EuPA Open Proteomics 8 (2015) 6–15

Contents lists available at ScienceDirect

EuPA Open Proteomics

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

Inter-laboratory evaluation of instrument platforms and experimental workflows for quantitative accuracy and reproducibility assessment

Andrew J. Percy[a], Jessica Tamura-Wells[a], Juan Pablo Albar[b,1], Kerman Aloria[c], Ardeshir Amirkhani[d], Gabriel D.T. Araujo[e], Jesus M. Arizmendi[f], Francisco J. Blanco[g,h], Francesc Canals[i], Jin-Young Cho[j], Núria Colomé-Calls[i], Fernando J. Corrales[k], Gilberto Domont[e], Guadalupe Espadas[l,m], Patricia Fernandez-Puente[g], Concha Gil[n], Paul A. Haynes[o], Maria Luisa Hernáez[n], Jin Young Kim[p], Arthur Kopylov[q], Miguel Marcilla[b], Mathew J. McKay[d], Mehdi Mirzaei[o], Mark P. Molloy[d], Leanne B. Ohlund[r], Young-Ki Paik[j], Alberto Paradela[b], Mark Raftery[s], Eduard Sabidó[l,m], Lekha Sleno[r], Daniel Wilffert[t], Justina C. Wolters[t], Jong Shin Yoo[p], Victor Zgoda[q], Carol E. Parker[a], Christoph H. Borchers[a,u,*]

[a] University of Victoria—Genome British Columbia Proteomics Centre, Vancouver Island Technology Park, #3101-4464 Markham St., Victoria, BC V8Z 7X8, Canada
[b] Proteomics Facility, Centro Nacional de Biotecnologia (CSIC), UAM Campus Cantoblanco, Darwin, 3, Madrid E-28049, Spain
[c] Proteomics Core Facility-SGIKER, University of the Basque Country (UPV/EHU), Leioa 48940, Spain
[d] Australian Proteome Analysis Facility, Macquarie University, Sydney NSW 2109, Australia
[e] Proteomics Unit of the Department of Biochemistry, Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro 21941-909, Rio de Janeiro, Brazil
[f] Department of Biochemistry and Molecular Biology, University of the Basque Country (UPV/EHU), Barrio Sarriena s/n, Leioa 48940, Spain
[g] Grupo de Proteómica-ProteoRed/Plataforma PBR2-ISCIII, Servicio de Reumatología, Instituto de Investigación Biomédica de A Coruña (INIBIC), Complexo Hospitalario Universitario de A Coruña (CHUAC), Sergas, Universidade da Coruña, As Xubias, A Coruña 15006, Spain
[h] CIBER-BBN/ISCIII, Instituto de Investigación Biomédica de A Coruña INIBIC-Hospital Universitario A Coruña, A Coruña 15006, Spain
[i] Laboratori de Proteòmica, Institut de Recerca Hospital Univ. Vall d'Hebron. Edifici Collserola, Pg. Vall d'Hebron, 119-129, Barcelona 08035, Spain
[j] Yonsei Proteome Research Center and Department of Biochemistry, College of Life Science and Biotechnology, Yonsei University, Seoul 120-749, South Korea
[k] Center for Applied Medical Research (CIMA), University of Navarra, ProteoRed-ISCIII, CIBERHED-ISCIII, IDISNA, Pío XII, 55; Ed. CIMA, Pamplona 31008, Spain
[l] Proteomics Unit, Center of Genomics Regulation (CRG), Carrer Dr. Aiguader 88, Barcelona 08003, Spain
[m] Proteomics Unit, Universitat Pompeu Fabra, Carrer Dr. Aiguader 88, Barcelona 08003, Spain
[n] Department of Microbiology II and Proteomics Unit UCM-PCM Complutense Unversity of Madrid, Plaza Ramón y Cajal s/n, Madrid 28040, Spain
[o] Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde NSW 2109, Australia
[p] Division of Mass Spectrometric Analysis, Korea Basic Science Institute, Cheongwon, South Korea
[q] Institute of Biomedical Chemistry, Pogodinskaya Street 10, Moscow 119121, Russian Federation
[r] Université du Québec à Montréal, CP 8888 Succ. Centre-Ville Montreal QC H3P 3P8, Canada
[s] Bioanalytical Mass Spectrometry, Wallace Wurth Building, University of New South Wales, Sydney NSW 2052, Australia
[t] Department of Pharmacy, Analytical Biochemistry, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands
[u] Department of Biochemistry and Microbiology, University of Victoria, Petch Building Room 207, 3800 Finnerty Rd., Victoria, BC V8P 5C2, Canada

ARTICLE INFO

Article history:
Received 27 March 2015
Received in revised form 1 June 2015
Accepted 10 June 2015
Available online 11 June 2015

ABSTRACT

The reproducibility of plasma protein quantitation between laboratories and between instrument types was examined in a large-scale international study involving 16 laboratories and 19 LC–MS/MS platforms, using two kits designed to evaluate instrument performance and one kit designed to evaluate the entire bottom-up workflow. There was little effect of instrument type on the quality of the results, demonstrating the robustness of LC/MRM-MS with isotopically labeled standards. Technician skill was a factor, as errors in sample preparation and sub-optimal LC–MS performance were evident. This highlights

* Corresponding author at: University of Victoria–Genome British Columbia Proteomics Centre, Vancouver Island Technology Park, #3101-4464 Markham St., Victoria V8Z 7X8, Canada. Fax: +1 250 483 3238.
E-mail address: christoph@proteincentre.com (C.H. Borchers).
[1] Deceased.

http://dx.doi.org/10.1016/j.euprot.2015.06.001

Keywords:
Accuracy
Human plasma
Quality control
Quantitative proteomics
Reproducibility
Workflow validation

the importance of proper training and routine quality control before quantitation is done on patient samples.

1. Introduction

Reproducible and accurate quantitation is a requirement for clinical and translational applications in targeted proteomics. The "gold standard" method in MS-based proteomics relies on multiple reaction monitoring (MRM) with stable isotope-labeled standards (SIS) incorporated within a bottom-up workflow [1]. Although MRM with labeled standards has been used for decades in small-molecule analysis (*e.g.*, hormones, drugs) [2,3], the use of this technique in proteomics introduces additional analytical variables related to sample preparation (*e.g.*, digestion) that increases the complexity of the sample analysis, thereby increasing the potential for interferences with the MRM transitions. To address this, standardization of proteomics methods has been encouraged by the Human Proteomics Proteome Organization's Plasma Proteome Project [4–10] and supported further by the US National Cancer Institute through the Clinical Proteomic Tumor Analysis Consortium [11,12]. This is necessary to enhance the global reproducibility of high quality data using different MS technologies.

As part of this standardization effort, we previously developed three standardization kits for instrument QC on a daily (Kit A) or monthly (Kits B and C) basis for LC–MS/MS platform (Kits A and B) and complete workflow (Kit C) assessment [13,14]. These kits contain a set of materials and analysis tools that enable value tracking and accuracy estimation by comparison with a set of reference values/ranges. Lyophilized pre-digested plasma standards that were spiked with a SIS peptide mixture (43 peptides from 43 human plasma proteins) are provided in Kits A and B, while raw starting materials (namely undepleted plasma, trypsin, and the SIS mix) are provided in Kit C for sample preparation and subsequent processing/analysis. In all cases, samples are processed by LC–MS/MS in a targeted or semi-targeted manner (*i.e.*, by using MRM on a triple quadrupole mass spectrometer or by using parallel reaction monitoring (PRM) on a hybrid quadrupole-Orbitrap mass spectrometer, respectively). These operative modes differ in that precursor-product ion pairs (*i.e.*, transitions) are sequentially monitored in MRM, whereas in PRM, full product ion spectra are collected from the collisional fragmentation of all target precursors in a given *m/z* window (see Fig. 1a and b in [15] for comparative schematics).

In this paper, we report the use of these three kits to evaluate the accuracy and reproducibility of a quantitative proteomics analysis of 43 high-to-moderate abundance plasma proteins in a bottom-up workflow, and to determine the source of errors if sub-

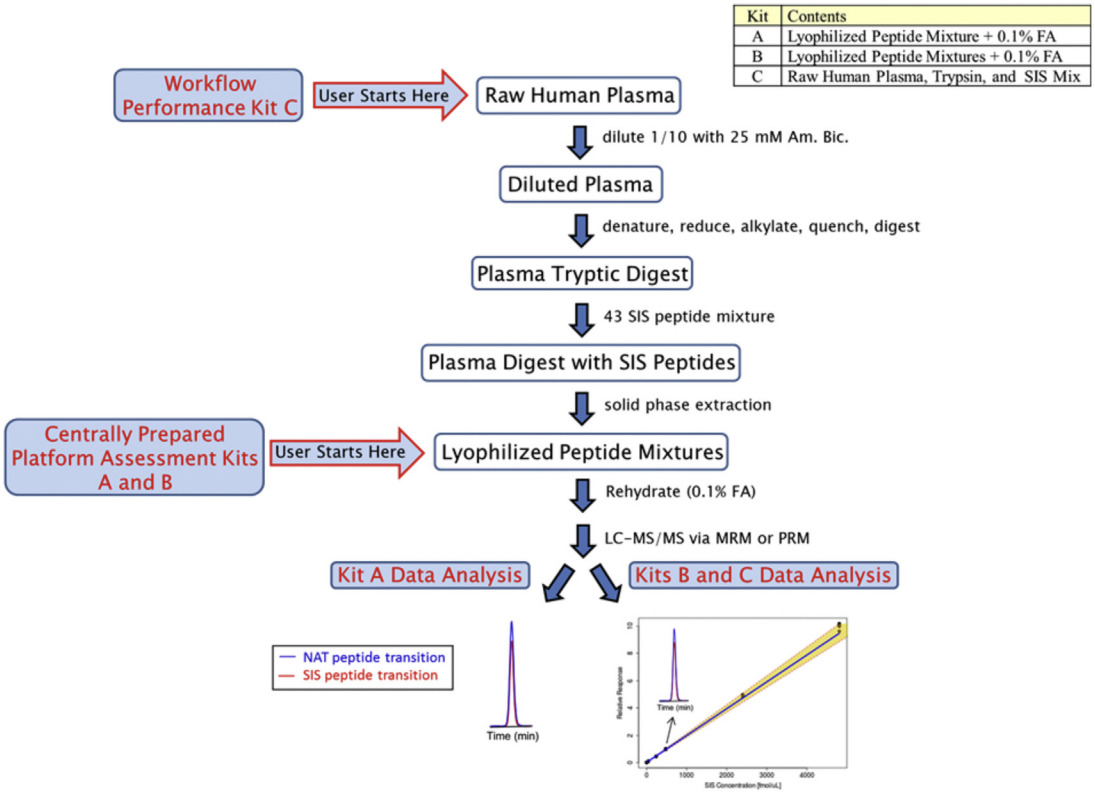

Fig. 1. Experimental workflows and starting points for the 3 QC kits. Kits A and B required simple rehydration of the lyophilized peptide mixture(s) prior to sample processing by LC/MRM-MS or LC/PRM-MS, whereas Kit C required the user to execute the entire workflow from 3 supplied starting materials (namely plasma, trypsin, and the SIS mix). Kit dispersal and data/quantitative analysis were performed at the UVic-Genome BC Proteomics Centre. Quantitation was facilitated by Qualis-SIS which generated standard curves (relative response vs. SIS concentration) from the SIS (red trace) and NAT (blue trace) response data of each peptide's quantifier transition. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

optimal results were obtained. These analyses were performed in different laboratories and on different instruments as part of an international multi-laboratory study which involved 15 participating laboratories and 19 total LC–MS/MS platforms from 4 vendors. Although this study is similar to a recently published multi-site study [16] in that both involved multiple sites and instruments from several vendors, there are significant differences. The previous multi-site study used immunodepleted plasma and a variety of different digestion procedures, with Lys-C/trypsin being ultimately recommended. In the study reported here, no affinity depletion was used. Thus, this current study involved a much more complex matrix. In addition, only trypsin was used for the digestion of the raw, undepleted, plasma, as Lys-C was deemed to be too costly for large-scale quantitative proteomics projects. Importantly, a major focus of our study was to determine the sources of variability, so that future action could be taken to reduce these errors.

2. Methods

The methods used are briefly described here with additional details being provided in Supporting information—Methods.

2.1. Materials, standards, and kit preparation

Forty-three proteotypic tryptic peptides (corresponding to 43 plasma proteins) were originally selected using bioinformatics. Synthesis of C-terminal [^{13}C] and/or [^{15}N]-labeled analogues of these proteotypic peptides was performed at the University of Victoria (UVic)-Genome British Columbia Proteomics Centre using Fmoc protection chemistry [17] on a Prelude or an Overture Robotic Peptide Synthesizer (Peptide Technologies; Seattle, WA, USA). After synthesis, the SIS peptides were purified by reversed-phase HPLC using an Agilent 1260 Infinity LC, with the peptides' identities being subsequently verified by MALDI-TOF-MS on an Ultraflex III mass spectrometer (Bruker Daltonics; Bremen, Germany). Characterization was done by amino-acid analysis, for composition determination, and capillary zone electrophoresis, for purity. The average purity of these SIS peptides was found to be 95%.

Human plasma was obtained from Bioreclamation (catalog no. HMPLEDTA2; Westbury, NY, USA). This biofluid represented a pooled sample collected from a group of healthy, race- and gender-matched donors between the ages of 18 and 50. The plasma samples were stored at −20 °C until use, in order to reduce degradation.

"Bond-breaker TCEP solution" (0.5 M tris(2-carboxyethyl) phosphine) was purchased from Thermo Scientific (Rockford, IL,

Table 1
List of targeted plasma proteins in the QC kit and their reference concentrations, as determined by LC/MRM-MS with SIS peptides.

No.	Protein	UniProt acc. No.	Peptide	Reference plasma protein conc. (μg/mL)
1	Afamin	P43652	DADPDTFFAK	30.71
2	Albumin_serum	P02768	LVNEVTEFAK	34966.77
3	Alpha-1-acid glycoprotein 1	P02763	NWGLSVYADKPETTK	29.13
4	Alpha-1-antichymotrypsin	P01011	EIGELYLPK	151.13
5	Alpha-1B-glycoprotein	P04217	LETPDFQLFK	262.36
6	Alpha-2-antiplasmin	P08697	LGNQEPGGQTALK	3.18
7	Alpha-2-macroglobulin	P01023	LLIYAVLPTGDVIGDSAK	96.42
8	Angiotensinogen	P01019	ALQDQLVLVAAK	15.58
9	Antithrombin-III	P01008	DDLYVSDAFHK	20.67
10	Apolipoprotein A-I	P02647	ATEHLSTLSEK	981.56
11	Apolipoprotein A-II	P02652	SPELQAEAK	33.52
12	Apolipoprotein A-IV	P06727	SLAPYAQDTQEK	21.59
13	Apolipoprotein B-100	P04114	FPEVDVLTK	186.45
14	Apolipoprotein C-I	P02654	TPDVSSALDK	0.54
15	Apolipoprotein C-III	P02656	GWVTDGFSSLK	1.81
16	Apolipoprotein E	P02649	LGPLVEQGR	22.51
17	Beta-2-glycoprotein I	P02749	ATVVYQGER	87.89
18	Ceruloplasmin	P00450	EYTDASFTNR	42.72
19	Clusterin	P10909	ELDESLQVAER	204.69
20	Coagulation factor XII a light chain	P00748	VVGGLVALR	9.58
21	Complement C3	P01024	TGLQEVEVK	298.48
22	Complement C4 gamma chain	P0C0L5	ITQVLHFTK	93.14
23	Complement C9	P02748	TEHYEEQIEAFK	6.86
24	Complement factor B	P00751	EELLPAQDIK	9.84
25	Complement factor H	P08603	SPDVINGSPISQK	152.82
26	Fibrinogen alpha chain	P02671	GSESGIFTNTK	664.56
27	Fibrinogen beta chain	P02675	QGFGNVATNTDGK	177.41
28	Fibrinogen gamma chain	P02679	DTVQIHDITGK	2.48
29	Gelsolin	P06396	TGAQELLR	7.20
30	Haptoglobin	P00738	VGYVSGWGR	1222.73
31	Hemopexin	P02790	NFPSPVDAAFR	422.30
32	Heparin cofactor II	P05546	TLEAQLTPR	55.23
33	Inter-alpha-trypsin inhibitor heavy chain H1	P19827	AAISGENAGLVR	36.72
34	Kininogen-1	P01042	TVGSDTFYSFK	55.33
35	L-selectin	P14151	AEIEYLEK	1.16
36	Retinol-binding protein 4	P02753	YWGVASFLQK	19.23
37	Plasminogen	P00747	LFLEPTR	4.89
38	Prothrombin	P00734	ETAASLLQAGYK	2.40
39	Serum amyloid P-component	P02743	VGEYSLYIGR	0.41
40	Serotransferrin	P02787	EDPQTFYYAVAVVK	180.37
41	Transthyretin	P02766	AADDTWEPFASGK	70.53
42	Vitamin D-binding protein	P02774	THLPEVFLSK	105.72
43	Vitronectin	P04004	FEDGVLDPDYPR	35.21

USA) and TPCK-treated trypsin was obtained from Worthington (Lakewood, NJ, USA). All remaining chemicals (*e.g.*, ammonium bicarbonate, dithiothreitol) and solvents (*e.g.*, acetonitrile, methanol) were of analytical reagent or LC/MS grade, and were purchased from Sigma–Aldrich (St. Louis, MO, USA).

Three kits were assembled at the UVic-Genome BC Proteomics Center and distributed at once to the 15 participating laboratories (see Fig. 1 for a workflow schematic). Kits A and B were prepared at the Centre by the addition of a mixture of 43 SIS peptides (from 43 human plasma proteins; see Table 1 for the target panel) to a plasma tryptic digest, followed by solid phase extraction (10 mg Oasis HLB sorbent; Waters; Milford, MA, USA) and lyophilization of the eluate. The use of Kit A required rehydration of a pre-digested plasma standard in 0.1% formic acid. The SIS peptide mixture used in this kit had previously been concentration-balanced to their endogenous peptide concentrations found in a pooled, de-identified plasma sample. Kit B is similar to Kit A, but required the addition of varying amounts of the SIS-peptide mixture to a constant amount of digested plasma, in order to generate 7-point standard curves spanning a 10,000-fold range (with standards labeled A–G from lowest to highest concentration). The analytical workflow Kit C contains three starting materials (raw plasma, trypsin, and the balanced SIS mix) for the preparation of the same 7 standard samples that were generated in Kit B.

The reference concentration range for the plasma proteins in the mixture was from 35 mg/mL (for human serum albumin) to 410 ng/mL (for serum amyloid P-component; see Table 1), with a range of protein molecular weights from 7 kDa (apolipoprotein C-I, P02654) to 513 kDa (apolipoprotein B-100, P04114) [18]. These kits were designed to be used with standard-flow and nano-flow LC/MS/MS platforms, operated in the MRM or PRM modes. Before distribution of the kits to the participants, the long-term storage properties and lot-to-lot kit variabilities were determined (see Supporting information—Fig. 1).

2.2. Study design

The kits were prepared at the Centre and were distributed simultaneously to the 15 participating laboratories for processing (see Table 2 for details). The 15 participants in this study were from 7 different countries, and used 17 LC–MS/MS platforms with 7 different mass spectrometers obtained from 4 different manufacturers. Including the 2 reference LC–MS/MS platforms (labeled R1a and R1b), 68% of the platforms were operated at nano-flow rates (usually at 300 nL/min), while 32% were operated at standard-flow rates (primarily at 400 μL/min).

The instructions given to the participants were to analyze the 2 platform assessment kits (Kits A and B) and the workflow kit (Kit C) by LC/MRM-MS or LC/PRM-MS. Each site performed interference testing on the complete panel of 43 peptides, or used a smaller panel of peptides as specified for their platform in the SOP, based on prior testing. Those peptides that passed the interference testing were targeted by MRM or PRM. The analysis order for the standard samples was specified in the SOP as follows: Kit A – Kit B – Kit A, then Kit A – Kit C – Kit A. Extracted ion chromatograms (XICs) obtained for the target quantifier peptides were compared across the 4 measurements in Kit A and were used to generate

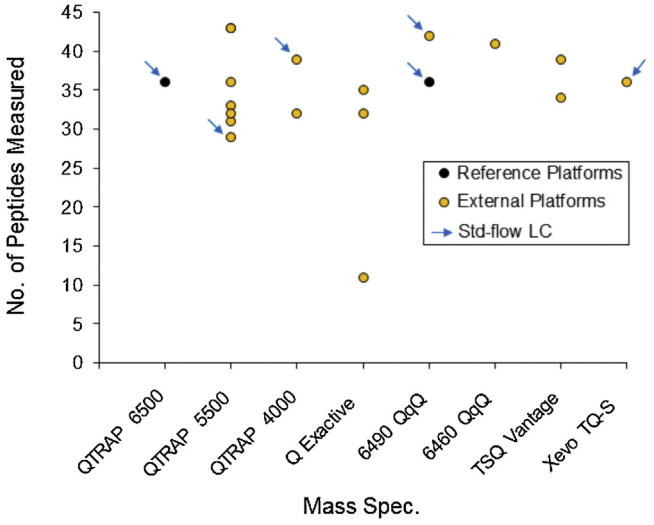

Fig. 2. Number of peptides detected in Kit A from the 4 measurements. Plotted are the results from all sites (including the 2 reference platforms) for the 8 different MS instruments utilized, with arrows to differentiate the standard-flow LC systems from the nano-flow. Please see Supplemental Fig. 2 for a plot of the differences between the peptides monitored and the peptides detected.

Table 2
Participants and instruments involved in the inter-laboratory standardization study of the platform/workflow QC assessment kits. The study included 19 LC–MS/MS platforms in 16 laboratories across 7 countries. General LC details were 0.2–0.4 μL/min for nano-flow and 0.2–0.4 mL/min for standard-flow, with an average of 3 or 1.8 μm particles packed onto columns of 50–100 μm × 100–150 mm for nano-flow and 2.1 × 100–150 mm for standard-flow. The reference instruments are labeled as site numbers R1a and R1b. Complete platform details are provided in Supporting information—Methods.

Site No.	Lab location	General LC conditions			Mass spectrometer
		Flow type	Chromatography	Gradient	
1	Country 1	Nano	HPLC	60	QTRAP 5500
2	Country 1	Nano	UHPLC	90	QTRAP 5500
3	Country 1	Nano	HPLC	90	QTRAP 5500
4	Country 1	Nano	HPLC	35	QTRAP 5500
5	Country 1	Nano	HPLC	90	QTRAP 5500
6	Country 2	Standard	UHPLC	30	QTRAP 5500
7	Country 3	Nano	UHPLC	60	QTRAP 5500
8	Country 1	Nano	HPLC	60	QTRAP 4000
9	Country 4	Standard	HPLC	35	QTRAP 4000
10	Country 1	Nano	UHPLC	71	Q Exactive
11	Country 5	Nano	HPLC	51	Q Exactive
12a	Country 6	Nano	UHPLC	75	Q Exactive
12b	Country 6	Standard	UHPLC	30	6490
13	Country 4	Nano	HPLC	60	6460
14	Country 3	Nano	UHPLC	80	TSQ Vantage
15a	Country 7	Nano	HPLC	60	TSQ Vantage
15b	Country 7	Standard	UHPLC	35	Xevo TQ-S
R1a	Canada	Standard	UHPLC	30	6490
R1b	Canada	Standard	UHPLC	30	QTRAP 6500

peptide standard curves for quantitation in Kits B and C. The peak verification was done at each site, but the quantitation (for Kits B and C) and data analysis (for all kits) was done at the UVic-Genome BC Proteomics Centre using Qualis-SIS software [19] and Microsoft Excel.

3. Results and discussion

3.1. Peptide detection and metric comparison using Kit A

The purpose of Kit A is to help isolate LC–MS issues through the monitoring of performance metrics over replicate injections. Here, a total of 4 measurements were made (once before and once after the use of Kits B and C), with the results used to evaluate system performance. Although it seems intuitive that the use of a platform with increased mass and chromatographic resolution would be less subject to interferences than one with lower resolution (of both types), as can be seen from Fig. 2, the number of reproducibly detected peptides was not correlated with the use of a nano-flow or standard-flow LC–MS/MS platform (see Supporting information—Fig. 2 for the site differences between the peptides monitored and detected). That this was *not* the case in this study speaks to the accuracy of the interference testing – achieved through ion ratio comparisons for each peptide's top 3 precursor-product ion pairs in buffer and matrix – that was done on these particular proteotypic peptide standards, making them suitable for use on a variety of instruments and platforms. To more closely examine this rather unexpected, but very promising, finding, two key performance metrics – retention time (RT, for LC evaluation) and relative response (RR, for MS assessment) – were explored.

To best assess the performance of the LC systems, the RT variability for a matching set of peptides was plotted for the 19 LC–MS/MS platforms (includes the reference sites R1a and R1b) according to the type of chromatography (HPLC *vs* UHPLC) and the flow rate (nano-flow *vs* standard-flow; see Fig. 3). Although lower variability was obtained with the standard-flow platforms (average %CVs were 0.20 with standard-flow UHPLC and 0.61 with standard-flow HPLC *vs* 2.06 with nanoUHPLC and 2.16 with nanoHPLC), sites 4 and 14 demonstrated that well-functioning nanoLC–MS/MS platforms are also capable of providing comparably low variabilities (average %CVs were 0.23 for site 4 and 0.28 for site 14). In fact, the RT variability for all reproducibly detected peptides at site 4 (36 peptides in total) and site 14 (39 peptides in total) was 0.29 and 0.47 %CV, respectively. The outliers in the dataset indicate a potential LC issue (*e.g.*, a system leak, restricted flow) that should be corrected before any "real" analytical measurements are made. To assist in this troubleshooting, column pressure and flow rate traces should be recorded as part of the acquisition parameters since fluctuations can reveal operation errors (*e.g.*, pressure dip – air bubble; improperly pulsing solvent delivery system – bad check valve). Overall, although there is a clear variation in both the distribution and the values of the RT %CVs, the average %CV was 1.6% across all 19 LC–MS/MS platforms.

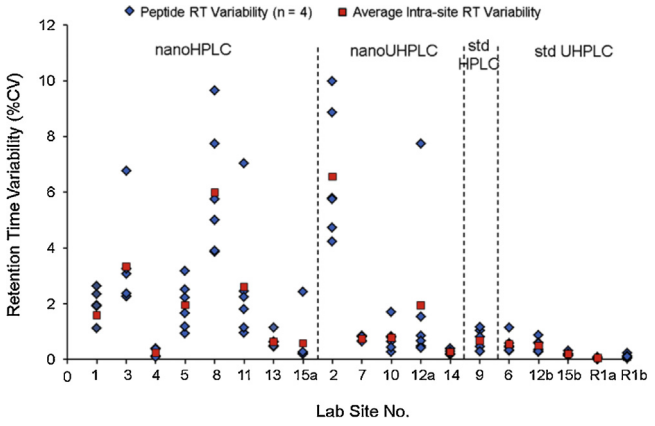

Fig. 3. Global RT variability comparisons of Kit A measured in quadruplicate across the 19 LC–MS/MS platforms. The sites are sorted according to chromatography and flow rate for the same panel of reproducibly detected peptides. Additional LC–MS/MS details and the origination sites (*e.g.*, R1a and R1b denote the reference instruments) are provided in Table 2 and Supporting information—Methods.

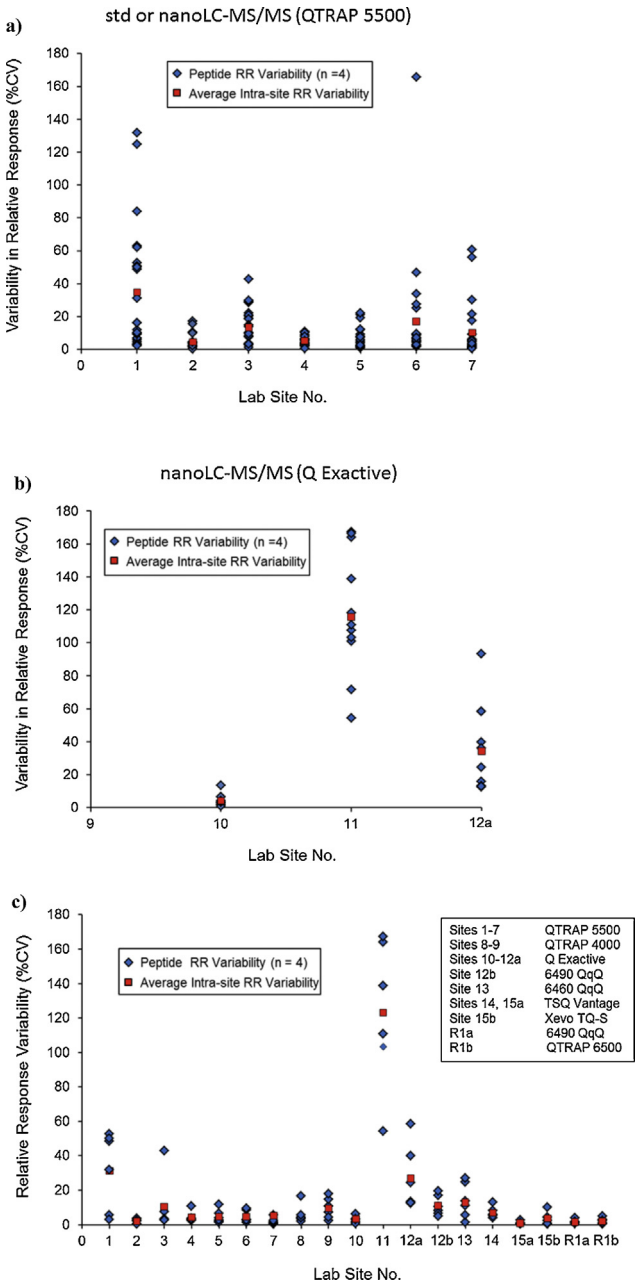

Fig. 4. Comparisons of RR variability for Kit A measured in quadruplicate across the 19 LC–MS/MS platforms. Variability comparisons for 7 laboratories using the QTRAP 5500 in a) and 3 laboratories using the Q-Exactive in b). c) Global variability comparison across all measured platforms. Data from an independent set of matching peptides is shown in each part, with each peptide's RR variability shown as blue diamonds and the average intra-site variability as red squares. As defined in Table 2, the reference instruments are labeled with site numbers R1a and R1b. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To assess variability at the MS level, the average %CV in RR (*i.e.*, the NAT/SIS peak area ratio, where NAT is defined as the endogenous peptide signal) for the quadruplicate measurements was compared on a peptide-by-peptide basis. Here, data collected from the same instrument was first compared (Fig. 4a for the QTRAP 5500s, Fig. 4b for the Q Exactives) before comparing the global variability across all sites (Fig. 4**c**). Regarding the local comparisons, Fig. 4a and b show unexpectedly high variability (>20% CV) for certain sites that was independent of the instrument employed (*e.g.*, site 1 *vs* site 4 in Fig. 4a; site 11 *vs* site 10 in Fig. 4b). This result was also observed in the global Fig. 4c plot when examining the range and average RR variability with a smaller set of matching peptides. Since the outliers had relatively stable RTs (the exception being site 11), the cause for the variability is purely MS-based (most likely due to interference with the transitions). Regarding site 11, a serious technical issue occurred since the RTs and RRs were both highly variable, with average %CVs of 8.6% and 118.3%, respectively, for their 11 reproducibly detected peptides. Considering all sites, however, although there was variation in both the distribution and the value of the RR %CVs, the average %CV was 14.2%.

3.2. Quantitative performance evaluation with Kit B

Since quantitation was done using peptide standard curves under strict qualification criteria (which involve, for example, precision and accuracy thresholds for the 5 replicates within a given concentration level), the number of proteins quantified was expected to be lower than would otherwise be expected based on the relative response data alone. The observed reduction between the number of detected and quantified peptides in Kits A and Kit B was an average of 8. Fig. 5 shows the results from Kit B, with the number of proteins quantified plotted as a function of instrument used. Although there are too few instruments involved for a rigorous statistical comparison, it appears that a major component of the variability is within-instrument rather than between instrument types. For instance, low sensitivity seems to be the cause of the reduced quantitation of the apparent outliers at select sites that utilized the QTRAP 4000 (site 9), Q Exactive (site 12a), and TSQ Vantage (site 14) instruments. This is attributed to the fact that the average dynamic range of the platforms utilizing those instruments at those sites was only 1–2 orders of magnitude, compared to the expected 3 orders of magnitude observed for the reference sites as well as for site 8 (QTRAP 4000), site 11 (Q Exactive), and site 15a (TSQ Vantage) using a matching set of quantified proteins. RR variability with these outlier sites is an additional issue that restricted the number of proteins quantified through regression analysis of standard curves. For instance, 25 peptides from the site 9 data had RR CVs >>20% at the top 3 levels of the curve (which reflect standards E, F, and G). Since these 3 upper levels must pass our precision and accuracy criteria in our low-to-high concentration removal strategy (defined initially in [20]), these peptides were automatically excluded by Qualis-SIS [19]. This provides an explanation for the large difference of 26 observed between the quantified and detected peptides in the site 9 Kit B analysis.

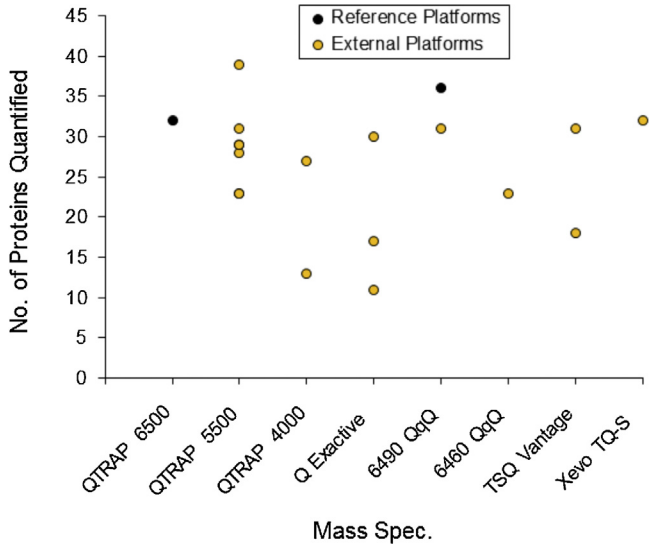

Fig. 5. Number of proteins quantified in Kit B by LC/MRM-MS or LC/PRM-MS as a function of instrument employed. The reference and external platforms are colour coded for comparison. The quantified values were obtained from Qualis-SIS using global, default parameters (*e.g.*, $1/x^2$ regression weighting, low-to-high concentration removal strategy, <20% deviation in a given levels precision and accuracy for the 5 technical replicates).

To further interpret the Fig. 5 results, the plasma protein concentrations were evaluated by examining the distribution of % CVs for the two instruments that were used by at least 3 sites (the QTRAP 5500 and the Q Exactive), as shown in Fig. 6a. From this figure, it is clear that the Q Exactive produced a greater proportion of peptides with a higher %CV than the QTRAP 5500, with the results from the QTRAP having a tighter distribution. For instance, the average CV for alpha-1-antichymotrypsin (P01011) was 21% for the LC–MS/MS platforms utilizing the Q Exactive, but 9% on the QTRAP 5500. Its average concentration, however, was comparable to the reference values with 148 μg/mL and 134 μg/mL obtained for the QTRAP 5500s and Q Exactives, respectively, compared to 151 μg/mL for the reference sites (R1a and R1b).

Closer examination of the distribution involved a site-by-site comparison to the reference concentrations for accuracy assessment. This revealed cases of both excellent and poor agreement. Fig. 6b shows a representative example of good correlation with a matching set of 28 plasma proteins. In fact, the excellent correlation with the site 5 Kit B results extended beyond the concentrations—the average coefficient of determination (R^2) was 0.990 *vs* 0.992 and the average RR variability for standard E was 3.8% *vs* 2.1% CV (for sites 5 and R1a, respectively, in both cases). The exception was in the RT variability (0.55% *vs* 0.08% CV from site R1a). The deviation in RT variability from the reference, however, is not unreasonable nor unexpected for site 5, based on the Kit A results from their nano HPLC/MRM-MS platform relative to the reference R1a values obtained on a standard-flow UHPLC/MRM-MS platform (Fig. 3).

A contrasting example is shown in Fig. 6c for site 12a (measured by nanoLC/PRM-MS on a Q Exactive). Here, the derived concentrations for 11 matching proteins showed poor agreement with the reference site due to a consistently low linear dynamic range—an average of one order of magnitude for these 11 proteins instead of the expected 3 orders of magnitude. For instance, the quantifier peptide for apolipoprotein E had a dynamic range of 10 at the reference site, compared to a dynamic range of 1000 at site 12a, despite comparable R^2 values (0.992 from site 12a *vs* 0.994 from the reference sites R1a and R1b). The cause is attributed to sub-optimal performance of the mass spectrometer, since 8 peptides could not be observed with standard E (which should have resulted in balanced RRs of approximately 1:1) and 11 peptides had high RR variability (>20% CV) at the 3 highest concentration levels, thus preventing the generation of standard curves. These problems were not observed with the other Q Exactives (*e.g.*, site 10, where the average dynamic range for the same peptides was 3 orders of magnitude). This confirms a problem in the performance of this particular instrument at site 12a.

To assess the accuracy on a global scale, the proteins quantified on all 19 LC–MS/MS platforms were compared (see Supporting

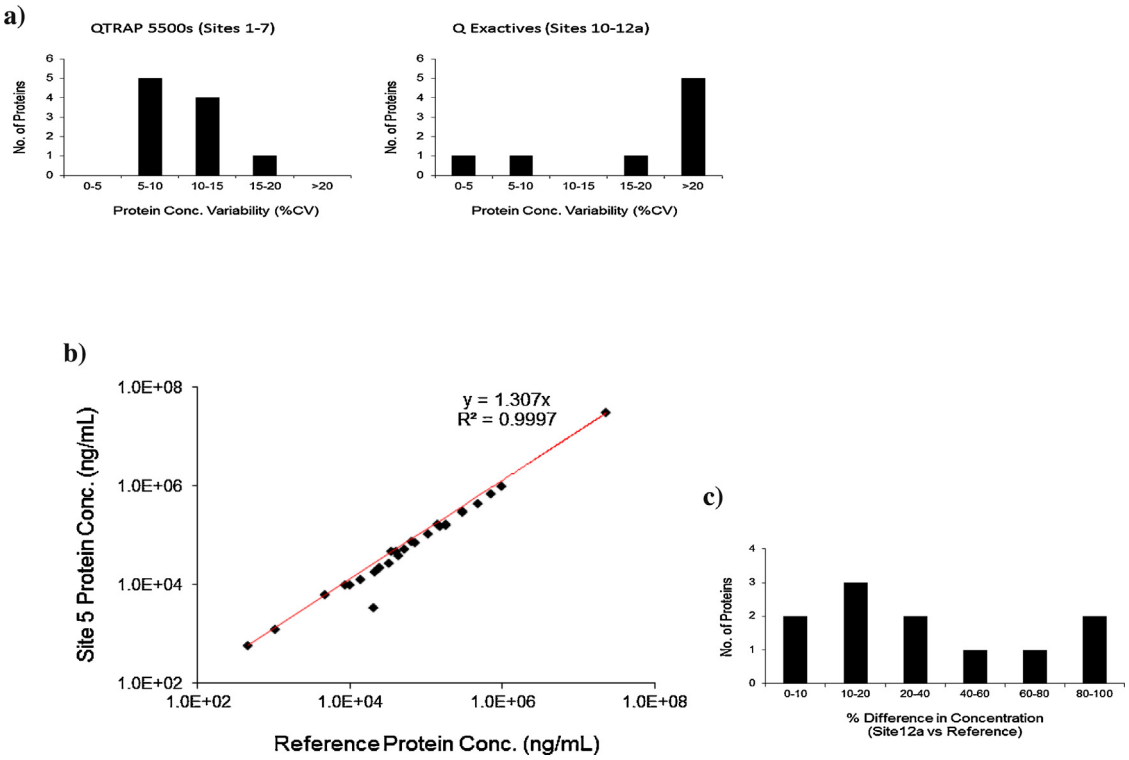

Fig. 6. Quantitative result comparisons from the LC/MRM-MS or LC/PRM-MS analysis of Kit B. a) Distribution of measured plasma protein concentration CVs for the QTRAP 5500 and the Q Exactive instruments against a common set of instrument-specific quantified proteins. To illustrate site-specific extremes, part b) shows an example of results from site 5 that correlated well to the reference values, whereas part c) shows poor agreement of site 12a's results with the references.

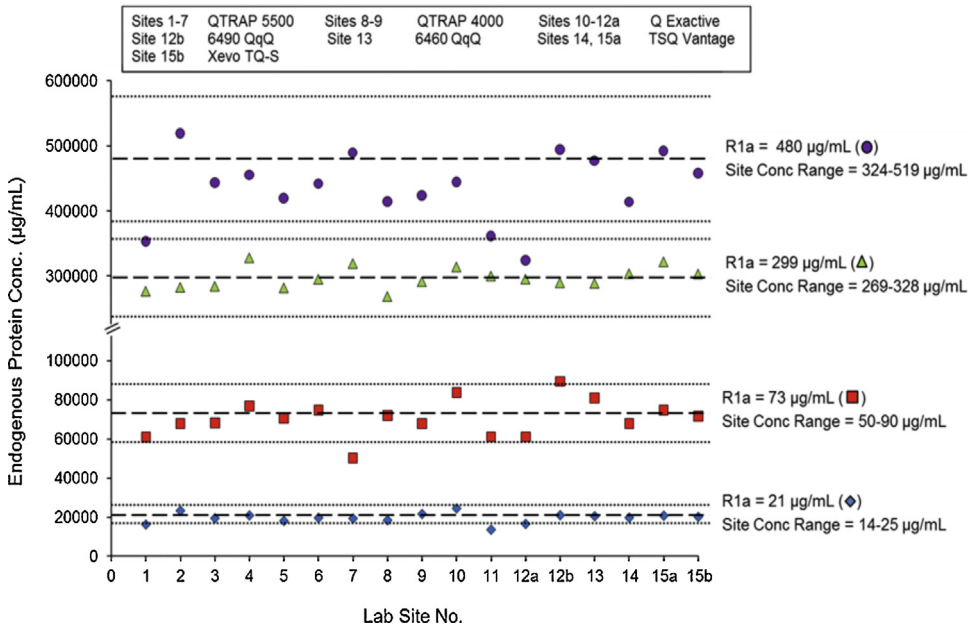

Fig. 7. Global quantitative comparisons from the LC–MS/MS measurement of Kit B. Accuracy assessment of 4 moderate-abundance plasma proteins (concentration range: 21–480 µg/mL) against their corresponding reference values (noted with dotted lines, ±20% accuracy limits shown as dotted plus signs), as determined by reference site R1a. The quantitation values correspond to hemopexin (purple circles), alpha-1B-glycoprotein (green triangles), transthyretin (red squares), and plasma retinol-binding protein (blue diamonds). Site details are provided in Table 2 and Supporting information—Methods. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

information Fig. 3a for a graphical overview). The quantitative comparison of 4 moderate abundance proteins (reference concentration range: 21–480 µg/mL) is shown in Fig. 7. As can be seen from this figure, the site-determined concentrations were typically within the acceptable accuracy limits of the reference values. Two exceptions were hemopexin from site 12a and transthyretin from site 7. In the former case, for example, poor linearity of the standard curve (y-intercept: −0.179, R^2: 0.974) contributed to this problem,

despite acceptable precision in RR and RT curve CVs. Overall, the correlation with the reference values was not dependent on the protein abundance level. For example, the high-abundance proteins haptoglobin (average determined concentration: 1.3 mg/mL) and apolipoprotein A–I (average determined concentration: 0.8 mg/mL) returned average inter-site CVs of 9.5% and 14.2%, respectively. This variability compared well to that of lower abundance proteins in the dataset, such as L-selectin (average determined concentration: 1.1 μg/mL) and plasminogen (average determined concentration: 5.1 μg/mL), which gave average inter-site CVs of 13.6% and 10.7% CV, respectively.

In further assessing the quantitative accuracy with Kit B, after excluding 3 outliers (as determined by Grubbs test at the 97.5% confidence interval [21]), the average inter-site variability in concentration across the 35 quantified proteins measured by at least 3 LC–MS/MS platforms was found to be 16.5% CV. Compliment C3, we note, had the lowest variability at 4.9% CV measured across 15 platforms. Its determined concentration of 293 μg/mL correlates well with our previous determinations in other projects using similar plasma lots and sample preparation protocols [22,23]. In fact, the measured concentrations for the reference are within 1.5-fold of those found previously, which strengthens the accuracy assessments. Overall, the strong correlation of the inter-site results was not unexpected since Kit B required only rehydration before sample processing, which limited the potential sources of variability. These results highlight the benefit of an optimally performing LC–MS/MS platform for maximum quantitative output and optimal quantitative accuracy. It also reinforced the utility of this kit for assessing LC–MS/MS platform performance and revealing the cause(s) of atypical variation.

3.3. Detection of procedural or instrumental errors with Kit C

As expected, sample preparation at the individual sites with Kit C resulted in larger variability in the number of proteins quantified than the results obtained by simply processing the centrally prepared standards in Kit B (Fig. 8a *vs.* Fig. 5, also Supporting information Fig. 3a *vs* 3b). Excluding the reference measurements, the total number of proteins quantified was reduced by a minimum of 1 (with sites 2, 10, and 15b being at the low end), while peptide detection using Kit A and quantitation in Kit C differed by a minimum of 4 (with sites 2 and 15b being at the low end).

The most evident outliers in Fig. 8 were the ones that failed to quantify a single protein with Kit C. In the case of the one circled in red, for example, the cause for this anomaly was due to both technician error and other errors (Fig. 8b). Regarding the former, the user seems to have inadvertently switched the labels of standards F and G prior to sample processing and analysis. This resulted in the representative plots shown in the 'Before Revision' panel of Fig. 8b. Subsequent correction of the user-supplied peptide input file prior to automated quantitation in Qualis-SIS enabled the quantitation of 14 plasma proteins (see the standard curve in the 'After Revision' panel of Fig. 8b for a representative example). While deviation from the standardized protocol at the trypsin solubilization step seems to have led to a problem with tryptic digestion efficiency (average NAT response: 5.5E + 05 for Kit C vs 6.6E + 06 for Kit B for a matching set of peptides), the number

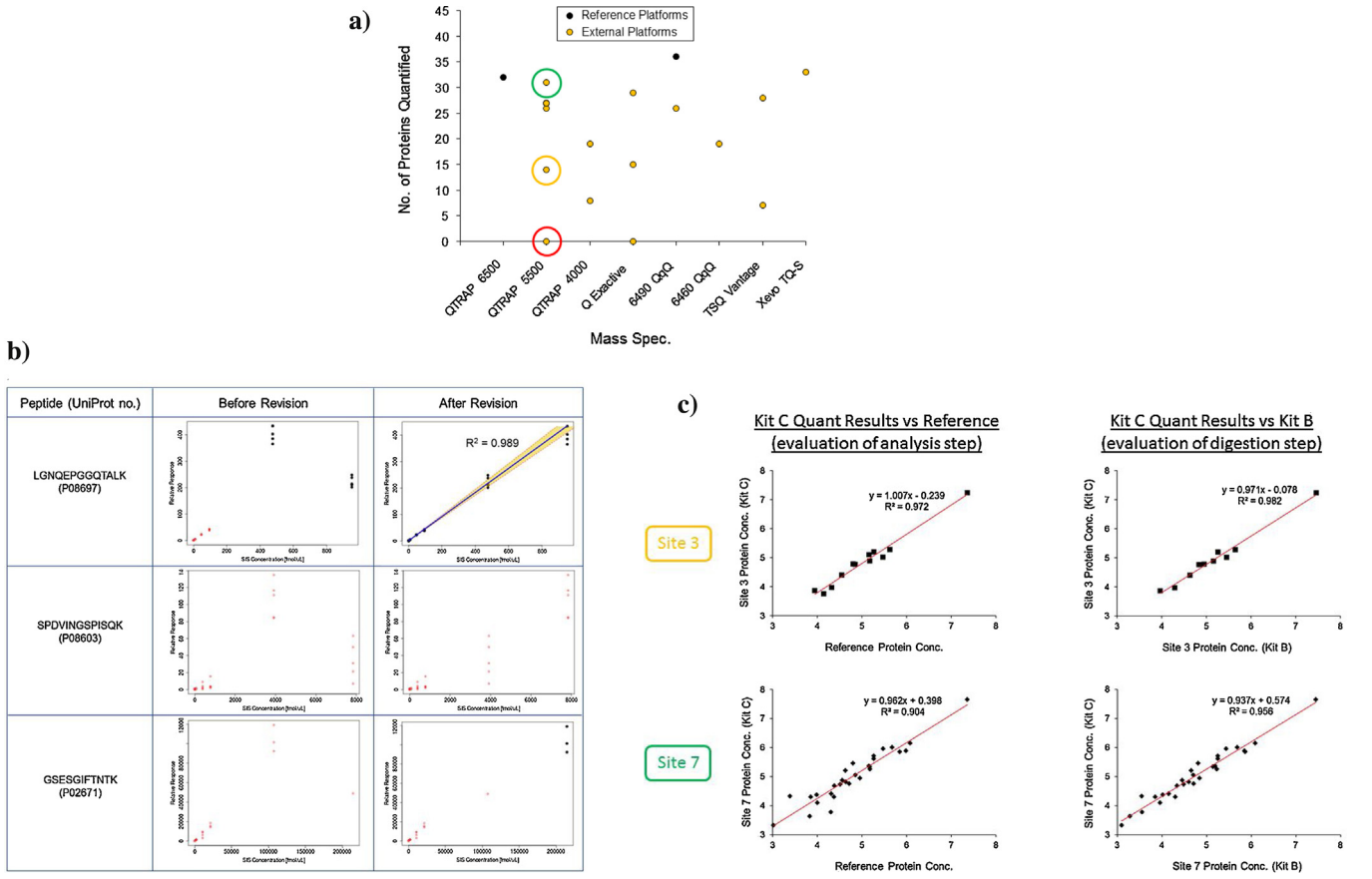

Fig. 8. Quantitative results from the QC workflow performance Kit C. a) Total number of proteins quantified on the 19 LC–MS/MS platforms measured across the 16 sites (external and internal). The colored circles represent the results from 3 different sites. b) Example of technician and other errors in the "red site's" results. c) Comparison of the quantitation results for Kit C from the yellow and green sites with a matching set of reference values, as well as each site's Kit C results with their values measured with Kit B. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of detectable peptides in Kit C relative to Kit B was the same at 29. The number of detectable peptides in Kit C relative to Kit B was the same at 29, yielding an average NAT response of 5.5E+05 *vs* 6.6E+06, respectively. Evidence of sub-optimal quantitative performance was revealed by poor precision (>20% CV; see the middle row of Fig. 8b) and the absence of complete technical replicate data (see the bottom row of Fig. 8b) in at least 1 of the 3 highest concentration levels of Kit C's standards. This caused their immediate disqualification, due to the violation of 3 qualification criteria (namely, the inclusion of all 5 replicates per level, a 20% CV threshold in a given level's precision, and the presence of 3 consecutive concentration levels). Upon closer examination of the calibration curve data, the RT variability was low with the user-prepared standards in Kit C (average CVs: 0.31% from standard E) and correlated well with the variability obtained with Kit A (average CVs: 0.82%, both for 26 matched peptides). This site also experienced a larger-than-expected RR variability of 18.6% CV, on average, from the Kit A analysis, which would also alert the user to an instrument problem. To place this high variability into perspective, site 2 and site 4 (both with nanoLC-MS/MS on a QTRAP 5500) had average CVs in RR of <6%, as expected, for a matching set of detectable peptides from Kit A.

Sub-optimal mass spectrometer performance was also the cause for poor quantitation at some other sites in Fig. 8 (*e.g.*, site 14, nanoUHPLC/MRM-MS on the TSQ Vantage), while procedural errors were the apparent cause of problems at site 9. In the latter case, an error was made in the preparation of the SIS peptide dilution series, as inferred from the incorrect SIS peptide ratios obtained for the 3 highest concentration levels of the standard curve (average ratios from standard G for Kit C *vs* Kit B: 1:5:6 *vs* the expected 1:2:5). This explains the narrow dynamic range observed with Kit C, as compared to the Kit B analysis (average dynamic range: 20 *vs* 885) at the same site. In contrast, when the procedures were carefully followed and the LC–MS/MS platform was performing optimally, accurate and reproducible quantitation was achieved (sites 3 and 7 in Fig. 8**c**). This good correlation extended beyond the concentrations. For example, the average curve R^2 values (from the relative response *vs* SIS peptide concentration plots) for the proteins shown in Fig. 8c were 0.967 and 0.993 for sites 3 and 7, respectively, compared to 0.994 from the references. Evidence of their well-functioning LC/MRM-MS platforms also came from the low RT and RR variabilities—the averages for these two sites were 3.9% and 2.1% CV for RR, and 0.61% and 0.15% CV for RT, both derived from standard E replicate data. This compared well to their Kit B results, which showed an average variability of 4.3% and 2.1% CV in RR, with 0.99% CV and 0.13% CV being obtained for RT (again with standard E replicate curve data). Since the kits were processed several weeks apart, the low degree of variability obtained with the site 3 and 7 data demonstrates the stability of their LC/MRM-MS platforms over that time period.

Also instructive is the global quantitative accuracy assessment of the Kit C measurements (see Supporting information Fig. 2b for a graphical overview and Fig. 9 for a quantitative comparison). As illustrated in Fig. 9, the majority of the determined concentrations for apolipoprotein E were within the accuracy limits of quantitation, the exception being site 5. This outlier site also generated concentrations that lay outside the accuracy limits of heparin cofactor II. Since their derived concentrations were below the lower accuracy limit for these proteins, as well as others not depicted in this figure (*e.g.*, 12 μg/mL determined for inter-alpha-trypsin inhibitor heavy chain H1 *vs* 36–44 μg/mL for the references with an accuracy range of 29–52 μg/mL), the cause is attributed to inefficient proteolytic digestion. This explanation can also be applied to other outlier sites in Fig. 9 that demonstrated acceptable assay attributes and performance metrics (*e.g.*, sites 8 and 15a). Nonetheless, the average inter-site variabilities for heparin cofactor II and apolipoprotein E were found to be 16.1% and 17.9% CV, respectively. These variabilities compared well to those obtained with Kit B (average inter-site CVs of 11.7% for heparin cofactor II and 19.6% for apolipoprotein E), with similar average concentrations (45.3 and 21.6 μg/mL for heparin cofactor II and apolipoprotein E, respectively, with Kit C vs. 49.2 and 23.9 μg/mL for these 2 proteins with Kit B).

Overall, this study reinforced the value of these kits for routine and comprehensive performance assessment. This study also emphasized the importance of carrying out this performance assessment in order to reveal the absence or presence of a procedural or instrumental error that would need to be corrected prior to performing a set of "real" quantitation experiments

4. Conclusions and future considerations

Standardization of proteomic methods and analytical platforms is essential for precise and accurate protein quantitation within and between laboratories. To this end, we have developed three standardization kits (A and B for platform assessment, C for complete workflow) for routine and comprehensive evaluation of various phases of an MRM-with-SIS-peptide assay. In this study, these three QC assessment kits (prepared at the UVic-Genome BC Proteomics Centre) were evaluated on 17 LC–MS/MS platforms across 15 external laboratories spanning 5 continents. The overall conclusions drawn from this study, which involved the quantitation of 43 plasma proteins in non-depleted human plasma, were that these 3 performance assessment kits can be successfully used

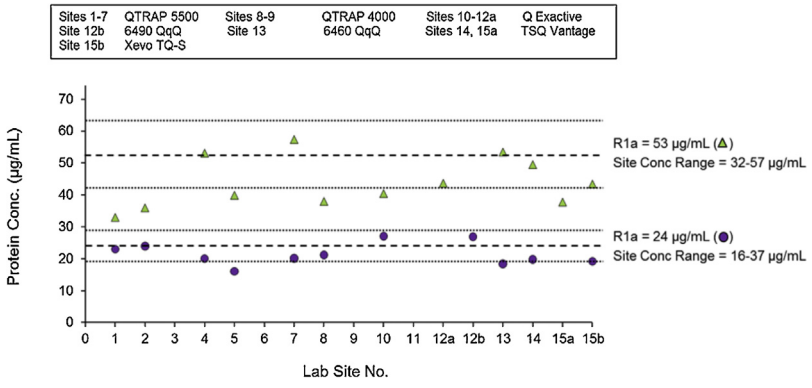

Fig. 9. Global quantitative accuracy comparisons using Kit C, as measured by 17 external LC–MS/MS platforms. Two moderate-abundance plasma proteins. quantified by a minimum of 11 platforms, are shown against their corresponding reference values (noted with dotted lines, ±20% accuracy limits shown as dotted plus signs), as determined by reference site R1a. The quantitation values correspond to heparin cofactor II (green triangles) and apolipoprotein E (purple circles). Site details are provided in Table 2 and Supporting information—Methods. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to evaluate laboratory procedures and platform performance, thereby enabling corrective methods or system maintenance to be undertaken. This study further showed that the accuracy of protein quantitation was independent of the LC–MS/MS platform, when performing optimally, but was dependent on the technical skills of the individual user.

The next step in this global standardization process involves training the users when errors were noted (be it from a sample preparation or LC–MS/MS processing), and then repeating the study to determine whether the quantitative accuracy and reproducibility can be improved. This study also highlighted the value of developing fully automated sample preparation procedures if MS-based assays are to find widespread clinical implementation. An additional consideration involves working with the instrument vendors toward the development of automated software solutions for system suitability. This will help non-expert users rapidly diagnose platform errors, which might arise unexpectedly during routine use. In this type of software, if the specifications for a given standard, such as QC Kit A, are not satisfied over replicate injections (due to either chromatographic properties such as column pressure, flow stability, and peak symmetry, or mass spectrometric performance issues such as electrospray stability, sensitivity, and MS/MS fragmentation), then the batch would be automatically aborted and the cause for this action would be reported in a transmission log. Such automated QA/QC measures are imperative for non-expert use of proteomic methods and for the use LC–MS/MS platforms in the clinic. Implementation of these procedures will also help to improve the quantitative reproducibility and accuracy of large-scale proteomic efforts, such as protein biomarker screening.

Conflict of interest

Christoph Borchers is the director of the UVic-Genome BC Proteomics Centre and the Chief Scientific Officer of MRM Proteomics, which is responsible for commercializing Centre-related technology developments into kits, such as the quality control ones discussed here, for use by the quantitative proteomics community.

Acknowledgements

The UVic-Genome BC Proteomics Centre would like to thank Genome Canada and Genome British Columbia for STIC funding and support, as well as Western Economic Diversification of Canada for instrumentation support. All laboratories from Spain are members of ProteoRed (Plataforma de Recursos Biomoleculares y Bioinformáticos) and are supported by grant PT13/0001 funded by Instituto de Salud Carlos III (ISCIII) and the European Regional Development Fund (FEDER). FC wishes to acknowledge support from Grant PRB2 (IPT13/0001 - ISCIII-SGEFI/FEDER). GD wishes to acknowledge support from grants FAPERJ E26-110.138/2013, FAPERJ E-26/111.697/2013, and FAPERJ E-26/0010.1582/2014. AK and VZ state that the Russian portion of the work was done in the framework of the State Academies Fundamental Research Program (2013-2020). Reynaldo Interior (Hospital for Sick Children; Toronto, ON, Canada), Jay Gambee (AAA Service Laboratory; Damascus, OR, USA), and David Chen (University of British Columbia; Vancouver, BC, Canada) are recognized for characterizing the 43 SIS peptides. PAH wishes to thank Ardeshir Amirkhani of the Australian Proteome Analysis Facility, Macquarie University, Sydney, Australia for assistance with sample preparation and data analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.euprot.2015.06.001.

References

[1] S. Ong, M. Mann, Mass spectrometry-based proteomics turns quantitative, Nat. Chem. Biol. 1 (2005) 252–262.
[2] D.H. Chace, T.A. Kalas, A biochemical perspective on the use of tandem mass spectrometry for newborn screening and clinical testing, Clin. Biochem. 38 (2005) 296–309.
[3] E.J. Want, B.F. Cravatt, G. Siuzdak, The expanding role of mass spectrometry in metabolite profiling and characterization, Chembiochem 6 (2005) 1941–1951.
[4] A.J. Rai, F. Vitzthum, Effects of preanalytical variables on peptide and protein measurements in human serum and plasma: implications for clinical proteomics, Expert Rev. Proteomics 3 (2006) 409–426.
[5] A.J. Rai, C.A. Gelfand, B.C. Haywood, D.J. Warunek, J. Yi, M.D. Schuchard, et al., HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples, Proteomics 5 (2005) 3262–3277.
[6] J. Malm, T.E. Fehniger, P. Danmyr, A. Végvári, C. Welinder, H. Lindberg, et al., Developments in biobanking workflow standardization providing sample integrity and stability, J. Proteomics 95 (2013) 38–45.
[7] G. Marko-Varga, A. Végvári, C. Welinder, H. Lindberg, M. Rezeli, G. Edula, et al., Standardization and utilization of biobank resources in clinical protein science with examples of emerging applications, J. Proteome Res. 11 (2012) 5124–5134.
[8] T.A. Addona, S.E. Abbatiello, B. Schilling, S.J. Skates, D.R. Mani, D.M. Bunk, et al., Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma, Nat. Biotechnol. 27 (2009) 633–641.
[9] N.L. Anderson, N.G. Anderson, T.W. Pearson, C.H. Borchers, A.G. Paulovich, S.D. Patterson, et al., A human proteome detection and quantitation project, Mol. Cell Proteomics 8 (2009) 883–886.
[10] S.E. Abbatiello, D.R. Mani, B. Schilling, B. Maclean, L.J. Zimmerman, X. Feng, et al., Design, implementation, and multi-site evaluation of a system suitability protocol for the quantitative assessment of instrument performance in LC-MRM-MS, Mol. Cell Proteomics 12 (2013) 2623–2639.
[11] A.G. Paulovich, D.D. Billheimer, A.-J.L. Ham, L. Vega-Montoto, P.A. Rudnick, D.L. Tabb, et al., A CPTAC inter-laboratory study characterizing a yeast performance standard for benchmarking LC–MS platform performance, Mol. Cell Proteomics 9 (2010) 242–254.
[12] CPTAC. http://proteomicscancergov/programs/CPTAC/. 2010.
[13] A.J. Percy, A.G. Chambers, J. Yang, A.M. Jackson, D. Domanski, J. Burkhart, et al., Method and platform standardization in MRM-based quantitative plasma proteomics, J. Proteomics 95 (2013) 66–76.
[14] A.J. Percy, A.G. Chambers, D.S. Smith, C.H. Borchers, Standardized protocols for quality control of MRM-based plasma proteomic workflow, J. Proteome Res. 12 (2013) 222–233.
[15] A.C. Peterson, J.D. Russell, D.J. Bailey, M.S. Westphall, J.J. Coon, Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics, Mol. Cell Proteomics 11 (2012) 1475–1488.
[16] S.E. Abbatiello, B. Schilling, D.R. Mani, L.J. Zimmerman, S.C. Hall, B. MacLean, et al., Large-scale inter-laboratory study to develop, analytically validate and apply highly multiplexed, quantitative peptide assays to measure cancer-relevant proteins in plasma, Mol. Cell Proteomics (2015) pii: mcp.M114.047050.
[17] M.A. Kuzyk, D. Smith, J. Yang, T.J. Cross, A.M. Jackson, D.B. Hardie, et al., Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma, Mol. Cell. Proteomics 8 (2009) 1860–1877.
[18] SIB_Swiss_Institute_of_Bioinformatics. Compute pI/Mw. http://web.expasy.org/compute_pi/.
[19] Y. Mohammed, A.J. Percy, A.G. Chambers, C.H. Borchers, Qualis-SIS. Automated standard curve generation and quality assessment for multiplexed targeted quantitative proteomic experiments with labeled standards, J. Proteome Res. 14 (2015) 1137–1146.
[20] A.J. Percy, J. Yang, A.G. Chambers, R. Simon, D.B. Hardie, C.H. Borchers, Multiplexed MRM with internal standards for cerebrospinal fluid candidate protein biomarker quantitation, J. Proteome Res. 13 (2014) 3733–3747.
[21] F.E. Grubbs, G. Beck, Extension of sample sizes and percentage points for significance tests of outlying observations, Technometrics 14 (1972) 847–854.
[22] A.J. Percy, R. Simon, A.G. Chambers, C.H. Borchers, Enhanced sensitivity and multiplexing with 2D LC/MRM-MS and labeled standards for deeper and more comprehensive protein quantitation, J. Proteomics 106 (2014) 113–124.
[23] A.J. Percy, A.G. Chambers, J. Yang, D.B. Hardie, C.H. Borchers, Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility, Biochim. Biophys. Acta 1844 (2014) 917–926.