to produce reliable and accurate information, technical issues must be addressed prior to and during quantitative studies, similar to the guidelines for pharmacological quantitative assays [6]. For this reason and others, emphasis is now placed on the need to report coefficients of variation, relative standard deviation (RSD) and other variance estimations in publications [7]. Additionally, the search for disease biomarkers is subject to an ongoing debate as to what is considered a significant alteration in abundance. Assessment of experimental variation, specifically for every peptide/protein, can

Abbreviations: LC, liquid chromatography; MS, mass spectrometry; MS^E, type of MS acquisition mode (expression mode); kDa, kilo Dalton; SD, standard deviation; RSD, relative standard deviation; QC, quality control; PLGS, Proteinlynx Global Server; FCS, fetal calf serum; GPS, L-glutamine, penicillin, streptomycin solution; DDA, data dependant analysis; RPMI, Roswell Park Memorial Institute; FPR, false positive rate; TPR, true positive rate.

☆ This paper is part of the special issue "Quantitative Analysis of Biomarkers by LC–MS/MS", J. Cummings, R.D. Unwin and T. Veenstra (Guest Editors).

* Corresponding author. Tel.: +44 1223 767799; fax: +44 1223 334162.
E-mail address: yl334@cam.ac.uk (Y. Levin).

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.11.007

Table 1
Elution gradient used for fibroblast and serum samples.

Time (min)	Flow (μl/min)	%A	%B	Curve
For fibroblast samples				
Initial	0.3	97	3	Initial
1	0.3	97	3	Linear
100	0.3	70	30	Linear
115	0.3	5	95	Linear
125	0.3	5	95	Linear
126	0.3	97	3	Linear
For serum samples				
Initial	0.3	95	5	Initial
1	0.3	95	5	Linear
80	0.3	70	30	Linear
90	0.3	5	95	Linear
104	0.3	5	95	Linear
105	0.3	95	5	Linear

be of great help in distinguishing significant changes from experimental artefacts.

In this manuscript we suggest the implementation of quality control (QC) samples into the experimental design of LC–MS/MS-based quantitative proteomic profiling studies. We show that platform performance, in some cases including variation in sample preparation steps, can be evaluated throughout the analysis period. We show that valuable information can be obtained from QC samples, such as intensity RSD and other performance ratings in an analyte specific manner.

The examples shown in this manuscript were analyzed using a label-free platform. The addition of QC samples however can be implemented to other quantitative techniques where a large set of samples is analyzed.

We present two separate studies, both using a nanoLC–MS^E platform, where QC samples were analyzed in conjunction with clinical ones. The first is a set of 28 fibroblast samples, which included 22 clinical samples and 6 quality controls. The second is a set of 39 serum samples, comprised of 32 clinical samples and 7 quality controls. Both studies were performed in order to identify protein disease biomarkers.

2. Experimental procedures

All study participants gave their written informed consent. All clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki.

Two sets of samples were used to describe the implementation and use of quality-control samples in quantitative proteomics. The first was a set of 11 clinical primary fibroblast obtained from schizophrenia patients and 11 samples collected from matched healthy volunteers. The QC samples were aliquoted from one clinical sample, after preparation and prior to nanoLC–MS^E analysis. One QC followed every forth randomized clinical sample, for a total of six QCs.

The second set consisted of a total of 39 sera samples, comprising 12 samples taken from patients diagnosed with Asperger Syndrome and 20 samples taken from matched healthy volunteers. A fraction of each clinical sample was pooled and then aliquoted. These aliquots served as the quality control samples (one QC for every fifth randomized clinical sample, including one at the end of the analysis). The seven QC samples were prepared and analyzed identically to randomized clinical samples. Because they were aliquoted prior to all preparation steps in this study, the experimental variation measured for these QCs included all experimental steps. The key of one QC to every five clinical sample was chosen on the basis of having enough QCs without prolonging the analysis.

2.1. Fibroblasts

Human fibroblasts were cultured from skin biopsies obtained from Asklepios-Med Bt (private practice and research centre), Hungary. Cells were isolated in supplemented RPMI 1640 medium (Sigma, St. Louis, MO) containing 10% fetal calf serum (FCS), 1% L-glutamine, penicillin, streptomycin solution (GPS) (Sigma, St. Louis, MO) in a 25-cm^2 culture flasks. After 2 weeks, fibroblasts were detached with 0.25% Trypsin/0.02% EDTA-solution (Sigma, St. Louis, MO) and further cultivated in 75-cm^2 flasks (5×10^5) in 5% CO_2 at 37 °C in supplemented RPMI 1640. Fibroblasts were detached with 0.25% Trypsin/0.02% EDTA-solution when cells reached 80–90% confluence. Media were changed every 3 days.

A sub-cellular proteome extraction kit (S-PEK, Merck KGaA, Darmstadt, Germany) was used to fractionate the samples into four cellular fractions [8]. Cell pellets (3×10^6) were collected and washed with wash buffer followed by total cell lysis, in accordance with the manufacturer's protocol. This resulted in four fractions for each sample: cytosolic, membrane, nucleus and cytoskeleton. The total protein concentration was determined for each fraction using a protein assay (Bio-Rad DC assay).

Buffer exchange into 50-mM ammonium bicarbonate (Sigma, St. Louis, MO) was performed for 200 μg of the cytosolic fraction of each sample using 4-ml spin concentrators with a 5 kDa molecular weight cut off (Agilent, Palo Alto, CA). Samples were then incubated with 2% Rapigest (Waters, Milford, MA) for 15 min at 80 °C. Reduction of proteins was then performed using final concentration of 5 mM dithiothreitol (Sigma, St. Louis, MO) and incubated at 60 °C for 30 min. Iodoacetamide (Sigma, St. Louis, MO) was then added to a final concentration of 10 mM in the dark and incubated at room temperature for 30 min. Finally, 4-μl sequencing grade trypsin (Promega, Madison, WI) was added and samples incubated at 37 °C overnight. Following digestion, 8.8 M HCl was added to a final concentration of 200 mM. Samples were stored at −80 °C until analysis. Prior to LC–MS^E analysis, 25 fmol/μl *Saccharomyces cerevisiae* Enolase tryptic digest was added to each sample, the MS signal of which was used later for normalization. Pure *Saccharomyces cerevisiae* Enolase digest was injected with identical LC and MS methods to assure correct peptides were selected for normalization (data not shown).

2.2. Serum samples

The total protein concentration of each serum sample was measured using a protein assay (Bio Rad) prior to preparation. Each sample was depleted of the 20 most abundant proteins using immunoaffinity chromatography (ProteoPrep20, Sigma, St. Louis, MO), loading a total of 560 μg of protein of each sample onto the depletion column. Buffer exchange was performed with 50 mM ammonium bicarbonate using spin columns (Millipore, Bedford, MA) with a 5 kDa molecular weight cut-off. The proteins were reduced using 5 mM dithiothreitol (Sigma, St. Louis, MO) at 60 °C for 30 min. 10 mM iodoacetamide was added (Sigma, St. Louis, MO) in the dark at room temperature for 30 min. The proteins were digested using Trypsin (Promega, Madison, WI), at a ratio of 1:50 (w/w Trypsin/Protein) for 16 h at 37 °C. Digestions were stopped by adding 2.3 μl of 8.8 M HCl to each sample. The samples were stored at −80 °C until LC–MS^E analysis. As for the fibroblast samples, 25 fmol/μl *Saccharomyces cerevisiae* Enolase tryptic digest was added to each sample. During data analysis, the intensity measurements of the Enolase peptides were used for normalization, in the same manner as the fibroblast sample set.

2.3. Liquid chromatography

For all chromatographic steps, HPLC grade solvents were used (H_2O from Sigma, St. Louis, MO; Acetonitrile from Fisher Scientific).

Each sample was injected and analyzed three times followed by a blank injection (to ensure there is no carry-over of peptides from one sample to the other in this sequential process). For each sample 0.6 μg of total protein digest was loaded using split-less nano Ultra Performance Liquid Chromatography (10 kpsi nanoAcquity, Waters, Milford, MA). Buffers used were A: H_2O + 0.1% formic acid; B: acetonitrile + 0.1% formic acid. Desalting of the samples was performed online with 100% buffer A for 3 min, using an online Reverse-Phase C18 trapping column (180 μm i.d., 20 mm length and 5 μm particle size) (Waters, Milford, MA). The samples were separated using a C18 nanoColumn (75 μm i.d., 200 mm length, 1.7 μm particle size) (Waters, Milford, MA), using the gradient listed in Table 1.

2.4. Mass spectrometry

The nanoUPLC was coupled online through a nanoESI emitter of 7 cm length and 10 μm tip (New Objective, Woburn, MA) to a Quadrupole Time-of-Flight Mass Spectrometer (Qtof Premier, Waters, Milford, MA). Data were acquired in MS^E (Expression) mode. In this mode, the quadrupole is set to transfer all ions while the collision cell switches from low to high collision energy intermittently throughout the acquisition time. In the low energy scans, collision energy was set to 4 eV while in the high energy scans it was ramped from 20 to 43 eV. This mode enables accurate mass measurement of both intact peptides as well as fragments, and conservation of the chromatographic profile for both intact peptides and fragments.

Mass accuracy was maintained throughout the analysis by the use of a LockSpray apparatus. A reference compound (Glu-Fibrinopeptide B, Sigma, St. Louis, MO) was continuously infused using the LockSpray and scanned intermittently every 30 s. During data processing, the analyte spectra were corrected based on the difference between the detected m/z peak and the theoretical m/z peak (785.8426 [m + 2H]+) of Glu-Fibrinopeptide B.

2.5. Data processing and protein identification

Raw data, acquired in continuum format, were processed using the ProteinLynx Global Server software version 2.3 (also known as Identity^E) (Waters, Milford, MA). Both quantitative and qualitative information were produced automatically by the software, using the default parameters.

2.6. Quantitative information

Intensity measurements were obtained by integration of the total ion volume of each extracted, charge-state-reduced, deisotoped and mass corrected ion across the mass spectrometric and chromatographic volume (Fig. 1), as opposed to two-dimensional integration of extracted ion chromatograms (XIC). The algorithm calculates the observed mass and intensity measurement deviation for every detected component. The chromatographic area associated with each component is calculated using an integration algorithm similar to the ApexTrack peak integration algorithm provided in the MassLynx software. If a particular component exists in more than one charge-state, the corresponding area for any given monoisotopic ion is reported as the summed area from all contributing charge states. The retention time is determined for each reported monoisotopic ion at the moment it reaches its maximum intensity (apex). This process is performed for both the low collision energy and the high energy scans (saved as separate channels). The ion detection thresholds were set as follows: low energy ion detection threshold of 250 counts; high energy ion detection threshold was set to 100.

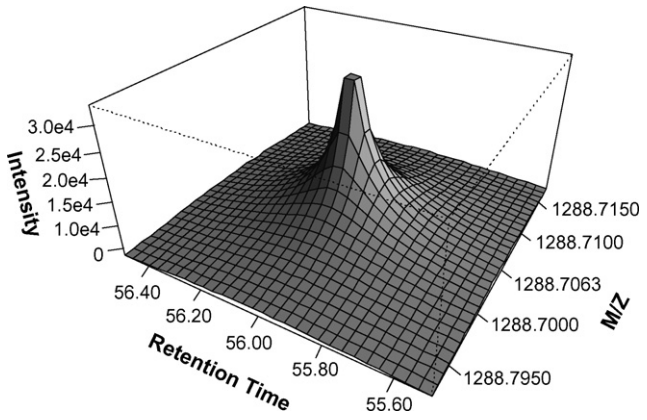

Fig. 1. The plot shows a theoretical peak showing the concept of integration across the mass spectrometric and chromatographic volumes of each charge-state-reduced, deisotoped mass corrected ion.

In this type of acquisition chromatographic profile is maintained reproducibly throughout the sample set, thus it is possible to align the precursor ions in time and then directly compare the intensities of precursor ions across all injections of all samples. The data was normalized based on the internal standard – digested *Saccharomyces cerevisiae* Enolase – that was added to each sample.

The data set was then filtered using the free software package R (www.r-project.org) and only peptides that were detected in at least two out of three injections of each sample and at least 70% of the samples were included in the analysis. Those peptides that did not pass these filtering criteria were excluded from the analysis as the quantitative information they generate is of low confidence due to poor replication. Protein quantitation was performed by summing the intensities of all peptides of a given protein, which were not identified as phosphorylated, that passed the above filtering criteria, and a minimum of two peptides.

Variation was calculated by dividing the standard deviation of protein intensity by the mean of protein intensity, multiplied by 100 (represented as relative standard deviation, % RSD). This calculation was used in both studies to assess the experimental variation (with no biological variation as the data was obtained from QC samples which were identical).

2.7. Protein identification

Proteinlynx Global Server version 2.3 (Identity^E) was also used for database searches. The database used for searching was the Human IPI version 3.34 (October 2007), appended with the sequence of *Saccharomyces cerevisiae* Enolase. The total number of entries in the database was 67,757.

The database search algorithm of the software was described by Li et al. [9] and Vissers et al. [10]. Briefly, the software detects the 250 most abundant peptides and performs an initial pass through the database in order to identify those peptides (with mass tolerance of 10 ppm of precursor ions and 20 ppm for fragment ions). It then calculates the precursor-ion mass tolerance, fragment-ion mass tolerance and chromatographic peak widths for these 250 peptides. These peptides are then depleted from the database and the remaining peptides are searched based on these criteria. The

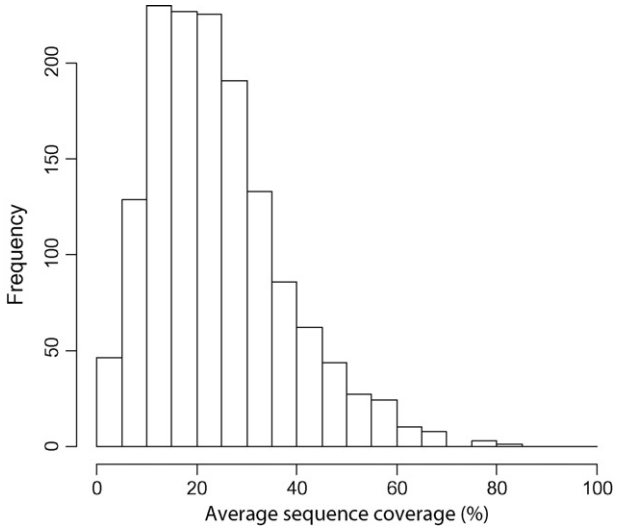

Fig. 2. A histogram showing the frequency of average sequence coverage for all 1446 identified proteins, across all injections of all fibroblast samples.

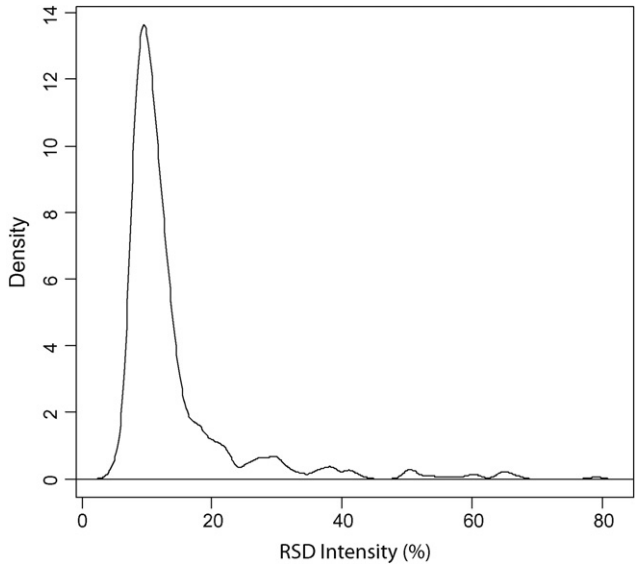

Fig. 3. Density plot of the RSD of intensity for the 846 proteins identified in the serum samples (based on the averaged intensity of peptides which passed the filtering criteria). It can be seen that most of the proteins have intensity RSD lower than 20%, with an average of 14%.

cycle continues to the next abundant peptides, which are identified and then depleted from the database. These tentative peptide identifications are ranked and scored by how well they conform to 14 predetermined models of specific, physicochemical attributes (such as retention time and fragmentation pattern). All tentative peptides are collapsed into their parent proteins utilizing only the highest scoring peptides that contribute to the total protein score. Once a protein has been securely identified, all top-ranked precursor ions and their corresponding product ions are removed from all other tentatively identified proteins. The remaining unidentified peptides – and tentatively identified proteins – are then re-ranked and re-scored, and the process is repeated until a 4% false positive rate is reached. The false discovery rate is determined by the number of random or reverse identifications identified (false positive rate, FPR) divided by the number of correct identifications, (true positive rate, TPR) expressed as a percentage. Therefore the false discovery rate (FDR) equals the FDR = FPR/TPR × 100.

All protein identifications were based on at least two peptides. Since this search algorithm is not probability base there is no need for a hard cut-off for selection of individual MS/MS spectra.

3. Results

3.1. Fibroblast samples

As in any label-free quantitative method, the samples in both studies were analyzed discretely (without pooling), sequentially and randomly. The fibroblast nanoLC–MS^E analysis was completed in 299 consecutive hours. The QC samples in this study were aliquoted from one clinical sample, following sample preparation described in the previous section, and prior to nanoLC–MS^E analysis. For this reason, the QCs in this study represent only variation derived from LC–MS analysis, and not from previous preparation steps. Every forth clinical sample was followed by a QC. The choice of the rate in which a QC would be analyzed was made so as not to prolong the analysis too much and yet analyze enough QC samples indicating the experimental variation across the entire study period.

The total number of proteins identified in the QCs in this study was 1446, identified by at least two peptides and the criteria outlined in the "Experimental Procedures". Among these, 846 proteins were identified with 6025 peptides that passed the criteria listed in the methods section (see supplementary data). The average sequence coverage of these proteins was 27 ± 13% (Fig. 2).

Among the valuable information that can be obtained from the QC samples is the variation in detected intensity, calculated as the (intensity standard deviation/average intensity) × 100 = relative standard deviation (RSD). This calculation was performed for the intensity measurements in the QC samples and is therefore an accurate indication of experimental variation since the QC samples were analyzed alongside the clinical ones and during the same period of time.

The intensity RSD was calculated specifically for every detected protein, based on the averaged intensities of all peptides per protein. The averaged RSD of intensity for the 846 proteins was 14%. Fig. 3 shows a density plot of these measurements. Eighty-eight percent of the proteins were detected with an RSD of 20% or lower. The same calculation can be performed for every detected peptide, however since the ultimate goal of the analysis is the relative quantitation of proteins this would not be practical information.

The QC samples also provide insight into the performance of the chromatography part of the system during the study. The retention time standard deviations (SD) were calculated for all peptides passing the filtering criteria (i.e. detection in two out of three injections of each sample and detection in at least four out six QCs). Fig. 4 shows a density plot of the retention time SDs as detected in the QCs. It can be seen that the retention time shift during the 12 consecutive days of analysis was less than ±0.7 min.

3.2. Serum samples

The total nanoLC–MS^E analysis time of the serum samples was 346 consecutive hours. Similar to the fibroblast study, one QC sample was analyzed following every five clinical samples for a total of seven QCs. However, unlike the fibroblast experiment, the QCs in this study were aliquoted from one serum sample *prior* to any preparation. This was done to account for the *overall* experimental variation (including sample preparation and nanoLC–MS^E analysis) measured during the analysis of the clinical samples.

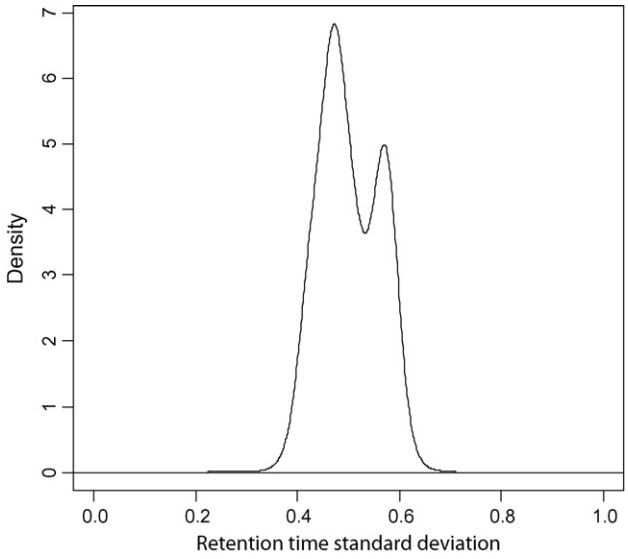

Fig. 4. Density plot of the retention time standard deviation for the 6025 peptides which passed the filtering criteria. It can be seen that most of the peptides had a standard deviation lower than 0.7.

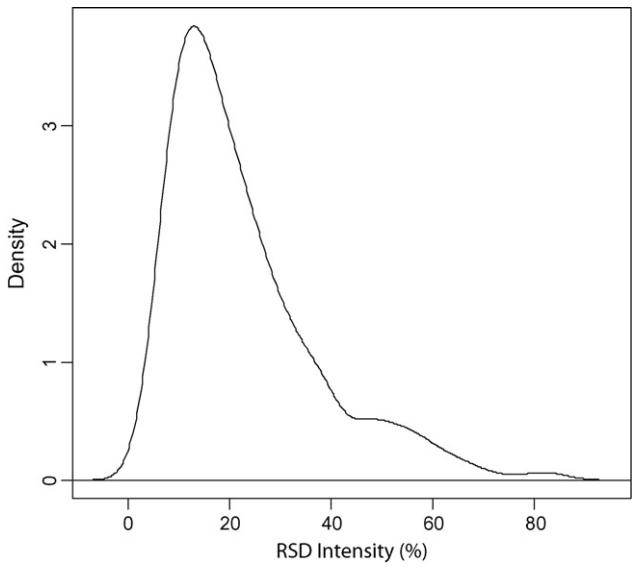

Fig. 6. Density plot of the RSD of intensity for the 307 proteins identified in the serum samples (based on the averaged intensity of peptides which passed the filtering criteria). It can be seen that most of the proteins have intensity RSD lower than 40%, with an average of 22%.

The total number of proteins identified in the serum QC samples was 579. All proteins were identified with at least 2 peptides and based on criteria outlined in the "Experimental Procedures". A total of 2854 peptides were detected in at least 2 out of 3 injections of each sample and in at least 5 out of 7 QC samples, which were used to identify 307 proteins (minimum of 2 peptides per protein, see supplementary data). The average sequence coverage of all detected proteins was $30.5 \pm 14\%$. Fig. 5 shows a histogram of the detected sequence coverage for all 579 proteins.

The RSD of intensity was calculated specifically for every detected protein, based on the averaged intensities of all peptides. Fig. 6 shows a density plot of these measurements. It can be seen that 88% of the proteins have a variation in measured intensity lower than 40% with a mean of 22% RSD.

The retention time standard deviations were calculated for the serum sample set as well. This calculation was performed for all peptides passing the filtering criteria. Fig. 7 shows a density plot of the SD of retention times as detected in the QCs. The retention time shift during the 14 consecutive days of analysis was ± 0.3 min for most peptides.

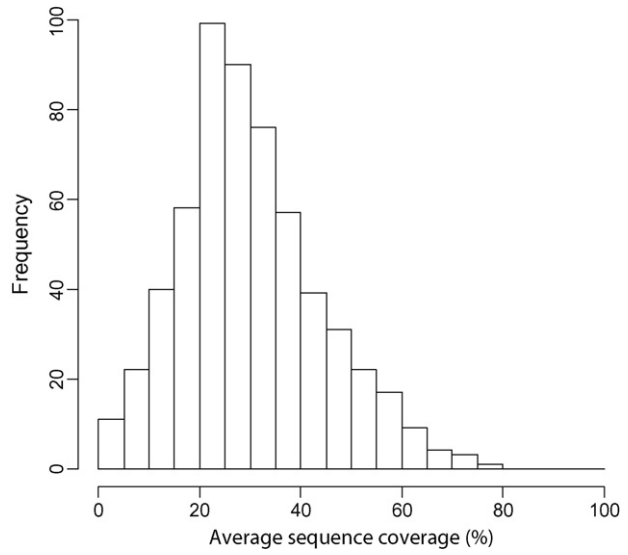

Fig. 5. A histogram showing the frequency of average sequence coverage for all 579 identified proteins, across all injections of all serum samples.

4. Discussion

Quantitative large-scale proteomic profiling is gaining momentum at an ever-increasing rate [11,12]. The valuable information that can be obtained from studies utilizing quantitative proteomic techniques has enabled researchers to break new ground in the understanding of complex molecular and cellular mechanisms. Furthermore, it is now used more and more for the purpose of disease biomarker discovery. The parallel process of quantitatively comparing hundreds of proteins across multiple samples enables the

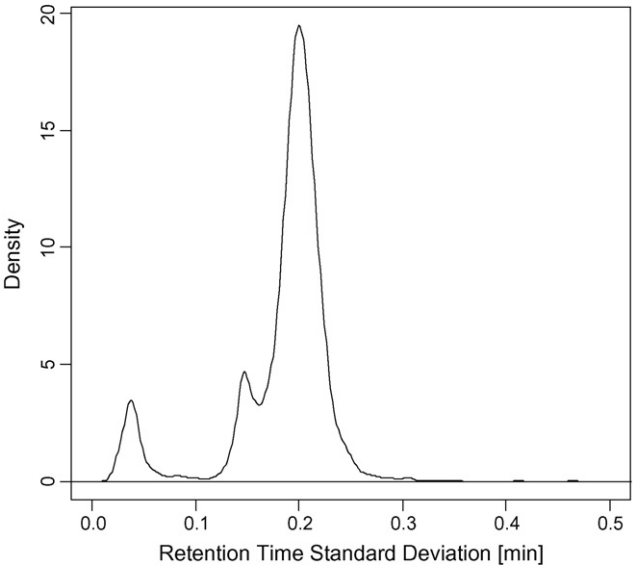

Fig. 7. Density plot of the retention time standard deviation for the 2854 peptides that passed the filtering criteria. It can be seen that most of the peptides had a standard deviation lower than 0.3 min.

discovery of novel disease markers that would have otherwise taken years to identify. However, it is now realized that the comparison of hundreds of variables (i.e. proteins) across a handful of observations (i.e. samples) deems such experiments to be statistically underpowered [7]. The consequence of this can in some cases be disappointing due to large numbers of identified candidate markers that do not hold up to validation [13]. New technologies and techniques now enable the quantitation of hundreds of proteins across dozens of samples [1–4] thus strengthening statistical power, enabling the use of complex statistical clustering tools and producing high quality data. However, such large-scale quantitative analyses should be embarked on with the utmost care. Regardless of the platform used, the preparation and analysis of large numbers of samples is a process that often takes several weeks for any given experiment. In order to be sure that experimental variation does not interfere or skew the quantitative analysis, this variation should be monitored in real time. Furthermore, this information can be easily extracted and reported thus enhancing the validity of results.

In this manuscript we suggest and implement the incorporation of quality control (QC) samples to large-scale quantitative proteomic profiling studies. These are identical samples of the same biological origin as the clinical samples being analyzed. Since they are identical, any variation in the detection of analytes in these samples would represent purely experimental variation. By adding these among clinical samples it is possible to monitor experimental variation during the study, in real-time.

We demonstrated this using two separate studies. In the fibroblast study, the QC samples represented only the nanoLC–MS^E variation, where the sera QCs represented the entire experimental procedure from crude serum to raw data.

We were able to show that the average variation of intensity measurement was only 14% RSD. This result reflects the integrity of the analytical platform as well as its reproducibility. Furthermore, it stresses the ability of the platform to maintain performance over long run periods.

In the case of the serum study, where a total of 39 samples were prepared and analyzed sequentially (including the QCs), the average variation was 22% RSD. The reason for slightly higher variation compared with the fibroblast samples is due to the fact that the serum QCs included variation caused by sample preparation. Furthermore, the sera samples were depleted of the 20 most abundant proteins, a procedure which introduces experimental variation, a step that is not necessary for the fibroblast preparation.

In both studies, the total number of proteins identified was comparable with methods based on two-dimensional liquid chromatography, although we used only one-dimensional separation [14]. The high duty cycle of the MS^E acquisition enabled efficient sampling of the data. We could show that for both studies the majority of the proteins could be identified based on peptides that replicated well across the samples with high sequence coverage.

Beyond the overall assessment of the variation, the QCs can be used to obtain the variation information of a specific protein of interest, given that it was detected in both the QCs and clinical samples (as is the case for many of the detected proteins. For the purpose of protein biomarker discovery, this type of information is useful to determine biologically significant alterations in protein abundance. Since the information of pure experimental variation is readily accessible, specifically at the protein level, it makes the task of determining biological alteration from experimental variation easier. For example, in the serum QC samples the protein having the highest RSD of intensity was the TrEMBL entry Q6PJA4 (IGHG1). Five peptides of this protein were identified with an averaged RSD of intensity of 84%. The variation is visualized in Fig. 8. Had this protein been identified as significant, with a fold change

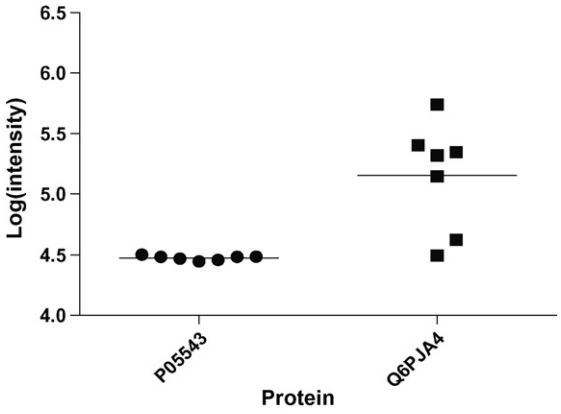

Fig. 8. The plot shows the log of intensities for proteins P05543 and Q6PJA4. P05543 has an RSD of 4% and Q6PJA4 84%. Both were detected in the sera QCs.

lower than the above variation in the clinical samples, one should question the validity of the result due to the high experimental variation. However, the protein with the lowest experimental variation detected in the serum QC samples had 4% RSD of intensity. This protein, thyroxine-binding globulin (Swissprot accession P05543), was identified with 14 peptides. Had this protein been identified as statistically significant, one could be more confident that it is due to biological and not experimental variation.

Our results also imply that the earlier the QCs are introduced to the analysis pipeline, the more experimental variation they represent and hence the information generated is a more realistic presentation of the experimental variation.

QC samples can also be used for normalizing intensities in a given study and also to normalize across separate sample sets as reported by Van Der Greef et al. [3]. However for this purpose, the QCs must be acquired at a higher rate than that proposed in this manuscript, which would mean much longer analyses.

In conclusion, the addition of QC samples is a straightforward task, regardless of the platform used. This additional step is not time consuming in view of the overall analysis and reveals extremely useful information specifically for every detected protein, as well as information of the overall performance of the platform throughout the study.

Acknowledgments

The research was kindly supported by the Stanley Medical Research Institute (SMRI). We also thank Psynova Neurotech® for centre support and for Ph.D. funding.

We thank the Nancy Lurie Marks Family Foundation and the Medical Research Council for funding the specimen collection.

Dr. Mark McDowal (Waters Corp.) is kindly acknowledged for his help and support as well as the service team from Waters corp. for their technical support.

We would also like to present our appreciation to Dr. Paul Guest, head of Proteomics at the Cambridge Center for Neuropsychiatric Research, for the helpful discussions and suggestions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2008.11.007.

References

[1] W. Wang, H. Zhou, H. Lin, S. Roy, T.A. Shaler, L.R. Hill, S. Norton, P. Kumar, M. Anderle, C.H. Becker, Anal. Chem. 75 (2003) 4818.

[2] G. Wang, W.W. Wu, W. Zeng, C.L. Chou, R.F. Shen, J. Proteome Res. 5 (2006) 1214.
[3] J. van der Greef, S. Martin, P. Juhasz, A. Adourian, T. Plasterer, E.R. Verheij, R.N. McBurney, J. Proteome Res. 6 (2007) 1540.
[4] Y. Levin, E. Schwarz, L. Wang, F.M. Leweke, S. Bahn, J. Sep. Sci. 30 (2007) 2198.
[5] K.C. Chan, D.A. Lucas, D. Hise, C.F. Schaefer, G.M.J. Zhen Xiao, K.H. Buetow, T.D.V. Haleem, J. Issaq, T.P. Conrads, Clin. Proteom. 1 (2004) 101.
[6] R.E. Aarnoutse, C.P. Verweij-van Wissen, E.W.J. van Ewijk-Beneken Kolmer, E.W. Wuis, P.P. Koopmans, Y.A. Hekster, D.M. Burger, Antimicrob. Agents Chemother. 46 (2002) 884.
[7] M.R. Wilkins, R.D. Appel, J.E. Van Eyk, M.C. Chung, A. Gorg, M. Hecker, L.A. Huber, H. Langen, A.J. Link, Y.K. Paik, Proteomics 6 (2006) 4.
[8] S.B. Man, K. Chan, Curr. Proteom. 4 (2007) 131.
[9] G.-Z. Li, D. Golick, M.V. Gorenstein, J.C. Silva, a.G.J.P.C.S.J. Vissers, Proceedings of the Human Proteome Organisation (HUPO) 5th Annual World Congress, Long Beach, CA, 2006.
[10] J.P.C. Vissers, J.I. Langridge, J. Aerts, Mol. Cell. Proteom. 6 (2007) 755.
[11] J. Cox, M. Mann, Cell 130 (2007) 395.
[12] W.J. Qian, J.M. Jacobs, T. Liu, D.G. Camp II, R.D. Smith, Mol. Cell. Proteom. 5 (2006) 1727.
[13] N. Rifai, M.A. Gillette, S.A. Carr, Nat. Biotechnol. 24 (2006) 971.
[14] T. Kislinger, A.O. Gramolini, D.H. MacLennan, A. Emili, J. Am. Soc. Mass Spectromet. 16 (2005) 1207.