PEAKS Online: Automated Quality Control (QC) Tool

Yiming Xiao, PhD, Application Scientist
Kyle Hoffman, PhD, Applications Manager

Bioinformatics Solutions Inc., Waterloo, Canada

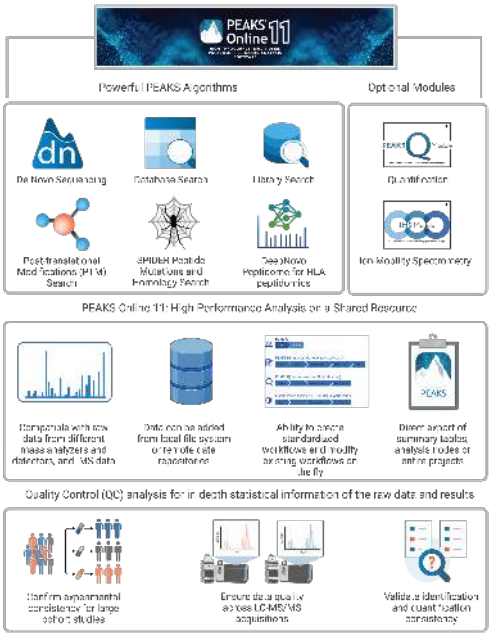
Aim:

To demonstrate the utility of the Automated Quality Control (QC) tool PEAKS Online 11.

Introduction:

LC-MS based protein/peptide quantification has drawn attention for physiopathology/pharmaceutical studies. For such research, especially clinical studies, obtaining a large number of samples is necessary to provide sufficient statistical power, i.e., over 50 samples per group are often necessary due to the high interindividual variation [1]. For cohorts with large sample sizes, performing a QC step is essential for ensuring data quality and validity of the results.

PEAKS Online has a automated QC tool that provides sophisticated QC analysis on top of protein/peptide identification and quantification results. All the statistical analyses are user-friendly and can be specified based on the user's requirements.



Case study: benchmarking data set analysis with PEAK Online:

A published data set [2] was used as an example to demonstrate an application of LFQ and QC analysis in PEAKS Online.

Study aims and background:

Label-free quantification using DDA and DIA (SWATH) can be compared to determine which acquisition method leads to more reproducible and accurate results. DDA is the more traditional way but also suffers from high missing values and under-sampling. DIA is more robust and can resolve DDA-related biases, however, DIA also has issues such as high false positive rate in complex MS2 spectra. In this study, the same samples were analyzed by high-resolution DDA or DIA (SWATH) and compared. When performing MS1 ion current extraction, high resolution DDA (HS-DDA) quantification was comparable to DIA in accuracy, precision, and better for lower abundance proteins. In this application note, the DDA data is used for presenting automated QC tool in PEAKS Online, to highlight attributes that fail in the analysis.

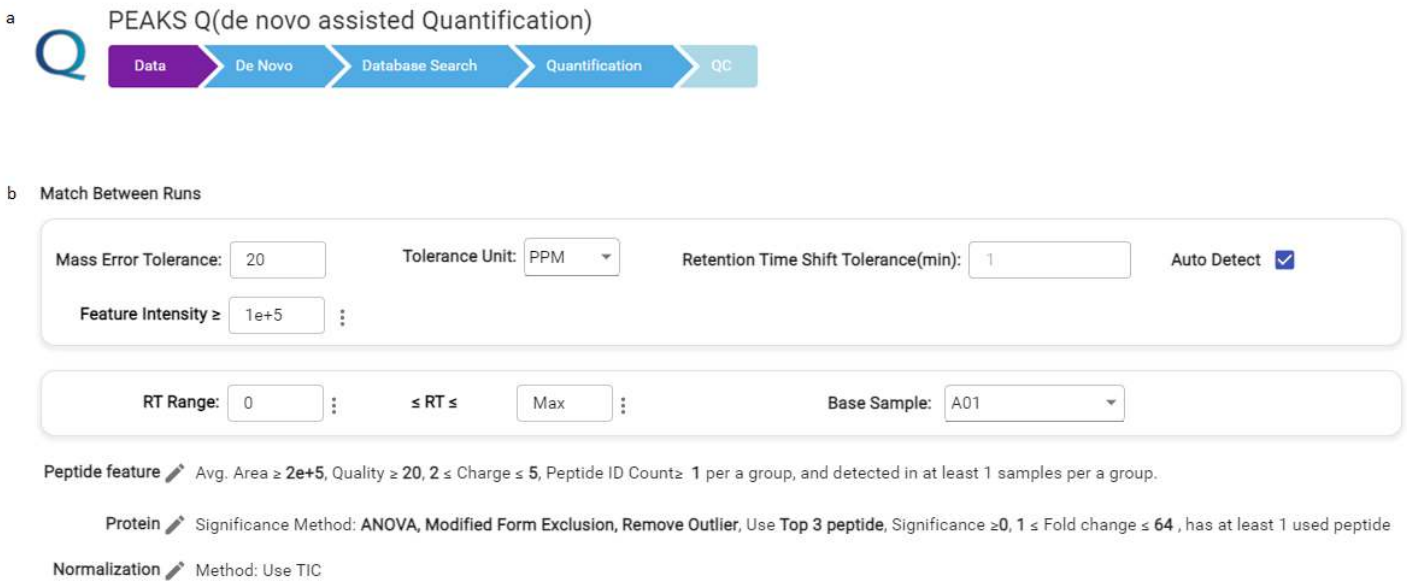


Fig 1. Protein/peptide quantification in PEAKS Online. (a) PEAKS Q workflow integrated with QC function. (b) LFQ parameters used.

Methods:

Five groups of samples were prepared with three different proteomes (human, E. coli and yeast) and run with five technical replicates (n=25). The human protein proportion was 60% across all samples. The portion of E. coli to yeast protein amounts were as follows:

A: 5%/35%, B: 7.5%/32.5%, C: 10%/30%, D: 15%/25%, E: 20%/20%.

DDA data was acquired with MS1 240000 resolution and MS2 15000 resolution. The gradient was 160 minutes long. A detail of LC-MS method can be found in [2].

Results:

MS data (5 runs *5 samples) was analyzed in PEAKS Online as LFQ with PEAKS Q module (Figure 1a). For the quantification part, match-between-run and TIC normalization were applied. The detailed search parameters are shown in Figure 1b.

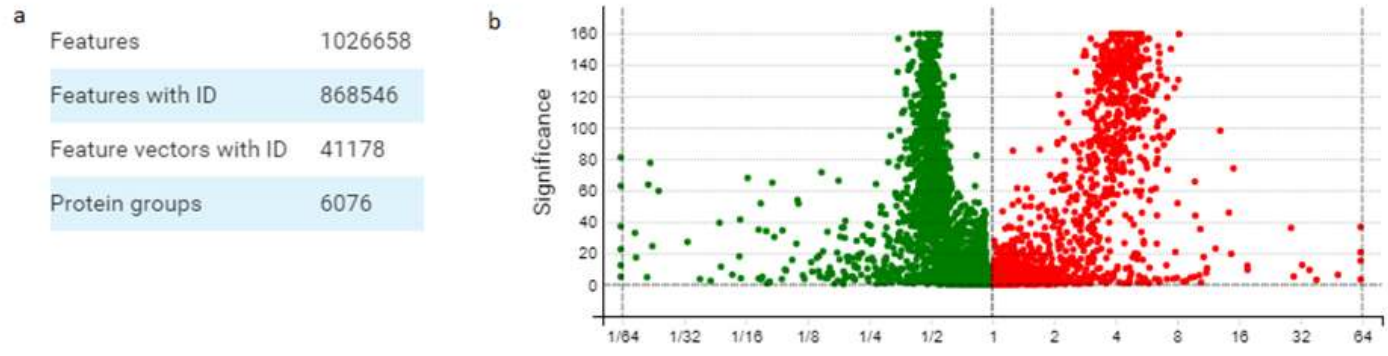


Fig 2. LFQ result. (a) Number of identified attributes. (b) The volcano plot for differentially expressed proteins.

While using at least two peptides per protein for quantification, 6076 protein groups are quantifiable. The number of features, features with identifications, and protein groups are listed in Figure 2a. The volcano plot shows the differential expression of proteins across all samples (Figure 2b).

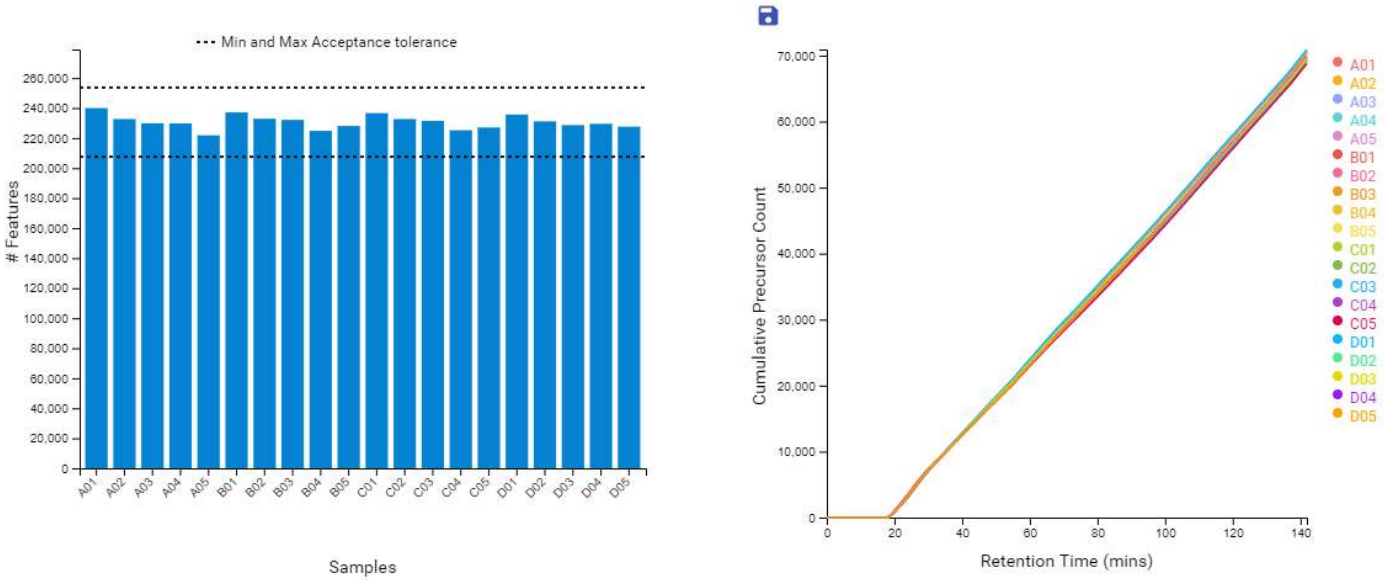


Fig 3. (a) Number of features found in each sample. (b) Cumulative precursor count plotted against retention time.

Results cont'd:

QC result of LC-MS data

In QC analysis, the sample average was set as the reference, and the acceptance tolerance was set as 10%. Any attribute that falls outside of the 10% acceptance tolerance is labelled in red (fail) in the QC result views (Table 1).

The QC result shows the number of MS1, MS/MS, features, MS2/MS1 ratio, full peak width, full peak width at half maximum (FWHM), and total base peak chromatogram (BPC) intensity. The data QC result shows that out of all 25 samples (E01-E05 are not shown in Table 1) and 7 different metric categories, only B04 and B05 have a lower BPC intensity compared to the average value (12.7% and 11.6%), causing this attribute to fail and fall outside of the 10% acceptance tolerance.

1 - 21 of 21

SAMPLE NAME	# MS1	# MS/MS	MS2/MS1 RATE	# FEATURES	FULL WIDTH (MIN)	FWHM (SEC)	TOTAL BPC
Average	5773	72711	12.60	230574	0.31	11.21	3.882e+9
A01	5641	73581	13.04	239956	0.31	11.31	4.022e+9
A02	5882	71706	12.19	232641	0.31	11.23	3.754e+9
A03	5874	71802	12.22	229834	0.31	11.22	3.890e+9
A04	5825	72155	12.39	229729	0.31	11.21	3.737e+9
A05	5871	71830	12.23	221794	0.31	11.18	3.664e+9
B01	5684	73493	12.93	237093	0.31	11.20	4.060e+9
B02	5792	72458	12.51	232858	0.31	11.20	3.867e+9
B03	5799	72375	12.48	232093	0.31	11.20	3.718e+9
B04	5778	72370	12.53	224801	0.31	11.19	3.389e+9
B05	5807	71887	12.38	228044	0.31	11.19	3.431e+9
C01	5680	73690	12.97	236578	0.31	11.20	4.172e+9
C02	5787	72654	12.55	232673	0.31	11.20	3.960e+9
C03	5834	72219	12.38	231485	0.31	11.20	3.824e+9
C04	5834	72473	12.42	225205	0.31	11.20	3.782e+9
C05	5863	71803	12.25	227010	0.31	11.19	3.691e+9
D01	5661	73921	13.06	235646	0.31	11.20	4.235e+9
D02	5819	72415	12.44	231106	0.31	11.20	4.034e+9
D03	5783	72649	12.56	228636	0.31	11.20	3.899e+9
D04	5724	72882	12.73	229480	0.31	11.21	3.696e+9
D05	5775	72734	12.59	227576	0.31	11.21	3.845e+9

Table 1. LC-MS Data QC Result

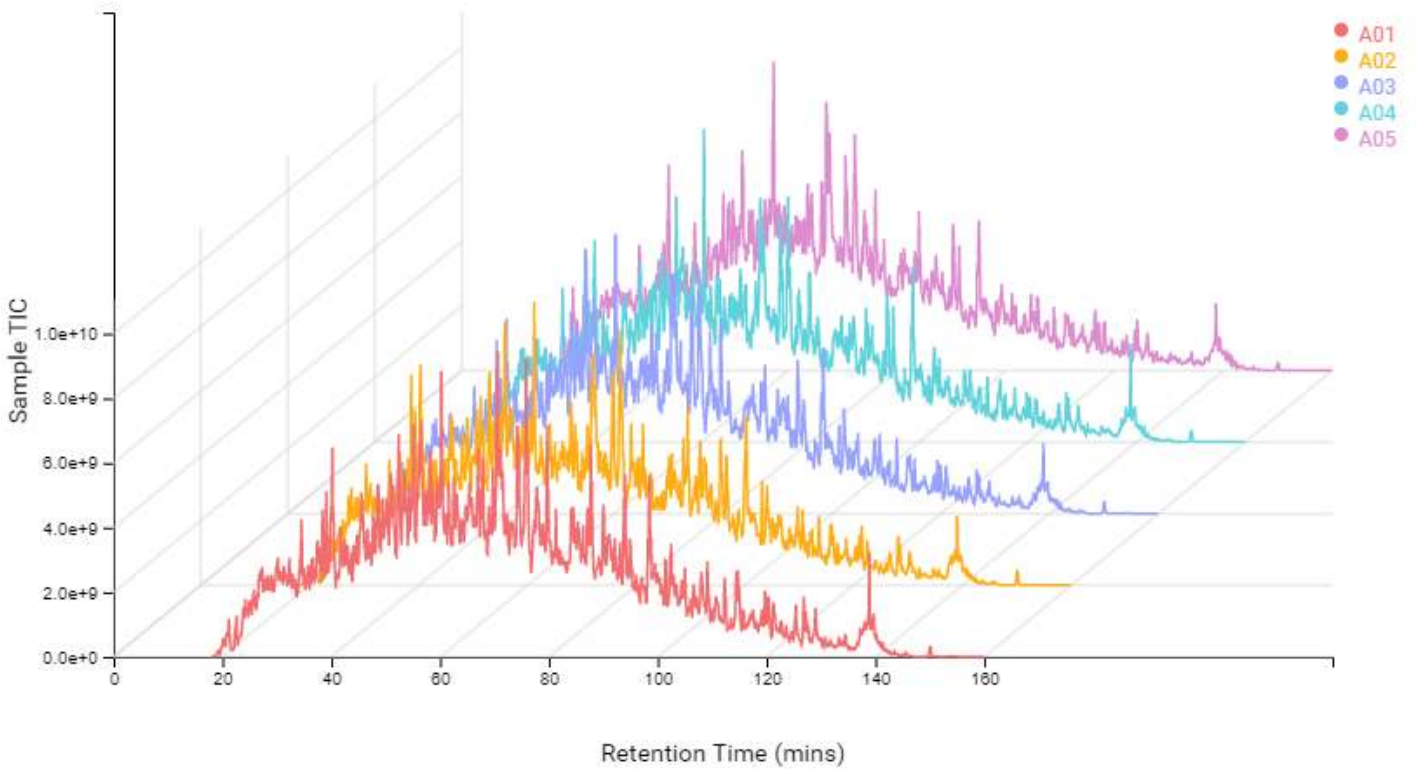


Fig 4. Sample TIC over retention time for samples A01 to A05.

The number of features (highlighted in blue in Table 1) in each sample is also presented in a bar chart with min and max tolerances (Figure 3a). The cumulative precursor count is shown in Figure 3b. A near linearity increment of precursor across retention time suggests the peptides were well separated by the LC gradient.

The Chromatogram of each sample (up to 5) overlay in Figure 4, provides an overview of total ion chromatograms (TIC) from the selected LC-MS runs. This visualization allows the user to assess differences in the chromatograms across sample runs.

1 - 21 of 21

SAMPLE NAME	# MS/MS	# IDENTIFIED FEATURES	# PSM	# IDENTIFIED PEPTIDES	# QUANTIFIED PEPTIDES	# SEQUENCES	# IDENTIFIED PROTEIN GROUPS	# QUANTIFIED PROTEIN GROUPS	# TOP PROTEINS	# ALL PROTEINS	# SCANS / # PEPTIDE RATE	ID RATE
Average	72711	47774	59700	42157	36387	41565	7049	6064	7139	7493	1.22	82.10%
A01	73581	50751	62123	44946	36424	44402	7088	6066	7179	7564	1.18	84.43%
A02	71706	46906	58171	41330	36229	40688	6794	6049	6881	7219	1.23	81.12%
A03	71802	46686	58623	41474	36386	40873	6908	6067	6994	7349	1.22	81.65%
A04	72155	47691	59931	42354	36400	41754	7067	6065	7160	7520	1.20	83.06%
A05	71830	45214	58250	39899	36236	39415	6793	6053	6885	7223	1.23	81.09%
B01	73493	49832	61447	44158	36438	43544	7159	6072	7246	7621	1.20	83.61%
B02	72458	48131	59619	42513	36356	41845	7089	6062	7180	7530	1.22	82.28%
B03	72375	47061	58603	41664	36443	41066	6938	6071	7016	7356	1.24	80.97%
B04	72370	45055	57661	39567	36275	39050	6842	6057	6926	7248	1.26	79.68%
B05	71887	46539	58916	41184	36368	40629	6924	6057	7012	7378	1.22	81.96%
C01	73690	49720	61302	43866	36412	43217	7142	6062	7232	7598	1.20	83.19%
C02	72654	47421	58909	41880	36418	41267	7044	6065	7131	7479	1.23	81.08%
C03	72219	47622	59532	42228	36441	41617	7101	6072	7182	7537	1.21	82.43%
C04	72473	47192	59254	41721	36294	41172	7190	6062	7275	7632	1.22	81.76%
C05	71803	46176	58616	40828	36413	40269	6908	6068	6986	7331	1.22	81.63%
D01	73921	49407	60493	43440	36413	42828	7205	6069	7311	7688	1.22	81.83%
D02	72415	48191	60110	42553	36415	41941	7131	6061	7218	7579	1.20	83.01%
D03	72649	47558	59381	42138	36420	41541	7120	6072	7208	7558	1.22	81.74%
D04	72882	47452	59553	41690	36438	41122	7109	6072	7206	7561	1.22	81.71%
D05	72734	47235	59792	41503	36427	40917	7014	6066	7099	7459	1.22	82.21%

Table 2. LFQ QC Result.

QC result of LFQ

Similar to the QC result of LC-MS data, a table is provided for the LFQ QC result (Table 2). The highlighted part (identified and quantified protein groups) is also presented in a bar chart (Figure 5a). The min and max acceptance tolerance is calculated based on the average number of quantified protein groups. The number of proteins with the corresponding covariance values (CV) are shown in Figure 5b. The CV value here reflects the variance in a protein's abundance across all samples.

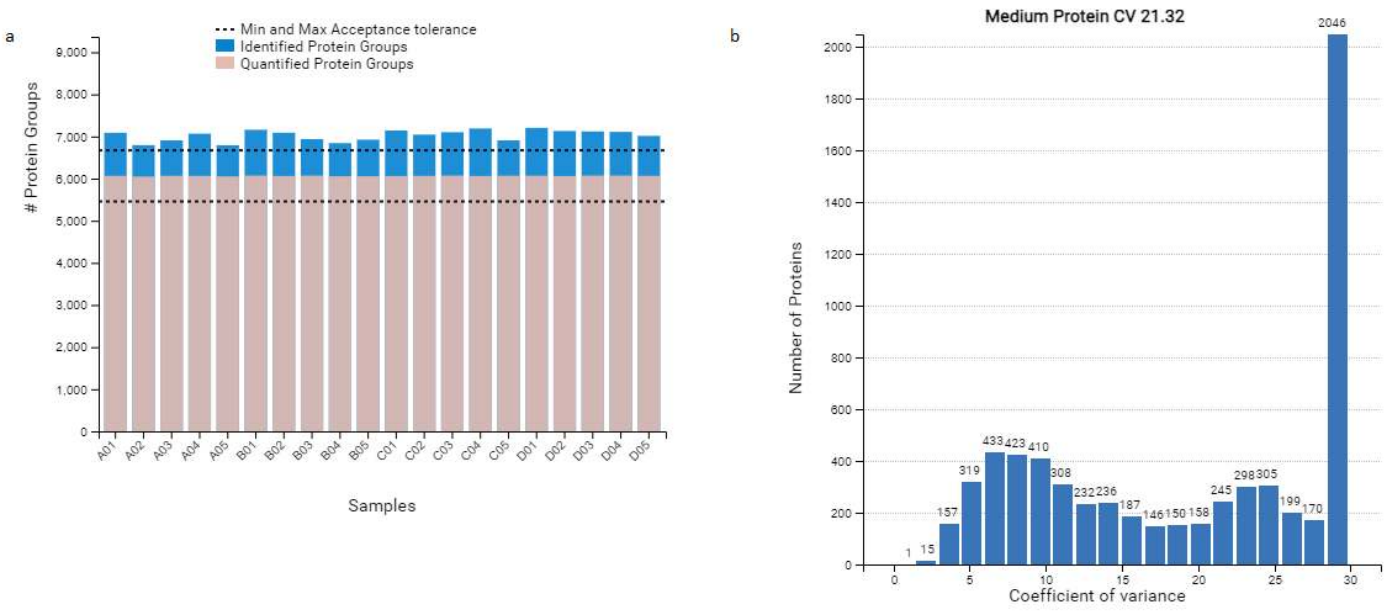


Fig 5. Number of identified and quantified protein groups and covariance values. (a) Identified and quantified protein groups after match between runs. (b) Number of proteins with corresponding CV values.

The peptide/protein identification reproducibility figures (Venn diagrams) shown in Figure 6 allow the user to quickly determine how many common or unique peptides/proteins are identified between samples. The Pearson correlation charts of samples shown in Figure 7 displays the correlation of peptide or protein intensities between the two selected samples. The user could pick any pair of samples to perform such examinations.

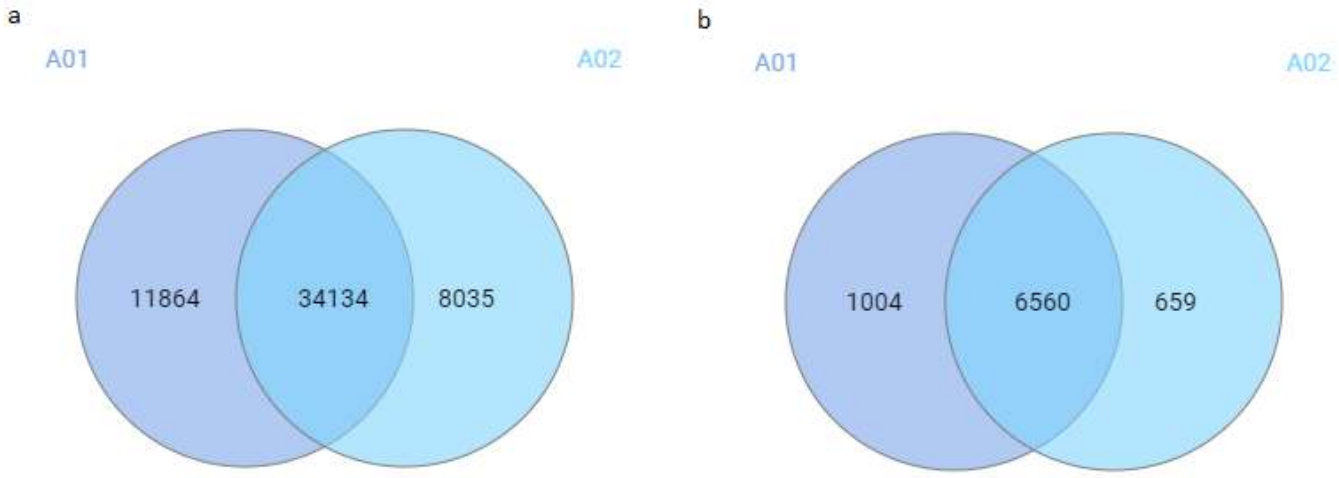


Fig 6. Sample reproducibility between A01 and A02. (a) Number of peptides unique or common in samples A01 and A02. (b) Number of proteins unique or common in samples A01 and A02.

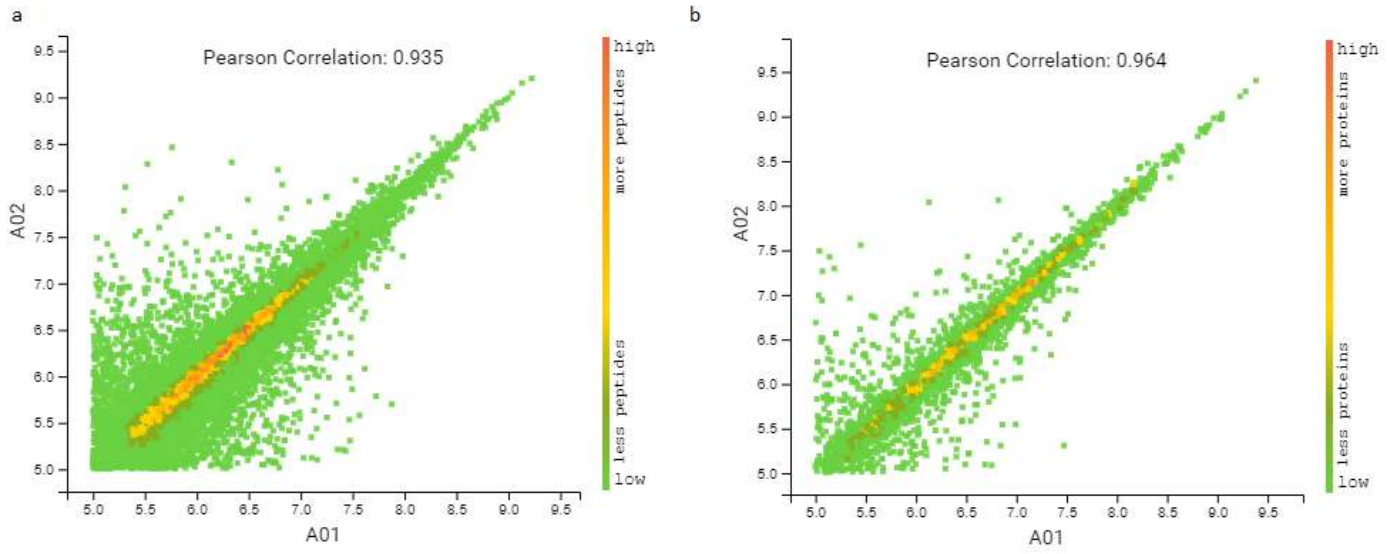


Fig 7. Pearson correlation of A01 and A02. (a) Correlation of peptides in samples A01 and A02. (b) Correlation of proteins in samples A01 and A02.

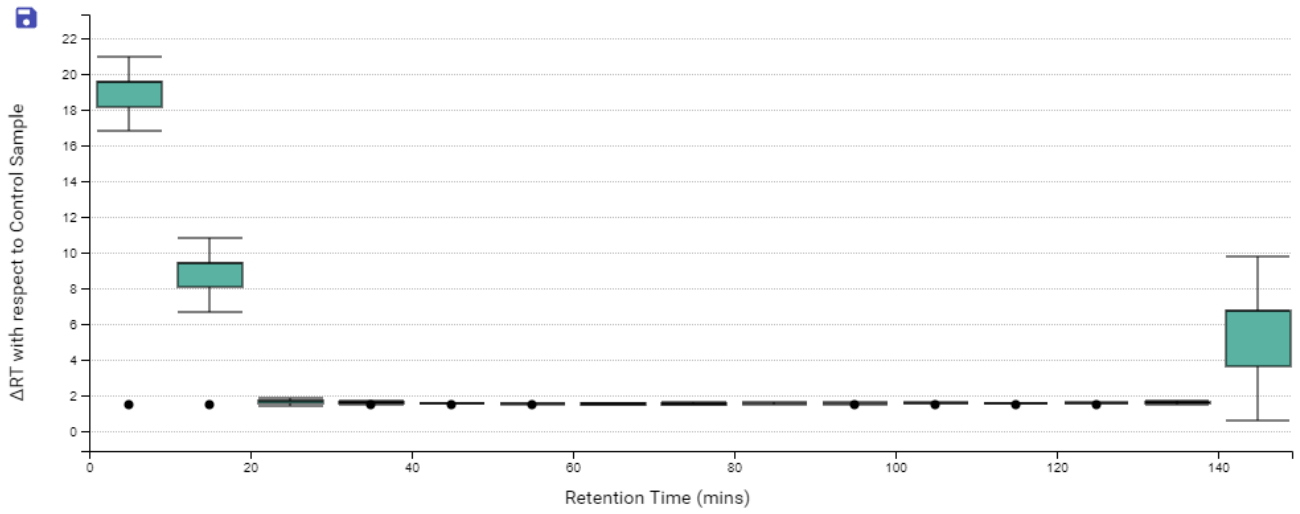


Fig 8. Box plot of the difference in the retention times of peptides from all samples across the LC gradient, with respect to a specified control sample.

Conclusion:

The new version of PEAKS Online integrates an automated QC tool for protein/peptide identification and quantification. By coupling QC information with data analysis, PEAKS Online can help users efficiently validate sample data in large cohorts and provide information for any potential troubleshooting.

References:

1. Overmyer, K. A. et al. Large-scale multi-omic analysis of COVID-19 severity. Cell Syst. 12, 23–40 (2021).
2. Wang, X. et al. Ultra-High-Resolution IonStar Strategy Enhancing Accuracy and Precision of MS1-Based Proteomics and an Extensive Comparison with State-of-the-Art SWATH-MS in Large-Cohort Quantification. Analytical Chemistry. 11, 4884-4893 (2021).

Bioinformatics Solutions, Inc.

140 Columbia St, Suite 202
Waterloo, Ontario N2L 3K8
Canada

Tel: (519) 885-8288
Fax: (519) 885-9075

sales@bioinfor.com
www.bioinfor.com