

Statistical analysis of multiple iTRAQ/TMT experiments labeled with variable reporter ion tags using ProteoIQ

Ningombam Sanjib Meitei¹, Arun Apte², Arijit Biswas¹, Mark Pitman², Parimal Samir³ and Andrew J. Link^{3,4}

¹PREMIER Biosoft, Indore, India, ²PREMIER Biosoft, Palo Alto, CA, USA, ³Dept of Biochemistry, ⁴ Dept of Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, TN, USA

Corresponding author e-mail: sanjib@premierbiosoft.com



1. Abstract

Isobaric tagging of peptides for reporter ion based quantitation are powerful techniques for mass spectrometry-based quantitative proteomics to analyze differential protein expression [1], [2]. Isobaric reagents are used to label different protein samples with isobaric tags. Samples are then pooled for LC-MS/MS analysis. The majority of the existing bioinformatics tools support statistical analysis of proteins from a single iTRAQ/TMT experiment. However, improved data analysis approaches are required to obtain accurate estimates of biological effects and to increase the power to detect meaningful differences from multiple iTRAQ experiments and large sample sizes [3]. We have developed software modules to perform statistical analysis on iTRAQ data collected from complex proteomics samples across multiple experiments and samples. Different normalization techniques to control for adverse effects of inter- and intra- experimental variability were developed and tested. We demonstrate the utility of the program using differential protein expression analysis of complex myotonic dystrophy in mouse and human patient muscle samples. Data were collected from multiple iTRAQ experiments comparing control and disease muscle samples for both myotonic dystrophy type 1 (DM1) and myotonic dystrophy type 2 diseases. ANOVA was performed to identify proteins that are significantly different across the control and disease groups.

2. Methods

2.1 Sample preparations and design of experiments

Cultured myoblasts (used as control) and skeletal muscle tissues were homogenized in lysis buffer (50mM HEPES, 50% Trifluoroethanol) using ultrasonication. Total protein was quantified using BCA assay (Pierce). Ten µg of protein each from the samples were digested with sequencing grade trypsin (Promega), desalted, labeled with 8plex iTRAQ and pooled according to scheme depicted in figure 1. The description of the samples are as follows.

• All human samples were analyzed by 8plex iTRAQ

• The following codebook explains the naming system of the samples:

- Prefix C: Control
- Prefix DM1: Myotonic dystrophy type 1
- Prefix DM2: Myotonic dystrophy type 2
- The letter F following C, DM1 or DM2: Female
- The letter M following C, DM1 or DM2: Male
- The number following F or M: Patient number

• For example:

- DM1M1- Myotonic dystrophy type 1 patient, male, patient number 1
- DM2F1- Myotonic dystrophy type 1 patient, female, patient number 1
- CM1- Control, male, subject number 1

• There is one sample labeled 'UND'. This sample was from patient suffering from an unrelated neurological disease.

• There are two control channels in DM2 Set1 and DM2 Set2. There is one control channel in DM1

2.2 MudPIT Methods

Pooled iTRAQ labeled samples were analyzed using MudPIT using an Eksigent 2-D nanoLC pump coupled to a nanoESI-LTQ-OrbitrapXL mass spectrometer (Thermo Scientific) as described in Hoek *et al* [5]. The precursor ions were analyzed in the Orbitrap followed by 4 collision induced dissociation (CID) fragment ion scans in the ion trap to identify peptides. The precursor ions were then fragmented by higher-energy collisional dissociation (HCD) to measure reporter ion intensities in the Orbitrap. For each precursor ion, the CID and HCD spectra were merged using Proteome Discoverer v1.3 (Thermo Scientific).

2.3 ProteoIQ software workflows

2.3.1 ProteoIQ schema for combining multiple iTRAQ/TMT experiments

In the iTRAQ/TMT method for relative quantitation of biological samples, each sample is labeled independently with isobaric reporter ion tag and then pooled for MS/MS analysis. In many situations, multiple iTRAQ experiments are combined to obtain accurate estimates of biological effects and to increase the power to detect meaningful differences. Figure 1 shows the schematic representation of ProteoIQ software facilitating combination of multiple iTRAQ/TMT experiments.

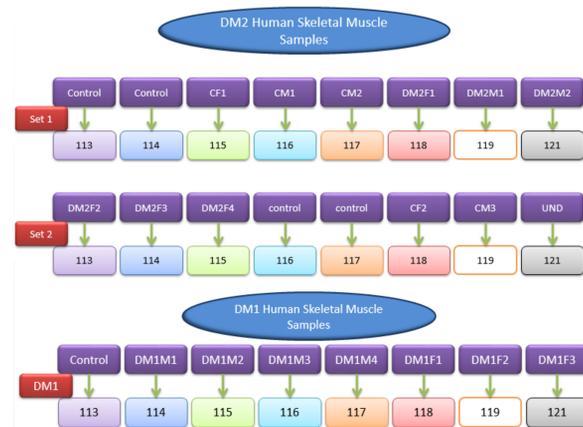


Figure 1: Design of experiments being employed for the 3 iTRAQ experiments.

| Reporter Ion Tags | | | | | | | | |
|-------------------|------|------|------|------|-----|------|-----|-----|
| Experiment No. | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 121 |
| 1 | Y11 | Y12 | Y13* | Y14 | Y15 | Y16 | Y17 | Y18 |
| 2 | Y21 | Y22 | Y23 | Y24 | Y25 | Y26* | Y27 | Y28 |
| 3 | Y31* | Y32 | Y33 | Y34 | Y35 | Y36 | Y37 | Y38 |
| ... | | | | | | | | |
| i | Yi1 | Yi2* | Yi3 | Yi4 | Yi5 | Yi6 | Yi7 | Yi8 |
| m | Ym1 | Ym2 | Ym3 | Ym4* | Ym5 | Ym6 | Ym7 | Ym8 |

Table 1: Typical schematic representation of ProteoIQ software combining multiple iTRAQ/TMT experiments. iTRAQ 8 is used to describe the schema. Yij: Sample labeled with jth reporter ion tag from the experiment number 'i' where i = 1,2,3,...,m and j represents a particular reporter ion tag e.g., 113, 114, 115, 116, etc. Reference sample for each experiment number is represented by Yij with the suffix *.

Figure 2 shows the typical ProteoIQ 3.0 software graphical based user interfaces (user interface hereon) facilitating users to combine 3 experiments that includes DM2 set 1, DM2 set 2 and DM1 set 1.

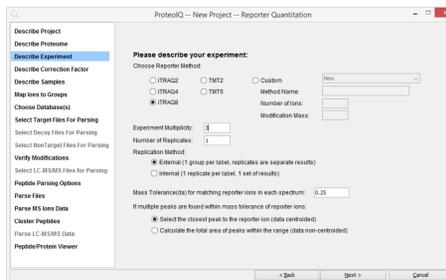


Figure 2(a): Typical ProteoIQ 3.0 user interface allowing users to specify how many experiments need to be combined. Here, we specify the value of Experiment Multiplicity as 3 to combine 2 experiments from DM2 patients and 1 experiment from DM1 patients.

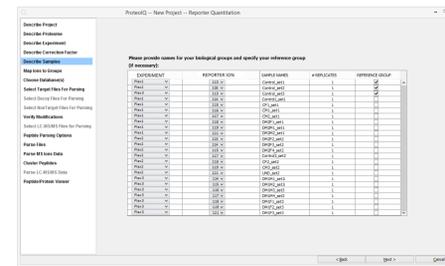


Figure 2(b): Typical ProteoIQ 3.0 user interface allowing users to specify reference samples for each of the experiments labeled with variable reporter ion tags.

2.3.2 Intra-experiment data normalization

Figure 3 shows the schematic representation on how proteins are measured using the ratio to reference sample (removing the variance in the reference channel), log2 normalization (ensuring magnitude independent) values and averaging across replicates (ensuring random observation).

2.3.3 Intra-experiment ANOVA workflow

ProteoIQ generates p-values for proteins based on quantitative information in samples and replicates. In ANOVA the F-statistic tests the significance of quantitation differences between samples against the reproducibility of their replicates. The F-statistic measures the ratio of explained error (difference across samples) to the unexplained error (difference across replicates). Typically p-values < 0.05 lead to the rejection of the null hypothesis and indicate a lower chance of quantitation differences between samples being due to random error.

2.3.4 Inter-experiment ANOVA workflow

Here, ProteoIQ generates p-values for proteins based on quantitative information in biological groups across experiments.

The F-statistic tests the significance of quantitation differences between biological groups. For example, we have 3 types of patients namely control i.e., healthy people, DM1 i.e., patients suffering from DM1 condition and DM2 i.e., patients suffering from DM2 condition. Typically p-value < 0.05 leads to the rejection of the null hypothesis and indicates significant quantitation differences between groups.

Figure 4 shows typical ProteoIQ software user interface that facilitates users to model experimental design i.e., define biological groups and assign samples from experiments as their replicates.

3. Results and Discussion

In this work MS/MS samples are loaded into ProteoIQ 3.0 after database searching is completed using the Sequest database search engine running under Proteome Discoverer [6,7]. To ensure the quality of the results, we select only the proteins and peptides that satisfy the following filters criteria.

- Proteins with at least 3 identified peptides
- 1% FDR threshold for proteins
- Peptides with peptide probability (calculated using Peptide Prophet algorithm) with predicted negatives < predicted positives
- Peptides with iTRAQ modifications
- Proteins with intra experiment ANOVA p-value less than 0.05

Table 2 shows the number of proteins and peptides for each of the experiment (before and after applying the filters).

| | Experiment 1 | | Experiment 2 | | Experiment 3 | |
|------------------------|---------------|--------------|---------------|--------------|---------------|--------------|
| | Before Filter | After Filter | Before Filter | After Filter | Before Filter | After Filter |
| Total no. of peptides | 16687 | 4562 | 20545 | 8393 | 22517 | 7116 |
| No. of unique peptides | 3022 | 1130 | 5184 | 2580 | 2378 | 735 |
| No. of protein groups | 449 | 169 | 637 | 252 | 225 | 56 |
| No. of proteins | 2937 | 1093 | 4124 | 1490 | 1449 | 375 |

Table 2: Experiment wise before- and after- filter statistics of peptides/proteins.

Inter-experiment ANOVA report

For further statistical analysis, the three experiments are combined and then modeled the design of experiment by defining three different groups based on the disease conditions namely Patients with DM1, Patients with DM2 and Control and specifying samples as their replicates (Table 3).

| # Replicates | Patients with DM1 | Patients with DM2 | Control |
|--------------|-------------------|-------------------|------------------|
| 1 | DM1M1 | DM2F1 | Control_114_Exp1 |
| 2 | DM1M2 | DM2M1 | CF1 |
| 3 | DM1M3 | DM2M2 | CM1 |
| 4 | DM1M4 | DM2F2 | CM2 |
| 5 | DM1F1 | DM2F3 | CF2 |
| 6 | DM1F2 | DM2F4 | CM3 |
| 7 | DM1F3 | | Control_117_exp2 |

Table 3: Design of experiment to test difference in protein quantitation across experiments.

On combining the three experiments, 1803 unique proteins are reported by the software. Out of these proteins, 235 proteins are observed at least in one of the samples of each experiment thereby facilitating the software to calculate inter-experiment ANOVA p-values. Table 5 shows the list of 63 proteins which are observed with p-value < 0.05.

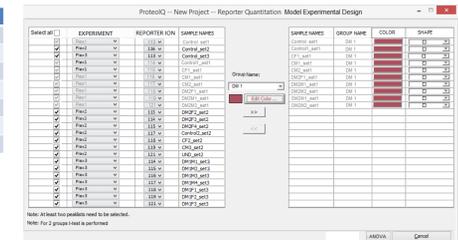


Figure 4: Typical ProteoIQ user interface showing the functionality to define groups and replicates thereby facilitating combination of proteins across multiple iTRAQ samples and perform ANOVA or any other statistical analysis.

| Sl. No. | Sequence Id | p-values | Sl. No. | Sequence Id | p-values | Sl. No. | Sequence Id | p-values |
|---------|-----------------|----------|---------|-----------------|----------|---------|-----------------|----------|
| 1 | ENSP00000224237 | 0.022635 | 22 | ENSP00000366962 | 0.002606 | 43 | ENSP00000406359 | 9.72E-05 |
| 2 | ENSP00000229239 | 0.003105 | 23 | ENSP00000367008 | 0.002606 | 44 | ENSP00000407461 | 0.002606 |
| 3 | ENSP00000244601 | 0.002606 | 24 | ENSP00000375391 | 9.72E-05 | 45 | ENSP00000408907 | 9.72E-05 |
| 4 | ENSP00000282050 | 0.033037 | 25 | ENSP00000375399 | 9.72E-05 | 46 | ENSP00000419425 | 0.016946 |
| 5 | ENSP00000289316 | 0.002606 | 26 | ENSP00000375736 | 0.002606 | 47 | ENSP00000422012 | 0.003143 |
| 6 | ENSP00000303408 | 0.002606 | 27 | ENSP00000380065 | 0.002875 | 48 | ENSP00000427976 | 0.023446 |
| 7 | ENSP00000310219 | 0.00427 | 28 | ENSP00000380067 | 0.003105 | 49 | ENSP00000430817 | 0.023446 |
| 8 | ENSP00000321744 | 0.002606 | 29 | ENSP00000380068 | 0.003105 | 50 | ENSP00000431702 | 0.02777 |
| 9 | ENSP00000324286 | 0.002606 | 30 | ENSP00000380070 | 0.003105 | 51 | ENSP00000435613 | 0.022626 |
| 10 | ENSP00000348706 | 0.002606 | 31 | ENSP00000380100 | 0.002806 | 52 | ENSP00000445633 | 0.002606 |
| 11 | ENSP00000348924 | 0.002606 | 32 | ENSP00000380177 | 0.002806 | 53 | ENSP00000445831 | 0.002606 |
| 12 | ENSP00000349430 | 0.002606 | 33 | ENSP00000380180 | 0.002606 | 54 | ENSP00000446007 | 0.022635 |
| 13 | ENSP00000350580 | 0.002606 | 34 | ENSP00000381736 | 0.033037 | 55 | ENSP00000446031 | 0.002606 |
| 14 | ENSP00000352442 | 0.002606 | 35 | ENSP00000382915 | 9.72E-05 | 56 | ENSP00000462132 | 0.002606 |
| 15 | ENSP00000353074 | 0.002606 | 36 | ENSP00000391842 | 0.026599 | 57 | ENSP00000463319 | 0.002606 |
| 16 | ENSP00000358151 | 0.002606 | 37 | ENSP00000393087 | 9.72E-05 | 58 | ENSP00000464073 | 0.002606 |
| 17 | ENSP00000358164 | 0.002606 | 38 | ENSP00000394670 | 0.018477 | 59 | ENSP00000464485 | 0.002606 |
| 18 | ENSP00000363071 | 0.002617 | 39 | ENSP00000401820 | 0.047961 | 60 | ENSP00000465477 | 0.033797 |
| 19 | ENSP00000364801 | 9.72E-05 | 40 | ENSP00000403530 | 9.72E-05 | 61 | ENSP00000467037 | 0.036325 |
| 20 | ENSP00000364802 | 9.72E-05 | 41 | ENSP00000404524 | 9.72E-05 | 62 | ENSP00000475942 | 0.002606 |
| 21 | ENSP00000366618 | 0.002606 | 42 | ENSP00000405975 | 0.018477 | 63 | ENSP00000476136 | 0.002606 |

Table 5: List of proteins with inter-experiment ANOVA p-values < 0.05

The difference in the protein quantitation levels are investigated using mouse models by comparing fold change in protein quantitation levels between HSA-CTG vs HSA-CCTG samples. 139 unique proteins are identified with fold change > 1.5 in Log₂ scale of normalized reporter ion intensity values (Data not shown).

Conclusion

ProteoIQ 3.0 allows users to perform statistical analysis of multiple iTRAQ/TMT experiments labeled with variable reporter ion tags. The above described functionality has also been extended to analyze isotopic label data. A free trial will soon be made available at <http://www.premierbiosoft.com>

Reference

- Ross et al. Mol Cell Proteomics. 2004; 3(12):1154-69.
- Thompson et al. Anal. Chem. 2003; 75 (8): 1895-904.
- Oberg et al. J Proteome Res. 2008; 7(1):225-33.
- Gerbaso et al. Molecular & Cellular Proteomics. 2007; 6(6): 1049-1058.
- Hoek, Kristen L., et al. PLoS one 10.2 2014; e0118528-e0118528.
- Eng JK et al. J Proteome Res. 2008; 7: 4598-4602.
- Eng JK et al. J Am Soc Mass Spectrom. 1994; 5: 976-989.