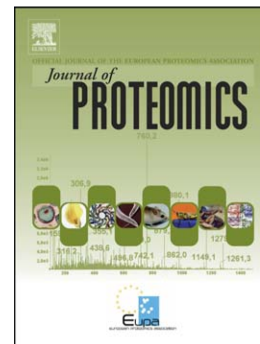
Accepted Manuscript



Multicenter Experiment for Quality Control of Peptide-Centric LC-MS/MS Analysis – A Longitudinal Performance Assessment with nLC Coupled to Orbitrap MS Analyzers

Alex Campos, Ramón Díaz, Salvador Martínez-Bartolomé, Jose Sierra, Oscar Gallardo, Eduard Sabidó, Maria López-Lucendo, J. Ignacio Casal, Carla Pasquarello, Alexander Scherl, Cristina Chiva, Eva Borras, Antonia Odena, Félix Elortza, Mikel Azkargorta, Nieves Ibarrola, Francesc Canals, Juan P. Albar, Eliandre Oliveira

PII: S1874-3919(15)30010-5
DOI: doi: 10.1016/j.jprot.2015.05.012
Reference: JPROT 2151

To appear in: *Journal of Proteomics*

Received date: 3 February 2015
Revised date: 7 May 2015
Accepted date: 11 May 2015

Please cite this article as: Campos Alex, Díaz Ramón, Martínez-Bartolomé Salvador, Sierra Jose, Gallardo Oscar, Sabidó Eduard, López-Lucendo Maria, Ignacio Casal J, Pasquarello Carla, Scherl Alexander, Chiva Cristina, Borras Eva, Odena Antonia, Elortza Félix, Azkargorta Mikel, Ibarrola Nieves, Canals Francesc, Albar Juan P., Oliveira Eliandre, Multicenter Experiment for Quality Control of Peptide-Centric LC-MS/MS Analysis – A Longitudinal Performance Assessment with nLC Coupled to Orbitrap MS Analyzers, *Journal of Proteomics* (2015), doi: 10.1016/j.jprot.2015.05.012

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Title:

Multicenter Experiment for Quality Control of Peptide-Centric LC-MS/MS Analysis – A Longitudinal Performance Assessment with nLC Coupled to Orbitrap MS Analyzers

Corresponding author:
Alex Campos
Proteomics Core Facility
Barcelona Science Park
c/ Baldiri Reixac, 10
Barcelona, Spain
Email: alex.proteomics@gmail.com
Phone: +34 670 526 871

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Authors:

Alex Campos [1,2], Ramón Díaz [1], Salvador Martínez-Bartolomé [3], Jose Sierra [2], Oscar Gallardo [4], Eduard Sabidó [5], Maria López-Lucendo [6], J. Ignacio Casal [6], Carla Pasquarello [7], Alexander Scherl [8], Cristina Chiva [5], Eva Borras [5], Antonia Odena [1], Félix Elortza [9], Mikel Azkargorta [9], Nieves Ibarrola [10], Francesc Canals [11], Juan P. Albar [3], Eliandre Oliveira [1]

[1] ProteoRed-ISCIII, Proteomics Platform, Barcelona Science Park, Barcelona, Spain
[2] Integromics, Madrid, Spain
[3] ProteoRed-ISCIII, Proteomics Facility, Centro Nacional de Biotecnología - CSIC, Madrid, Spain
[4] ProteoRed-ISCIII, CSIC/UAB Proteomics Laboratory, Instituto de Investigaciones Biomédicas de Barcelona, Spanish National Research Council, Barcelona, Spain
[5] ProteoRed-ISCIII, Proteomics Unit, Universitat Pompeu Fabra (UPF) and Centre de Regulació Genòmica (CRG), Barcelona, Spain
[6] ProteoRed-ISCIII, Proteomics Facility and Functional Proteomics Laboratory, Centro de Investigaciones, Biológicas, Madrid, Spain
[7] Proteomics Core Facility, CMU, University of Geneva, Switzerland
[8] Department of Human Protein Sciences, CMU, University of Geneva, Switzerland
[9] ProteoRed-ISCIII, Proteomics Platform, CIC bioGUNE, CIBER*ehd*, Technology Park of Bizkaia, Derio, Spain
[10] ProteoRed-ISCIII, Centro de Investigación del Cáncer and Instituto de Biología Molecular y
Celular del Cáncer, CSIC-University of Salamanca, Salamanca, Spain
[11] ProteoRed-ISCIII, Proteomic Laboratory, Vall d'Hebron Institute of Oncology-VHIO, Vall d'Hebron University Hospital, Barcelona, Spain

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Proteomic technologies based on mass spectrometry (MS) have greatly evolved in the past years, and nowadays it is possible to routinely identify thousands of peptides from complex biological samples in a single LC-MS/MS experiment. Despite the advancements in proteomic technologies, the scientific community still faces important challenges in terms of depth and reproducibility of proteomics analyses. Here, we present a multicenter study designed to evaluate long-term performance of LC-MS/MS platforms within the Spanish Proteomics Facilities Network (ProteoRed-ISCIII). The study was performed under well-established standard operating procedures, and demonstrated that it is possible to attain qualitative and quantitative reproducibility over time. Our study highlights the importance of deploying quality assessment metrics routinely in individual laboratories and in multi-laboratory studies. The mass spectrometry data have been deposited to the ProteomeXchange Consortium with the data set identifier PXD000205.

Significance

Over the past years, the scientific community has shown an increased concern about failures in the reliability and reproducibility of published research articles. In particular, proteomics technologies are more mature and have become more accessible to the scientific community. To assure quality of the proteomics data, performance assessment of LC-MS/MS system is critical. We show here that the implementation of routine performance assessment promotes standardization across multiple laboratories in a long-term basis. Using a reasonably complex proteome sample, we show that it is possible to detect minor decays in LC-MS/MS system performance.

Introduction

In the past years, proteomic platforms combining enzymatic protein digestion followed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) data-dependent acquisition analysis have become the mainstay of screening proteomics profiling [1]. Despite the technological advancements achieved throughout these years, the proteomics community still faces important challenges related to the sensitivity and reproducibility of proteomic analysis. In particular, the problem of variability may significantly impact comparative studies aimed to profile biological states or phenotypes [2].

There may be several reasons for the lack of reproducibility across and within proteomics laboratories including sample preparation [3], peptide/protein fractionation, and LC-MS/MS analysis [4]. In addition, inappropriate data analysis can introduce false results, consequently affecting the reproducibility across proteomic platforms [5]. Regardless the goal of any proteomics study, its practical utility will ultimately depend upon the quality of the obtained data. In this context, quality control (QC) and performance metrics have recently gained much attention within the proteomics community [6-10].

To guarantee the quality of LC-MS/MS results, individual proteomics

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

laboratories have deployed and reported different QC metrics to help monitoring their overall performance such as signal stability, identification sensitivity, and repeatability [11]. Between 2009 and 2010, the Clinical Proteomic Technology Assessment for Cancer (CPTAC) network published a number of reports on a series of multicenter studies aimed to evaluate repeatability and reproducibility of unbiased [4, 12, 13] and targeted [14] proteomic experiments. In particular, Rudnick et al. [13] described the implementation of QC metrics for monitoring performance of different LC-MS/MS system components in unbiased proteomics experiments. A total of 46 performance metrics were described aiming to monitor LC and electrospray source (ESI) stability, MS1- and MS2-level signal stability, dynamic sampling of precursor ions for fragmentation, and peptide-to-spectrum identification. Many of these performance metrics were subsequently deployed in Paulovich et al. [12] that evaluated LC-MS/MS performance in four independent ion trap-based systems using yeast *Saccharomyces cerevisiae* standard prepared under a well-described Standard Operating Procedure (SOP).

The CPTAC work raised awareness across the proteomics community of the importance of deploying QC metrics. In the past years, several bioinformatics tools have been developed to facilitate assessing performance of LC-MS/MS systems [15-24]. Some of these tools have been expertly reviewed in Bereman [25] and are beyond the scope of the present study. Instead, we herein take a step further to the CPTAC study and evaluate long-term quality of sample preparation and LC-MS/MS analysis across 6 independent laboratories operating hybrid ion trap/orbitrap mass spectrometers. We analyze the performance of proteomics platforms over a six months period, and highlight the main sources of variability within and across platforms. This study also highlights the importance of deploying quality control metrics for monitoring performance of sample preparation and LC-MS/MS analysis. Finally, we thoroughly annotated the dataset generated in this study, and made it available to the scientific community through the ProteomeXchange repository [26].

Material and methods

Generation of Yeast Protein Standard Sample

An SOP for preparation of the yeast standard sample was developed as previously described in Paulovich *et al.* [12]. In brief, *S. cerevisiae* strain BY4741 (MATa, *leu2Δ0, met15Δ0, ura3Δ0, his3Δ1)* was grown in a 10-liter batch of rich (yeast extract peptone dextrose) medium at 30°C in a fermentor to an *A600* of 0.93. After harvesting by continuous centrifugation, the cell pellet was washed three times with ice-cold water. The cells were lysed by incubation with 10% ice-cold trichloroacetic acid (v/v) for 1 h at 4°C. The protein precipitate was collected by centrifugation, washed twice with 90% ice-cold acetone, and pelleted again. The resulting material was aliquoted, lyophilized and stored at -80°C.

Preparation Yeast Standard Sample

The present study is divided in Study A and Study B, originally designed to allow evaluation of (1) LC-MS/MS variability within and across proteomics platforms, and (2) repeatability of protein digestion procedures, respectively.

ACCEPTED MANUSCRIPT

Complete instructions to this study can be found in Supplementary material (Study Guidelines). Briefly, sample preparation for both sub-studies was as follows.

Lyophilized yeast lysate (1 mg) was reconstituted in 8 M urea, 10 mM DTT in 50 mM NH_4HCO_3, and incubated for 45 min. Then, cysteine residues were alkylated with 30 mM iodoacetamide (final concentration) for 30 min. Prior to trypsinization, an additional volume of 50 mM NH_4HCO_3 was added to the sample to reduce the urea concentration to 1 M. A total of 50 µg of trypsin was added to sample and incubated overnight (16 h) at 32°C with gentle swirling. Following digestion, samples were acidified with formic acid (FA) and subsequently desalted using a 500 mg tC18 Sep-Pak SPE cartridge. C18 cartridges were conditioned with 3 ml of 100% MeCN, followed by 3 ml of 50% MeCN/0.1% FA, and finally 10 ml of 0.1% TFA. Peptides were then loaded onto the conditioned C18 cartridge, washed with 5 ml of 0.1% TFA, and eluted with 1 ml of 60% MeCN/0.1% FA. After reducing the concentration of MeCN in the SpeedVac, aliquots of 50 µg of yeast-digested protein sample was prepared, lyophilized to dryness and mailed to study participants. The digested yeast protein sample was named ‘sample A’, referring to the sample used in Study A. Along with ‘sample A’, each laboratory also received the ‘sample B’, an aliquot of undigested dried yeast proteins. All samples were delivered to the participant laboratories in dry-ice within 36 hours.

Analysis Guidelines

In the study A, participant laboratories were asked to inject 200 ng of digested yeast sample onto their LC-MS/MS system at least once a month during a period of at least 6 months. In study B, participant laboratories were asked to perform their own protein digestion procedure (Supplementary material – Digestion protocol comparative table) once a month and then inject 200 ng of this digested sample using the same method used in study A. In general, LC-MS/MS analyses were performed using a 110-minutes gradient followed by high organic solvent flow; participant laboratories were allowed to make minor modifications to the LC protocol in order to adapt to specific laboratory’s system particularities. Thorough description of LC and MS settings used by each participant laboratory is given in Supplementary material (LC-MS/MS settings comparative table).

Data Analysis

To enable inter-laboratory comparisons, data analysis was centralized and performed under standard procedures using a common Quality Control (QC) workflow deployed in the OmicsHub® Proteomics (OHP) platform from Integromics (www.integromics.com). An overview of the OHP analysis workflow is shown schematically in Fig. 1, and further described in Supplementary material (Quality control data analysis workflow). Briefly, raw MS data files submitted to the study repository was converted to mzML format. While MS2 data was searched with X!Tandem Cyclone, and the resulting matches were statistically validated with PeptideProphet. MS1 data was processed with OpenMS to detect peptide features. In our lexicon, a peptide feature is defined as a peptide ion

ACCEPTED MANUSCRIPT

signal detected at the MS1 level integrated over its elution time (extracted ion current). Ultimately, a feature is denoted by its isotopic pattern in mass-to-charge dimension, by the elution profile in retention time dimension and by the total intensity (i.e., the area under the curve of the chromatographic elution period). Following its detection, features were annotated using MS2 identification data. Finally, OHP generates a QC metrics report describing different aspects of proteomics analysis such as proteome coverage, reproducibility and instrument stability.

All participating laboratories were anonymized and coded according to the MS instrument used for analysis (OrbiXL and OrbiVL for Orbitrap XL and Orbitrap Velos, respectively). Raw and search result data have been deposited in the Proteome Exchange Consortium (http://proteomecentral.proteomexchange.org) [26] via the PRIDE partner repository [27] with the dataset identifier PXD000205.

Results

Study Organization

The present study was conceived with the primary goal of assessing quality of proteomics data generated by six independent peptide-centric proteomics platforms for a period of 6 months. To accomplish this goal, a trypsin-digested yeast protein sample was prepared (study A sample) and distributed to each laboratory along with a non-digested yeast protein sample (study B sample) (Fig. 1). Because protein digestion in study A sample was carried out centrally, potential differences between laboratories are expected to be due to differences in the LC-MS/MS settings and performance. On the other hand, in study B, the protein sample was digested individually in each laboratory. Potential differences between studies A and B within a given laboratory can therefore be attributed to differences in protein digestion efficiency since all LC-MS/MS settings for studies A and B were kept identical throughout the study. The total number of QC experiments carried out by each participant laboratory ranged from 8 to 21 and from 4 to 8 for studies A and B, respectively (Fig. 1). In the following sections, we summarize our key findings and provide technical insights to better understand and pinpoint the main sources of variation in proteomics data.

Identification Counts

Arguably, the number of PSMs and peptides confidently identified in a QC sample are the most commonly used performance metrics by MS-based proteomics laboratories. We evaluate the efficiency of proteomics platforms in terms of identification counts at the PSM, ion, peptide and protein levels. In our lexicon, ***PSMs*** are defined as the collection of all confidently (PeptideProphet probability >= 0.95) identified MS2 spectra in a QC experiment and differ from ***ions*** in that the later is the list of ‘unique PSMs’. In other words, two or more PSMs with identical peptide sequence, charge state and modifications (if any) give rise to one ion. On the other hand, ions with the same peptide sequence, regardless the charge state and whether any amino acid is modified (e.g., oxidation, pyroglutamic acid), is counted as one peptide. Fig. 2 summarizes the number of PSM (nPSM), ion (nION), peptide (nPEPT), and protein (nPROT)

ACCEPTED MANUSCRIPT

identified in study A within each laboratory. It is notably that sites OrbiVL1, OrbiVL2, OrbiVL3 and OrbiVL5 achieved similar nPSM, nION and nPEPT, while OrbiVL4 and OrbiXL1 generated significantly less identifications. To determine whether the means of number of identifications throughout the study were statistically significant between laboratories, we used the Tukey-Kramer test together with comparison circles (Fig. 2 right-side of box plots). The methodology used to draw Comparison Circle and calculate Tukey-Kramer statistics is provided in Supplementary material (Quality control data analysis workflow). It is plausible that the OrbiXL1 platform exhibited an overall poorer performance due to the MS instrument itself; while the OrbiXL1 platform deployed a first generation Orbitrap XL MS model, the other 5 platforms made use of a newer generation of Orbitrap MS, namely Velos.

In general, MS methods are designed to achieve high proteome coverage, in a reproducible and quantitative fashion. We here noted that a higher number of PSMs did not necessarily translated into a higher number of peptide identifications. Overall, precursor oversampling (the ratio between nPSM and nPEPT or nION) was consistently high for most laboratories throughout the study, except at OrbiVL2 and OrbiVL4 sites, which yielded ratios close to one (Supplementary Fig. 1). These results suggest that, in the QC experiments performed in these 2 laboratories, identical PSMs were barely selected for MS/MS fragmentation more than once. The most plausible explanation for this effect lies on MS acquisition settings, in particular the high dynamic exclusion setting used at OrbiVL2 and OrbiVL4 sites (60s and 180s, respectively). It is also noteworthy that these two laboratories set their maximum “FT full max MS ion injection time” to 10 ms, which was considerably below what other laboratories used in this study (see Supplementary material - LC-MS/MS settings comparative table). The effect of this setting on precursor oversampling is not clear to us. Supplementary Fig. 1 details the attained number of identifications at every QC experiment in this study for each laboratory.

In a typical peptide-centric LC-MS/MS analysis, thousands of full-scan MS (MS1) and data-dependent MS/MS (MS2) spectra are generated. Throughout this study, most laboratories consistently yielded similar number of MS1 spectra (nMS1). Nevertheless, the number of MS2 spectra (nMS2) varied substantially within most laboratories (see Supplementary Fig. 2a). The average QC experiment nMS2:nMS1 ratio for each laboratory varied from 2 to 5, likely due to individual MS acquisition settings (see Supplementary material - LC-MS/MS settings comparative table). The exception was the OrbiVL4 site that yielded on average twice more MS1 spectra than nMS2 spectra (Supplementary Fig. 2c). It is likely that the low MS2 sampling rate influenced the identification performance (Fig. 2) observed at OrbiVL4 site as Supplementary Fig. 2b shows an evident correlation between nMS2 and nPSM and/or nPEPT.

An MS/MS spectrum may remain unidentified for several reasons. For example, unaccountable peptide modifications will not be identified even though a high-quality spectrum is generated. Despite many peptides may not confidently identified by (or selected for) MS/MS, it is likely that those unidentified peptides within the sensitivity range of the MS analyzer will give rise to a detectable three-

ACCEPTED MANUSCRIPT

dimensional LC-MS signal denoting the peptide elution time range, monoisotopic m/z peak and ion current (area under the curve). The 3D LC-MS signal of a peptide is termed *peptide feature*, and is summarized by its elution time apex (RT), m/z of the centroided monoisotopic peak (m/z), total peptide ion current (Intensity), and the ion charge state. In our QC analysis workflow, features detected with a charge state between 2+ and 4+ were kept for downstream analysis. Detected features were annotated with a confidently identified MS/MS (see Supplementary material – Quality control data analysis workflow). The median number of features detected (nFEAT), which includes annotated (nFEATA) and unannotated (nFEATU) features, throughout this study within laboratories ranged from 7769 to 30991 (Supplementary Fig. 3). Our study shows that it is possible to detect as many as 40000 features from 200 ng of digested yeast proteins (on column) in a single 2-hours LC-MS/MS analysis. Although it is unlikely that every detected feature would render an unique peptide identification, it is also evident that nION is lower than nFEAT, evidencing the undersampling problem in peptide-centric approaches. Supplementary Fig. 3c shows the proportion of annotated features (i.e., nFEATA over nFEAT, herein pFEAT) in every QC experiment laboratory wise. We found that the overall median pFEAT within each site ranged from 18 to 24%, and the maximum pFEAT throughout across all sites was 31%.

Reproducibility

The goal of many proteomics studies is to compare protein identification and/or abundance levels between two or more experimental conditions. To this end, reproducible proteomics data is essential for generating reliable results. Here we evaluated longitudinal reproducibility of proteomics experiments at the qualitative and quantitative levels. Qualitative reproducibility was determined as the fraction of identifications (i.e., feature, ions, peptides, or proteins) in common (intersection) between two QC experiments given the total number of unique identifications (union) in both QC experiments. Thus, the overlap between experiments X and Y is calculated as:

$$XY\ Overlap\% = (X \cap Y) / (X \cup Y) * 100 \qquad (\textbf{\textit{Equation 1}})$$

Noteworthy, only identifications with at least one confident PSM (PeptideProphet probability ≥ 0.95) were used in this analysis. Supplementary Fig. 4 shows pairwise identification overlap between every QC experiment (within each site) at the protein, peptide, ion and feature levels. As expected, higher overlap was attained at the protein level followed by peptide and then ion levels. Fig. 3 summarizes pairwise identification overlap calculated between QC experiments within either study A (A vs. A) or study B (B vs. B), and between both studies (A vs. B). Fundamentally, the overlap within study A assesses technical reproducibility (LC-MS/MS analysis) whereas overlap within study B evaluates experimental reproducibility (i.e., sample preparation and LC-MS/MS). On the other hand, the overlap between QC experiments from studies A and B can be seen as the reproducibility in terms of sample preparation between two independent laboratories. Our results demonstrate that reasonable long-term

ACCEPTED MANUSCRIPT

repeatability can be attained technically and experimentally. The overall median of peptide overlap medians (Fig. 3) in study A across laboratories was 51.5% (with an interquartile range of 45.6 – 56%). Tabb *et al.* reported a median peptide overlap value of 53% (interquartile range of 44 – 58%) across 6 independent laboratories following triplicate injection of 600 ng on column of a yeast standard sample [4].

Reproducibility was also evaluated at the quantitative level by calculating intensity correlation (Pearson's correlation coefficient) of annotated features in common between QC experiments pairwisely (Supplementary Fig. 5). Fig. 4 shows the distribution of calculated correlation coefficients throughout the entire study within the 3 possible study comparison levels (A vs. A, A vs. B, and B vs. B). Despite the general notion that attainable feature intensity correlation coefficients would be superior in technical replicates (A vs. A comparisons) compared to experimental replicates (B vs. B comparisons), we noted that this presumption did not hold true at OrbiVL1 and OrbiVL3 sites (see *Comparison Circles* along with the box-and-whiskers plots in Fig. 4). Remarkably, we found significant evidence of reproducibility deterioration of feature intensity over time. Supplementary Fig. 5 shows feature intensity correlation coefficients between pairs of QC experiments for each laboratory. It is noteworthy that QC experiments performed months apart may show better quantitative and qualitative reproducibility compared to other QC experiments performed within a shorter period of time.

Quality Assurance of Sample Handling

It has been demonstrated that differences in sample handling can have significant impact on proteome coverage and experimental reproducibility []. After receiving custody of samples for proteomics analysis, laboratory staff should ensure sample integrity and proper handling. To assess the efficiency of sample handling, we benchmarked study B against study A. In other words, the results obtained from study A in each laboratory was deemed the 'gold standard', which was used as reference for the results obtained in study B. Since study A required minor sample handing and LC-MS/MS settings used in both studies were identical, we assumed that potential differences in the results generated for both samples should be mainly due to sample handling including digestion protocols in each site. Supplementary Fig. 6 reports the number of identifications for each QC experiment in study B. Fig. 5 summarizes the identification results within studies A and B, along with *comparison circle analysis* to show whether differences in number of identifications between both studies are statistically significant. The degree of identification overlap between studies A and B was laboratory-dependent as shown with the comparison circles in Fig. 3. It implies that some laboratories, most notably OrbiVL3, generated a slightly different subset of peptides and proteins. Although we initially surmised that dissimilarities between studies A and B were due to differences in digestion protocols, we remark that the OrbiVL3 site deployed urea-based protein digestion protocol (Supplementary material – Digestion protocol comparative table) similar to OrbiVL1 and OrbiVL2 sites that in turn showed higher overlap between both

ACCEPTED MANUSCRIPT

studies. In addition, the sample used in study A was also prepared using urea-based digestion protocol similar to that used at OrbiVL3 site. It is also noteworthy the high identification overlap attained between studies A and B at OrbiVL5 site despite the fact that this site deployed a different TFE-based digestion protocol. In our view, identification differences between laboratories open up an opportunity to increase proteome coverage in multilaboratory studies.

Protein digestion efficiency is often evaluated in terms of digestion completeness. Here we calculated the proportion of PSMs with 1 or 2 trypsin-missed cleavage sites (pMC1 and pMC2, respectively) throughout study B (Supplementary Fig. 7). The median pMC1 in Study B ranged from 2.5 to 32.4%, with an overall median of 10.57% (Fig. 6). The low digestion efficiency observed at OrbiXL1 site (median pMC1 and pMC2 was 32.4% and 5.5%, respectively) highlights the importance of deploying quality metrics to monitor sample preparation.

Undesirable peptide modifications due to inappropriate sample handling will likely affect proteomics results qualitatively and quantitatively. We monitored the proportion of PSMs containing peptide modifications such as oxidized methionine (pOxMet) as a quality metric for sample handling. Although not every oxidized methionine residue can be attributed to inappropriate sample handling [28, 29], we argue that monitoring changes in peptide modifications may provide important clues about sample integrity. In this study, we found that the pOxMet was significantly lower in freshly prepared digestion samples (study B pOxMet median=0.5%, SD=0.7) compared to digested samples stored for longer periods of time (study A pOxMet median=1.5%, SD=2.5), except for the OrbiVL2 site (Supplementary Fig. 8).

Performance Metrics for Liquid Chromatrography and Electrospray Ionization

Optimal performance of both liquid chromatography (LC) separation and electrospray ionization (ESI) are crucial for generating high quality MS data. Some of the observed variability in study A could be explained by potential instability of the LC and/or ESI systems. Here, we deployed two simple performance metrics that help monitoring stability of LC and ESI components.

Liquid Chromatography – After surveying the 6 study participants, we noted a lack of consensus criteria regarding making the decision to replace LC columns due to its lack of efficiency. Here, we use chromatography peak width to assess performance of LC column. In general, optimal chromatography width is important to reduce oversampling, and improve quantitative signal accuracy. To this end, we calculated the full width at half maximum (FWHM) for all detected features (CWF), or annotated features (CWAF) only (Supplementary Fig. 9). The median of CWAF median for all QC experiments within each laboratory ranged from 14.9 to 24.3 seconds. Noteworthy, after excluding OrbiVL4 and OrbiXL1 sites, the CWAF median was quite consistent and ranged from 14.9 to 17.4 seconds. Interestingly, we observed a direct correlation between CWAF broadening and the number of identifications and features (Supplementary Fig. 10). This could therefore explain the poor performance observed for OrbiVL4 and

ACCEPTED MANUSCRIPT

OribXL1 sites in terms of number of identifications (Fig. 2) and features detected and annotated (Supplementary Fig. 3).

Electrospray ionization – ESI is commonly neglected when it comes to troubleshooting LC-MS/MS systems. To evaluate the robustness of ESI throughout the study, we calculated the MS1 total ion count signal drop (TICdrop) metric. That is, for each LC-MS/MS analysis, TICdrop computes the number of times the MS1 total ion count signal varies more than 10-fold between consecutive MS1 scans. It therefore assumes that a drastic decrease of TIC signal (10x) between consecutive MS1 scans causes ESI instability. In this study, we observed a direct correlation between the TICdrop metric and nFEATA, as QC experiments with a relative high number of TICdrops also showed reduced number of features annotated (Supplementary Fig. 11). This observation emphasizes the importance of monitoring ESI stability, particularly in label-free intensity-based experiments where the detected peptide features make up the basis of quantitative comparison across samples.

Conclusions

Multi-laboratory collaborative studies are often difficult to manage due to general differences across proteomics platforms. The Spanish proteomics laboratory network, ProteoRed, designed the present study to systematically evaluate performance across 6 consortium proteomics platforms in a long-term fashion. The study was performed under well-established guidelines using identical yeast standard sample, and similar mass spectrometry instrumentation (i.e., Thermo Orbitrap Velos and XL) and liquid chromatography systems (i.e., Waters NanoAcquity or Proxeon Easy). Laboratories were allowed to set their preferred data-dependent acquisition LC-MS/MS method; and throughout the study, all analyses were carried out using the same LC-MS/MS settings. Since key system variables were held constant throughout the study, we here attributed most observed differences to the inherent variability of proteomics platforms.

In this study, most laboratories demonstrated fair reproducibility over time. Best platform performance in terms of peptide and protein identification overlap across technical replicates was 55% (interquartile range of 53.5 – 57.5%) and 71% (interquartile range of 69.2 – 72.4%), respectively. To the best of our knowledge, this represents the current state-of-the-art in data dependent acquisition LC-MS/MS analysis. This observation is corroborated by a recent multi-laboratory study using similar workflow (i.e., yeast standard sample, LC-MS/MS systems and analysis workflow), and that reported a median of 53% (interquartile range of 44 – 58%) of peptide overlap following triplicate injection of 600 ng on column.

In the past years, technological and bioinformatics innovation have been crucial to increase further the depth of proteome analysis [30]. Despite some minor experimental differences, we compared the best performance laboratories in this study with recent reports deploying similar or novel LC-MS/MS technologies. We found that the number of peptides (~8000) and proteins (~1600) routinely identified by the best performance laboratories was similar to what was reported by Thakur *et al.* [31] where Orbitrap Velos were also used.

ACCEPTED MANUSCRIPT

More recently, Hebert *et al.* [32] reported a significant increase in the number of proteins identified in a single-shot LC-MS/MS analysis using the newly introduced Orbitrap hybrid mass spectrometer (Orbitrap Fusion). The authors claimed almost comprehensive identification of the yeast proteome (approximately 34,000 unique peptides and 4000 proteins) using a 70-minutes LC separation

The ProteoRed-ISCII consortium has engaged in the Chromosome-centric Human Proteome Project (C-HPP), focusing on the characterization of Chromosome 16 (spHPP) [33, 34]. In this context, the present study provided the consortium laboratories with the opportunity to evaluate their performance, and identify strengths and weaknesses in their platform. We observed that the best performance laboratories in this study have more years of experience working with LC-MS/MS analysis. Variability and poor performance observed in sporadic QC experiments in this study were at least partially explained by malfunction of one or more of the LC-ESI-MS/MS components. Here, we highlighted the importance of deploying QC metrics to monitor system performance in a routine basis, and note that although the present study was performed on Orbitrap instruments, most of the performance metrics used here are also deployable in other LC-MS/MS instrument setups. Currently, a large number of LC-MS/MS performance metrics have been proposed. Although we believe that many of them are valuable and may help to pinpoint malfunction of different parts of a LC-MS/MS system, we also reckon that it is unpractical to many proteomics practitioners to monitor a large number of metrics on a regular basis. To this end, a small number of user-friendly QC metrics may facilitate its widespread use. In particular, we found the commonly used ‘number of PSMs confidently identified (herein, nPSM) and the number of features detected and their elution width (i.e., nFEAT and CWAF, respectively) among the most informative metrics in terms of long-term evaluation of LC-MS/MS system suitability. Overall, these metrics provide a good snapshot of the performance of both LC and MS components.

We also urge instrument manufacturers to incorporate such QC metrics in their MS acquisition software. Finally, the data presented here and made available through ProteomeXchange repository should provide the proteomics community with a unique dataset for longitudinal studies of proteomics experiments. We believe that it may foster the development of novel data analysis strategies aimed to increase depth of proteome identification.

ACKNOWLEDGEMENTS:

The authors wish to dedicate this manuscript to the memory of Juan Pablo Albar. We thank Juan Antonio Vizcaino, Attila Csordas, Rui Wang, and Florian Reisinger for assistance with PRIDE/ProteomeXchange submission process. This work was supported by the Instituto de Salud Carlos III (ProteoRed grant #2005X747).

ACCEPTED MANUSCRIPT

REFERENCES:

[1] Mallick P, Kuster B. Proteomics: a pragmatic perspective. Nature biotechnology. 2010;28:695-709.
[2] Lilley KS, Deery MJ, Gatto L. Challenges for proteomics core facilities. Proteomics. 2011;11:1017-25.
[3] Burkhart JM, Premsler T, Sickmann A. Quality control of nano-LC-MS systems using stable isotope-coded peptides. Proteomics. 2011;11:1049-57.
[4] Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJ, Bunk DM, et al. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. Journal of proteome research. 2010;9:761-76.
[5] Falick AM, Lane WS, Lilley KS, MacCoss MJ, Phinney BS, Sherman NE, et al. ABRF-PRG07: advanced quantitative proteomics study. Journal of biomolecular techniques : JBT. 2011;22:21-6.
[6] Kinsinger CR, Apffel J, Baker M, Bian X, Borchers CH, Bradshaw R, et al. Recommendations for mass spectrometry data quality metrics for open access data (corollary to the Amsterdam Principles). Molecular & cellular proteomics : MCP. 2011;10:O111 015446.
[7] Martens L, Vizcaino JA, Banks R. Quality control in proteomics. Proteomics. 2011;11:1015-6.
[8] Albar JP, Canals F. Standardization and quality control in proteomics. Journal of proteomics. 2013;95:1-2.
[9] Kocher T, Pichler P, Swart R, Mechtler K. Quality control in LC-MS/MS. Proteomics. 2011;11:1026-30.
[10] Eisenacher M, Schnabel A, Stephan C. Quality meets quantity - quality control, data standards and repositories. Proteomics. 2011;11:1031-6.
[11] Matzke MM, Waters KM, Metz TO, Jacobs JM, Sims AC, Baric RS, et al. Improved quality control processing of peptide-centric LC-MS proteomics data. Bioinformatics. 2011;27:2866-72.
[12] Paulovich AG, Billheimer D, Ham AJ, Vega-Montoto L, Rudnick PA, Tabb DL, et al. Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS platform performance. Molecular & cellular proteomics : MCP. 2010;9:242-54.
[13] Rudnick PA, Clauser KR, Kilpatrick LE, Tchekhovskoi DV, Neta P, Blonder N, et al. Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses. Molecular & cellular proteomics : MCP. 2010;9:225-41.
[14] Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nature biotechnology. 2009;27:633-41.
[15] Bereman MS, Johnson R, Bollinger J, Boss Y, Shulman N, MacLean B, et al. Implementation of statistical process control for proteomic experiments via LC MS/MS. Journal of the American Society for Mass Spectrometry. 2014;25:581-7.

ACCEPTED MANUSCRIPT

[16] Bittremieux W, Kelchtermans P, Valkenborg D, Martens L, Laukens K. jqcML: an open-source java API for mass spectrometry quality control data in the qcML format. Journal of proteome research. 2014;13:3484-7.
[17] Bittremieux W, Willems H, Kelchtermans P, Martens L, Laukens K, Valkenborg D. iMonDB: Mass Spectrometry Quality Control through Instrument Monitoring. Journal of proteome research. 2015.
[18] Ma ZQ, Polzin KO, Dasari S, Chambers MC, Schilling B, Gibson BW, et al. QuaMeter: multivendor performance metrics for LC-MS/MS proteomics instrumentation. Analytical chemistry. 2012;84:5845-50.
[19] Pichler P, Mazanek M, Dusberger F, Weilnbock L, Huber CG, Stingl C, et al. SIMPATIQCO: a server-based software suite which facilitates monitoring the time course of LC-MS performance metrics on Orbitrap instruments. Journal of proteome research. 2012;11:5540-7.
[20] Scheltema RA, Mann M. SprayQc: a real-time LC-MS/MS quality monitoring system to maximize uptime using off the shelf components. Journal of proteome research. 2012;11:3458-66.
[21] Such-Sanmartin G, Sidoli S, Ventura-Espejo E, Jensen ON. KYSS: mass spectrometry data quality assessment for protein analysis and large-scale proteomics. Biochemical and biophysical research communications. 2014;445:702-7.
[22] Sweredoski MJ, Smith GT, Kalli A, Graham RL, Hess S. LogViewer: a software tool to visualize quality control parameters to optimize proteomics experiments using Orbitrap and LTQ-FT mass spectrometers. Journal of biomolecular techniques : JBT. 2011;22:122-6.
[23] Walzer M, Pernas LE, Nasso S, Bittremieux W, Nahnsen S, Kelchtermans P, et al. qcML: an exchange format for quality control metrics from mass spectrometry experiments. Molecular & cellular proteomics : MCP. 2014;13:1905-13.
[24] Wang R, Fabregat A, Rios D, Ovelleiro D, Foster JM, Cote RG, et al. PRIDE Inspector: a tool to visualize and validate MS proteomics data. Nature biotechnology. 2012;30:135-7.
[25] Bereman MS. Tools for monitoring system suitability in LC MS/MS centric proteomic experiments. Proteomics. 2015;15:891-902.
[26] Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nature biotechnology. 2014;32:223-6.
[27] Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic acids research. 2013;41:D1063-9.
[28] Jamieson DJ. Oxidative stress responses of the yeast Saccharomyces cerevisiae. Yeast. 1998;14:1511-27.
[29] Liang X, Kaya A, Zhang Y, Le DT, Hua D, Gladyshev VN. Characterization of methionine oxidation and methionine sulfoxide reduction using methionine-rich cysteine-free proteins. BMC biochemistry. 2012;13:21.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[30] Perez-Riverol Y, Wang R, Hermjakob H, Muller M, Vesada V, Vizcaino JA. Open source libraries and frameworks for mass spectrometry based proteomics: a developer's perspective. Biochimica et biophysica acta. 2014;1844:63-76.
[31] Thakur SS, Geiger T, Chatterjee B, Bandilla P, Frohlich F, Cox J, et al. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Molecular & cellular proteomics : MCP. 2011;10:M110 003699.
[32] Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, et al. The one hour yeast proteome. Molecular & cellular proteomics : MCP. 2014;13:339-47.
[33] Segura V, Medina-Aunon JA, Guruceaga E, Gharbi SI, Gonzalez-Tejedo C, Sanchez del Pino MM, et al. Spanish human proteome project: dissection of chromosome 16. Journal of proteome research. 2013;12:112-22.
[34] Segura V, Medina-Aunon JA, Mora MI, Martinez-Bartolome S, Abian J, Aloria K, et al. Surfing transcriptomic landscapes. A step beyond the annotation of chromosome 16 proteome. Journal of proteome research. 2014;13:158-72.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

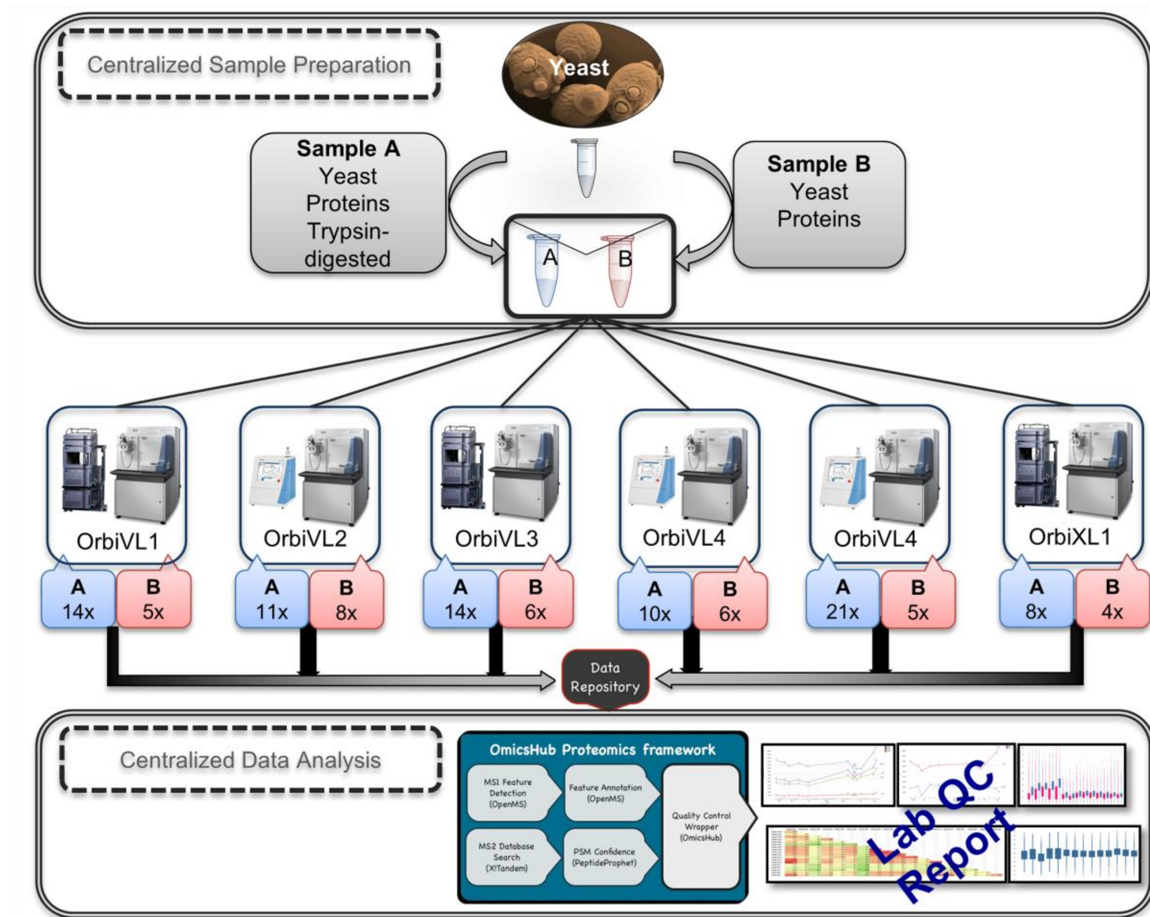
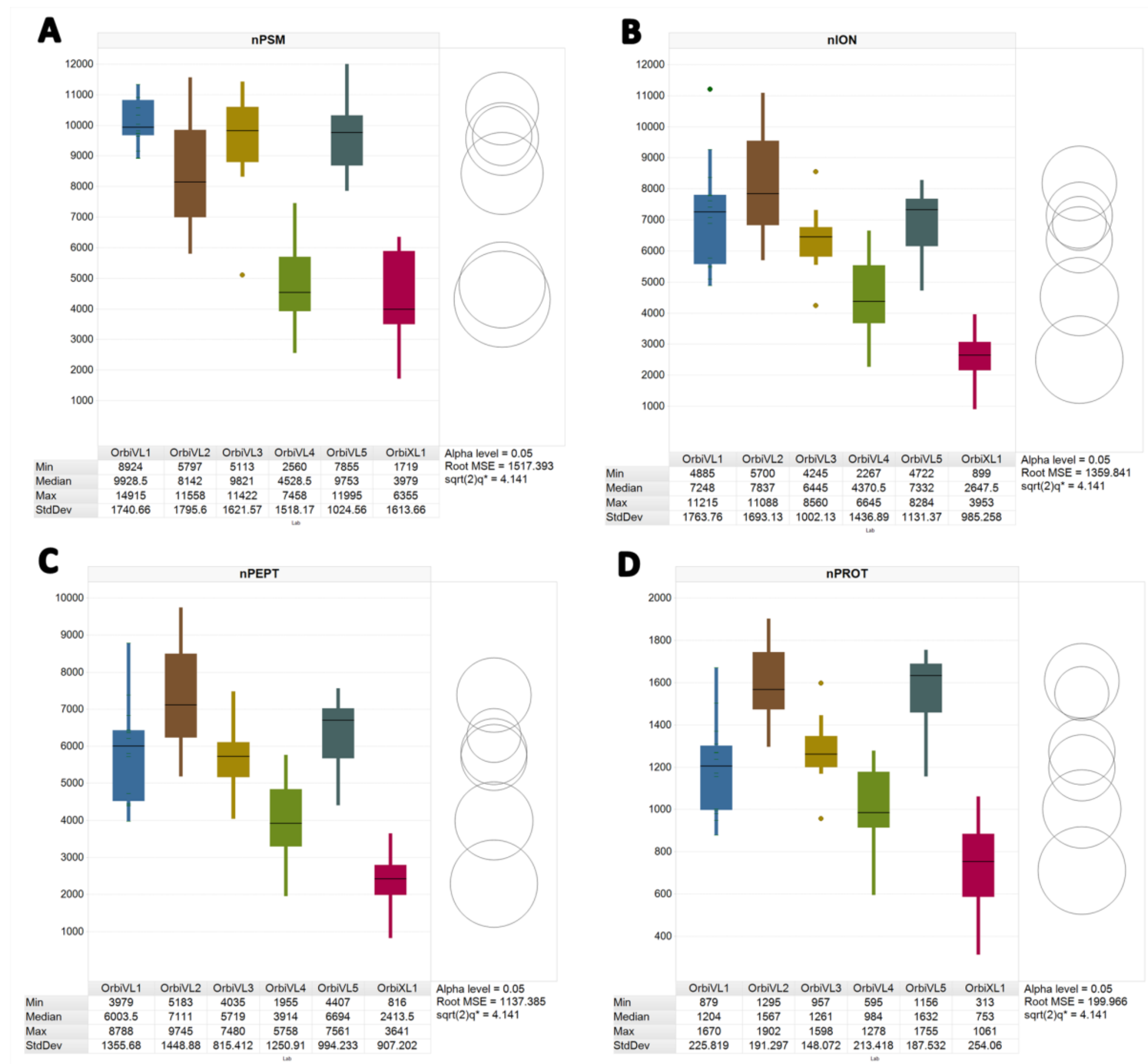


Figure 1

ACCEPTED MANUSCRIPT



	OrbiVL1	OrbiVL2	OrbiVL3	OrbiVL4	OrbiVL5	OrbiXL1
Min	8924	5797	5113	2560	7855	1719
Median	9928.5	8142	9821	4528.5	9753	3979
Max	14915	11558	11422	7458	11995	6355
StdDev	1740.66	1795.6	1621.57	1518.17	1024.56	1613.66

	OrbiVL1	OrbiVL2	OrbiVL3	OrbiVL4	OrbiVL5	OrbiXL1
Min	4885	5700	4245	2267	4722	899
Median	7248	7837	6445	4370.5	7332	2647.5
Max	11215	11088	8560	6645	8284	3953
StdDev	1763.76	1693.13	1002.13	1436.89	1131.37	985.258

	OrbiVL1	OrbiVL2	OrbiVL3	OrbiVL4	OrbiVL5	OrbiXL1
Min	3979	5183	4035	1955	4407	816
Median	6003.5	7111	5719	3914	6694	2413.5
Max	8788	9745	7480	5758	7561	3641
StdDev	1355.68	1448.88	815.412	1250.91	994.233	907.202

	OrbiVL1	OrbiVL2	OrbiVL3	OrbiVL4	OrbiVL5	OrbiXL1
Min	879	1295	957	595	1156	313
Median	1204	1567	1261	984	1632	753
Max	1670	1902	1598	1278	1755	1061
StdDev	225.819	191.297	148.072	213.418	187.532	254.06

Figure 2

ACCEPTED MANUSCRIPT

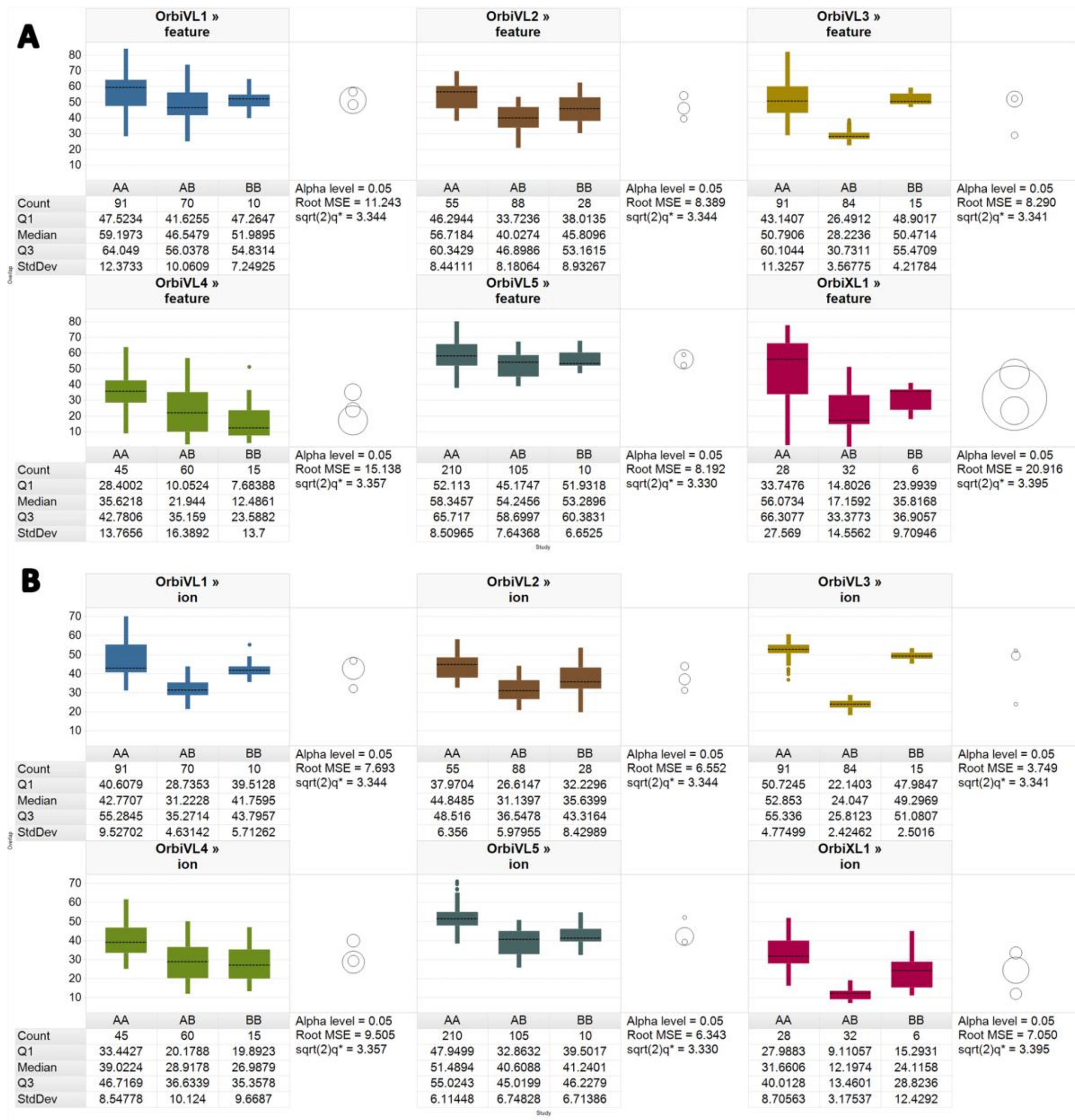


Figure 3ab

ACCEPTED MANUSCRIPT

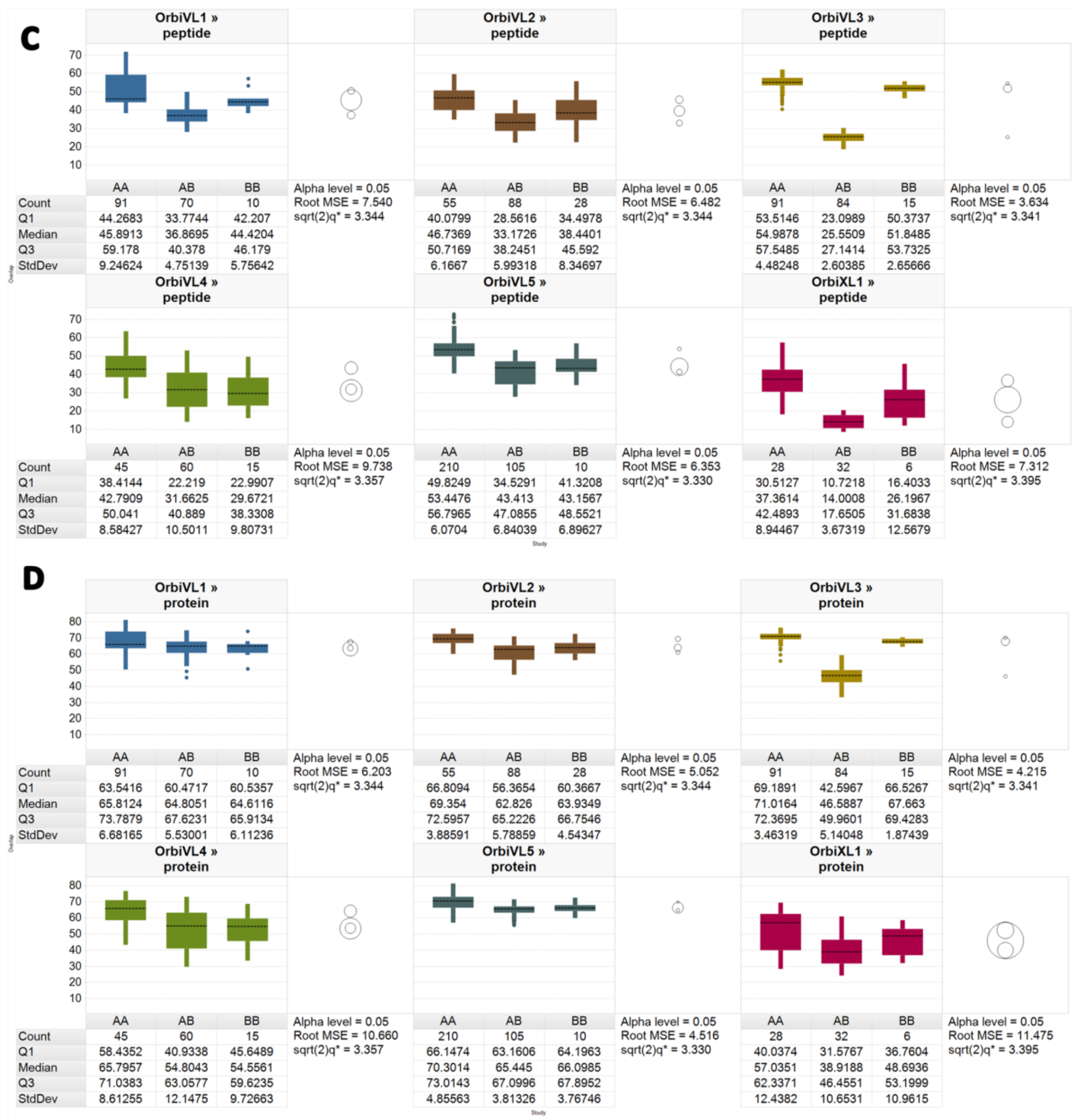


Figure 3cd

ACCEPTED MANUSCRIPT

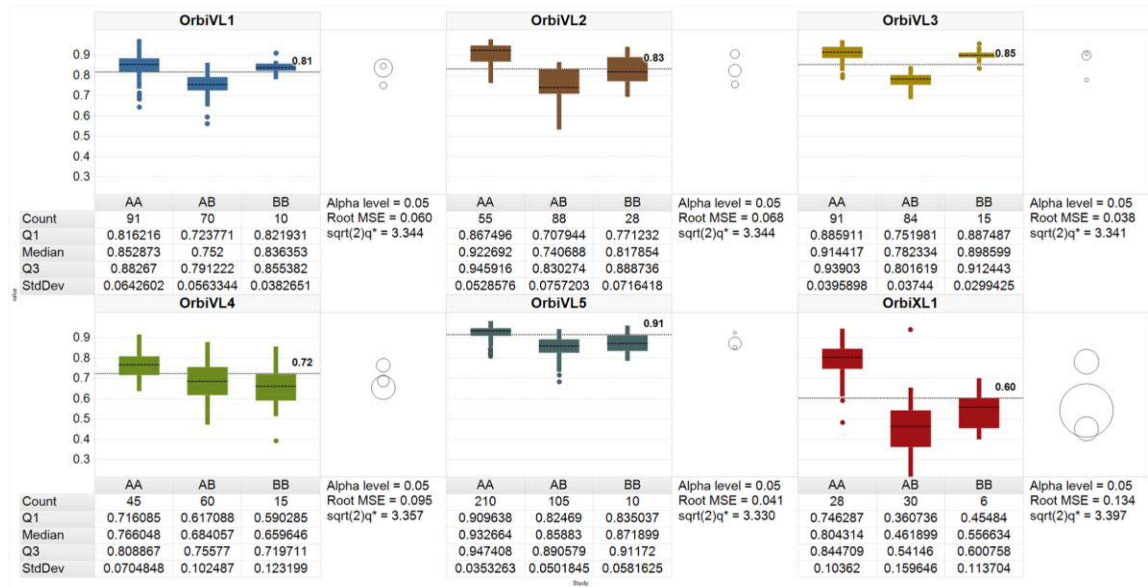


Figure 4

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

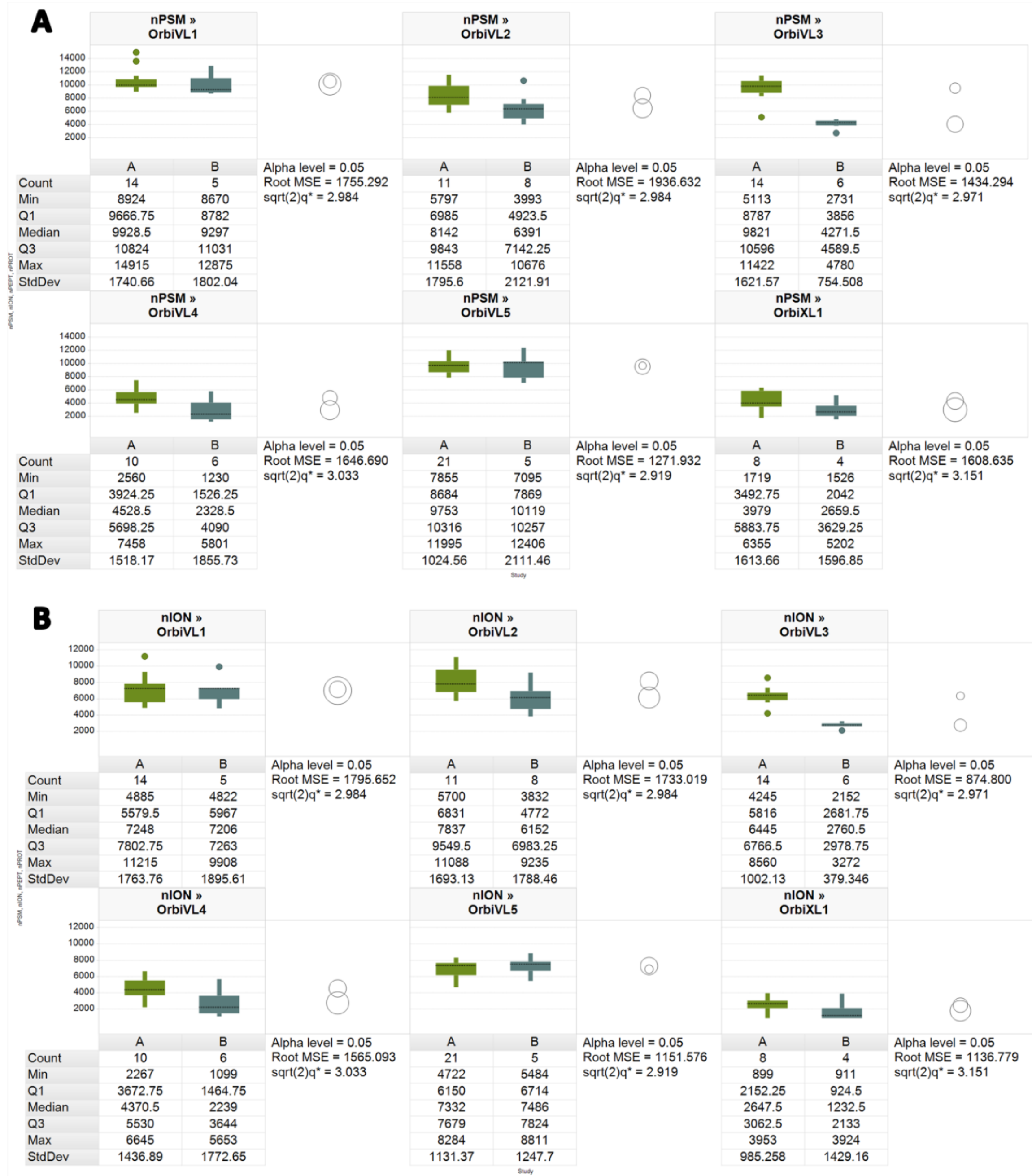


Figure 5ab

ACCEPTED MANUSCRIPT

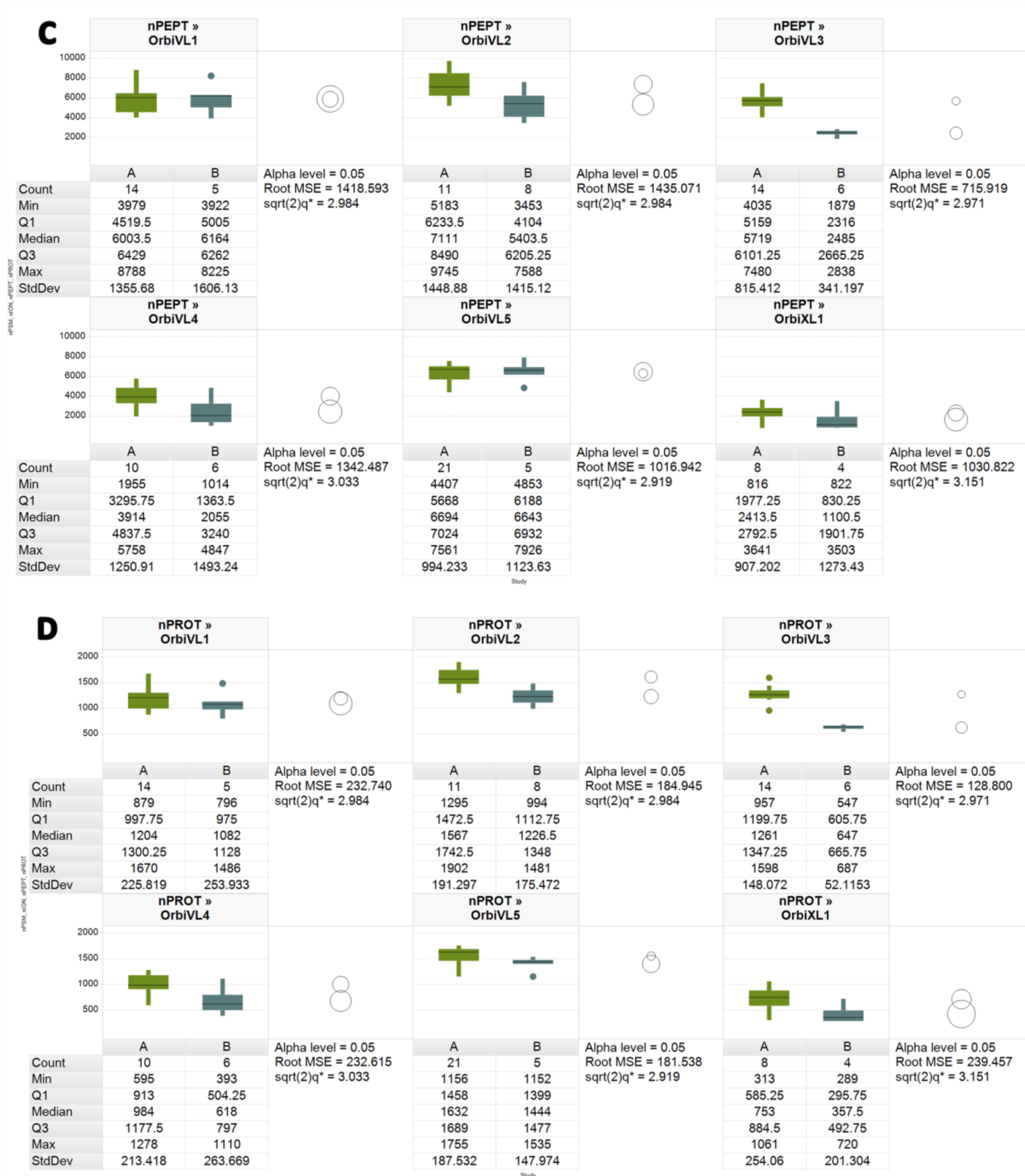
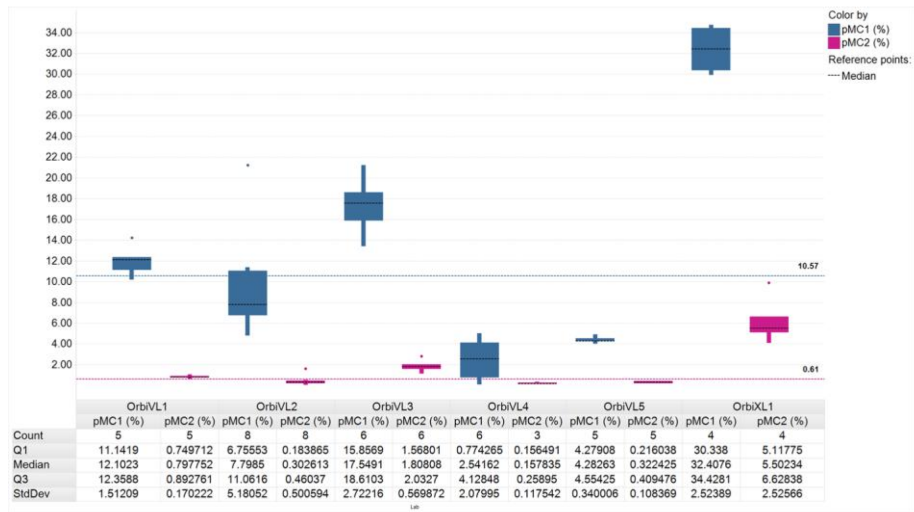


Figure 5cd

ACCEPTED MANUSCRIPT

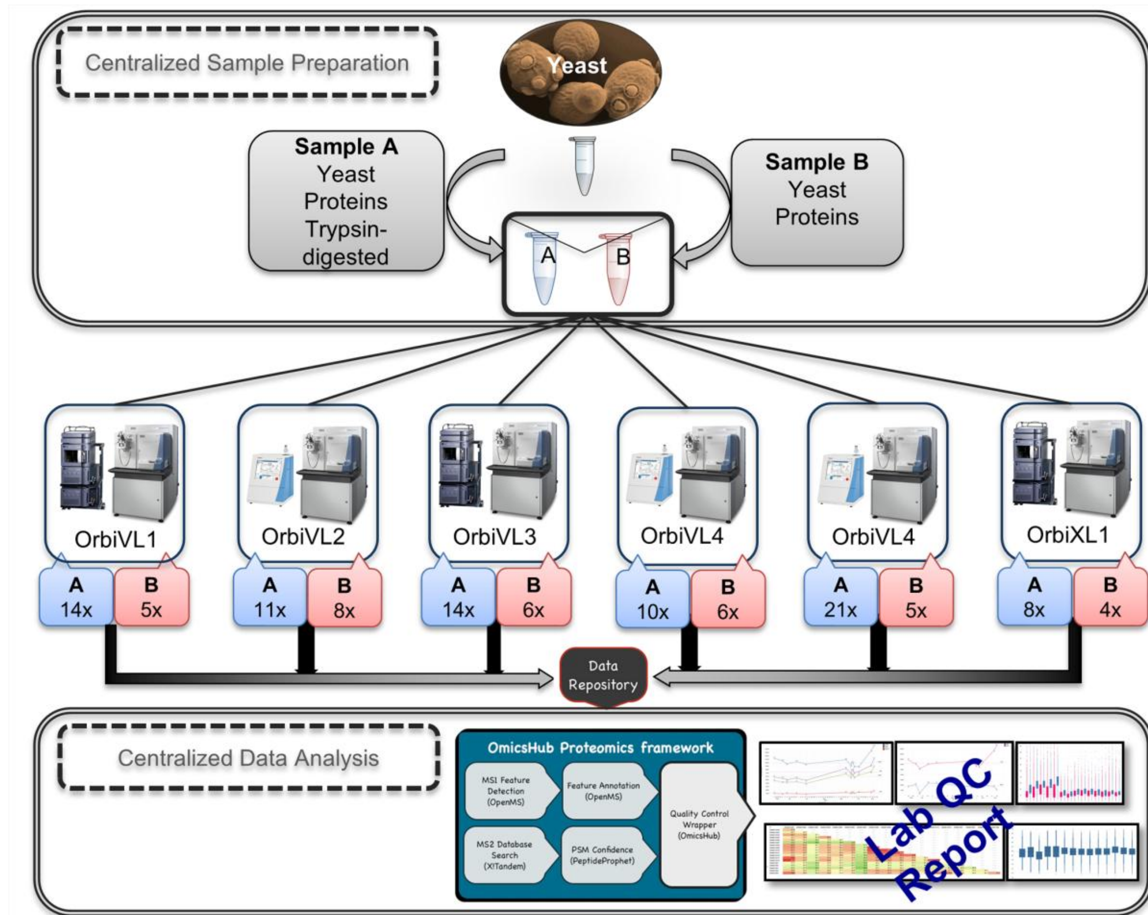


	OrbiVL1		OrbiVL2		OrbiVL3		OrbiVL4		OrbiVL5		OrbiXL1	
	pMC1 (%)	pMC2 (%)	pMC1 (%)	pMC2 (%)	pMC1 (%)	pMC2 (%)	pMC1 (%)	pMC2 (%)	pMC1 (%)	pMC2 (%)	pMC1 (%)	pMC2 (%)
Count	5	5	8	8	6	6	6	3	5	5	4	4
Q1	11.1419	0.749712	6.75553	0.183865	15.8569	1.56801	0.774265	0.156491	4.27908	0.216038	30.338	5.11775
Median	12.1023	0.797752	7.7985	0.302613	17.5491	1.80808	2.54162	0.157835	4.28263	0.322425	32.4076	5.50234
Q3	12.3588	0.892761	11.0616	0.46037	18.6103	2.0327	4.12848	0.25895	4.55425	0.409476	34.4281	6.62838
StdDev	1.51209	0.170222	5.18052	0.500594	2.72216	0.569872	2.07995	0.117542	0.340006	0.108369	2.52389	2.52566

Figure 6

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



Graphical abstract

ACCEPTED

ACCEPTED MANUSCRIPT

Conflict of interest

The authors have no financial conflicts of interest around the publication of this manuscript.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights

1) The LC-MS/MS performance of 6 laboratories was evaluated during 6 months.
2) QC metrics were evaluated to check performance of shotgun proteomics workflows.
3) Protein identification repeatability overtime is overall fair, IQR: 69.2 – 72.4%.
4) Lack of systematic LC-MS quality assessment can drop identifications down to 30%.

ACCEPTED MANUSCRIPT