Label-free Quantification with PEAKS

Wen Zhang, Ziaur Rahman, Yifang Chen, Lei Xin, and Baozhen Shan
Bioinformatics Solutions Inc., Waterloo, Canada

Aims

Label-free quantification (LFQ) data analysis

Summary:

A published dataset [1] was used to illustrate LFQ functions and performance in PEAKS. PEAKS8.5 provides samples-based perspective view of protein and peptide MS1 peak area, MS2 spectral counts and protein coverage information. Furthermore, the LFQ algorithm implemented in PEAKS Q achieved higher accuracy and smaller variance compared to MaxQuant.

Introduction

➢ Label-free Quantification

Protein abundances across a large number of samples can be compared via label-free quantitative method. Compared to label-based approaches, label-free quantification workflows are simpler, since no isotopic labeling is involved. In addition, less complicated data analysis is required. However, to achieve accurate and precise quantification, experimental consistencies in sample handling, liquid chromatograph (LC) separation and mass spectrometry (MS) performance are particularly crucial.

➢ Proteome Quantification with PEAKS

PEAKS is a data analysis tool for bottom-up quantitative proteomics. Specifically, the add-on PEAKS Q module supports quantitative analysis of both label-free and label-based, e.g. SILAC, TMT, iTRAQ, et al. data. For LFQ, two major approaches are commonly used: spectral counting and MS1 signal intensities. PEAKS can analyze data generated by either method, with a particular emphasis on the later one.

In PEAKS DB searching result, MS2 spectrum counts acquired for each protein or peptide are displayed to indicate its abundance in each individual sample. In addition, sample-based peptide and protein peak area and protein coverage are also provided. These newly-added information into PEAKS8.5 aims to provide users a more comprehensive view of the data in each individual sample.

In PEAKS Q module, protein relative quantification is performed based on the extracted ion chromatograms of the whole isotopic envelope on MS1 level. PEAKS LFQ algorithm also extracts and uses the LC retention time and MS features to align different runs [2], thus transferring IDs to matched features without IDs to maximally retrieve quantification information and improve results.

Methods

A published CPTAC dataset [1] was used as an example to illustrate LFQ performance in PEAKS. Details of this dataset was summarized as below.

➢ Study Aims

The CPTAC Study 6 OrbiO@65 dataset [1] was used in this study to show LFQ performance in PEAKS and the results were compared to the MaxQuant LFQ outputs.

➢ Experimental Design

CPTAC Study 6 included 5 groups of mixtures of yeast proteins with spiked-in Sigma UPS1 proteins at increasing concentrations shown in Table 1. Each group had 3 technical replicates. These samples were run on different mass spectrometers and the OrbiO@65 dataset generated on a LTQ-Orbitrap instrument was used in this study.

Table 1. Sample compositions [1].

Sample Mixture	Yeast Protein	UPS1 Human Protein
A (× 3 replicates)	60 ng/µL	0.25 fmol/µL
B (× 3 replicates)	60 ng/µL	0.74 fmol/µL
C (× 3 replicates)	60 ng/µL	2.2 fmol/µL
D (× 3 replicates)	60 ng/µL	6.7 fmol/µL
E (× 3 replicates)	60 ng/µL	20 fmol/µL

➢ Data analysis

OrbiO@65 data were searched against a database comprised of SwissProt human and *Saccharomyces cerevisiae* sequences (v2017_05) in PEAKS Studio 8.5 [3] and MaxQuant 1.5.3.30 [4]. Carbamidomethylation was set as a fixed modification. Oxidation (M) and pyro-glu (from Q or E) were added as variable modifications as suggested by the original publication [1]. A precursor mass tolerance of 15 ppm and a fragment mass tolerance of 0.5 Da were applied. Semi-tryptic peptides were included and a maximum of 3 variable modifications was used. 1% FDR was applied at peptide-spectrum match (PSM) level. Normalized LFQ intensities from PEAKS and MaxQuant were used for further data analyses.

Results and Discussion

➢ Sample-based View in PEAKS8.5 DB Searching Result

- View of protein MS1 area, MS2 spectral counts (#Spec) and coverage in each sample

Accession	Coverage	Area D-1	Area D-2	Area D-3	#Spec D-1	#Spec D-2	#Spec D-3	Coverage D-1	Coverage D-2	Coverage D-3	#Unique	#Peptides
Proteins												
P02787|TRFE_HUMAN	64%	1.19E7	2.99E7	1.38E7	43	47	34	42%	43%	35%	66	73
P10636|TAU_HUMAN	45%	1.56E7	3.27E7	2E7	42	36	27	32%	28%	27%	50	54
P06732|KCRM_HUMAN	59%	4.82E6	8.89E6	7.5E6	19	19	19	44%	40%	40%	33	35
P04040|CATA_HUMAN	59%	8.25E6	2.35E7	1.3E7	32	41	35	38%	38%	38%	58	60
P12081|SYHC_HUMAN	59%	2.24E6	5.61E6	4.7E6	22	24	23	25%	30%	27%	34	45

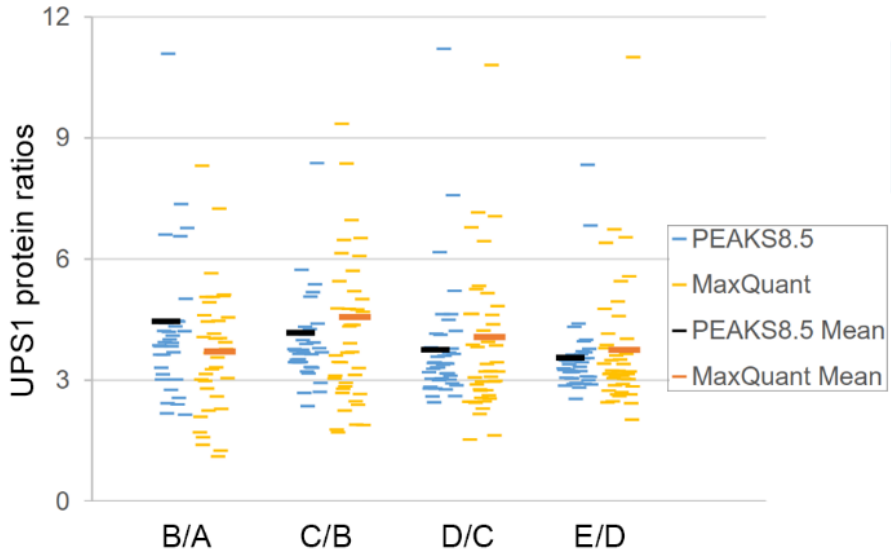
Figure 1. Sample-based proteome data in PEAKS8.5 DB searching result

➢ Higher Accuracy of PEAKS8.5 LFQ Algorithm

- Top 3 peptides selected for LFQ
- Retention time alignment to assign IDs to matched features without MS2 (identified in one, quantified in all samples)

Sigma UPS1 protein mixtures were spiked in at increasing amounts, i.e. 3 fold more, from group A to E. The fold changes of quantified UPS1 proteins reported by PEAKS8.5 and MaxQuant between the neighboring groups were shown in Figure 2. The UPS1 protein ratios in each group were calculated from the average ratios of the three replicates. At lower

amounts, i.e. between group A and B, MaxQuant gave protein ratios closer to 3 than PEAKS. That is likely because MaxLFQ uses median peptide ratios for protein ratio calculations, thus eliminating outliers that are more often observed for low abundant proteins [5]. However, PEAKS gave protein ratios closer to 3 than MaxQuant in the other 3 comparisons (Figure 2). Furthermore, the coefficient of variation (CV) in general was smaller with PEAKS LFQ.



UPS1 protein ratios with PEAKS LFQ				
Groups	B/A	C/B	D/C	E/D
Mean	4.46	4.18	3.76	3.56
CV	50.07	48.82	40.59	28.21

UPS1 protein ratios with MaxLFQ				
Groups	B/A	C/B	D/C	E/D
Mean	3.72	4.56	4.07	3.75
CV	41.92	69.46	57.79	42.12

Mean: average of all quantified UPS1 protein ratios between the specified 2 groups.

CV: CV of all quantified UPS1 protein ratios between the specified 2 groups.

Figure 2. Ratios of UPS1 protein levels between groups from LFQ normalized intensities in PEAKS and MaxQuant.

Conclusions

PEAKS8.5 provides samples-based perspective view of protein and peptide MS1 peak area, MS2 spectral counts and protein coverage information. Furthermore, the LFQ algorithm implemented in PEAKS Q achieved higher accuracy and smaller variance compared to MaxQuant.

Reference:

[1] Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., *et al.*, Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of proteome research* 2010, *9*, 761-776.
[2] Lin, H., He, L., Ma, B., A combinatorial approach to the peptide feature matching problem for label-free quantification. *Bioinformatics* 2013, *29*, 1768-1775.
[3] Zhang, J., Xin, L., Shan, B., Chen, W., *et al.*, PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & cellular proteomics : MCP* 2012, *11*, M111 010587.
[4] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 2008, *26*, 1367-1372.
[5] Cox, J., Hein, M. Y., Luber, C. A., Paron, I., *et al.*, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* 2014, *13*, 2513-2526.