

Project 6115: BioID hyperlopit TMT expression

Daichi Shonai, Scott Soderling

Report for TMT Differential Protein Analysis by LC-MS/MS (27 Jan 2023)

Objective:

Examine the protein compositions from different centrifuged fractions (fraction4 ~ 10) of three independent study groups, each of which has 7 fractions from both WT and transgenic BioID labeled (Tg).

Proteomics Facility Scientific Contributors:

Duke Proteomics Core Facility: Tricia Ho (sample preparation, data collection, data analysis, report writing), Greg Waitt (data analysis), Erik J. Soderblom (study design, data collection, data analysis, report writing)

Sample Preparation. The Duke Proteomics and Metabolomics Core Facility (DPMCF) received 42 samples (3 of each WT_F4, WT_F5,... WT_F10 and Tg_F4, Tg_F5,... Tg_F10) which were kept at -80C until processing. Samples were supplemented to 3% SDS and were spiked with 1 or 2pmol casein as in internal quality control standard. They were reduced with 10 mM dithiothreitol for 30 min at 32C and alkylated with 20 mM iodoacetamide for 45 min at room temperature, then supplemented with a final concentration of 1.2% phosphoric acid and 437 μ L of S-Trap (Protifi) binding buffer (90% MeOH/100mM TEAB) (**Table 1**). Proteins were trapped on the S-Trap micro cartridge, digested using 20 ng/ μ L sequencing grade trypsin (Promega) for 1 hr at 47C, and eluted using 50 mM TEAB, followed by 0.2% FA, and lastly using 50% ACN/0.2% FA. All samples were then lyophilized to dryness.

TMT Labeling Procedure. Each sample was resuspended in 85 μ L 200 mM triethylammonium bicarbonate, pH 8.0 (TEAB). 65 μ L were taken of each sample and the remaining was combined to form an SPQC pooled sample. Fresh TMT16plex reagents (Thermo, 0.5 mg for each 16-plex reagent) was resuspended in 20 μ L 100% acetonitrile (ACN) and two sets were split among the three sample groups. Samples were incubated for 1 hour at RT. After 1-hour reaction, 3.25 μ L of 5% hydroxylamine was added and incubated for 15 minutes at room temperature to quench the reaction. Sample were combined then lyophilized to dryness.

Offline Fractionation Procedure. Each set of samples was subjected to Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (ThermoFisher Scientific, see https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FMSG%2Fmanuals%2F84868_highph_rp_peptidefract_UG.pdf). Fractions 1 and 7; Fractions 2 and 8 were combined, resulting in a total of 6 fractions per sample set. Samples were frozen and lyophilized to dryness.

LC-MS/MS Analysis. Quantitative LC/MS/MS was performed on 3 μ L (25%) for each fraction, using an MClass UPLC system (Waters Corp) coupled to a Thermo Fusion Lumos high resolution accurate mass tandem mass spectrometer (Thermo) equipped with a FAIMS ion-mobility device via a nanoelectrospray ionization source. Briefly, the sample was first trapped on a Symmetry C18 20 mm x 180 μ m trapping column (5 μ L/min at 99.9/0.1 v/v water/acetonitrile), after which the analytical separation was performed using a 1.8 μ m Acquity HSS T3 C18 75 μ m x 250 mm column (Waters Corp.) with a 90-min linear gradient of 5 to 30% acetonitrile with 0.1% formic acid at a flow rate of 400 nanoliters/minute (nL/min) with a column temperature of 55C. Data collection on the Fusion Lumos

mass spectrometer was performed in a data-dependent acquisition (DDA) mode with an $r=120,000$ (@ m/z 200) full MS scan from m/z 375 – 1575 with a target AGC target of 100% ions was performed at CVs -40v, -60v and -80v. MS/MS scans were acquired at $r=50,000$ with AGC target of 200% and max accumulation time of 120 ms. The total cycle time was 1s per CV voltage with total cycle times of 3 sec between like full MS scans. A 45s dynamic exclusion was employed to increase depth of coverage. The total analysis cycle time for each sample injection was approximately 2 hours.

Quantitative Data Analysis. Following UPLC-MS/MS analyses, data were imported into Proteome Discoverer 3.0 (Thermo Scientific Inc.) where quantitative signals correlating the TMT16 reporter ions (m/z 126, 127N, 127C, etc.) were extracted. In addition to quantitative signal extraction, the MS/MS data was searched against the SwissProt *M. musculus* database (downloaded in Aug 2022), a common contaminant/spiked protein database (bovine albumin, bovine casein, yeast ADH, etc.), and an equal number of reversed-sequence “decoys” for false discovery rate determination. Sequest database search algorithm was utilized to perform the database searches. Database search parameters included fixed modification on Cys (carbamidomethyl), Lys (TMT), peptide N-termini (TMT) and variable modification on Met (oxidation). Precursor mass tolerances were 2.0 ppm and product ion mass tolerances were 0.02 da with full trypsin enzyme rules required. Peptide Validator and Protein FDR Validator nodes in Proteome Discoverer were used to annotate the data at a maximum 1% protein false discovery rate based on q-value calculations. Note that peptide homology was addressed using razor rules in which a peptide matched to multiple different proteins was exclusively assigned to the protein has more identified peptides. Protein homology was addressed by grouping proteins that had the same set of peptides to account for their identification. A master protein within a group was assigned based on % coverage.

Raw peptide intensity values are presented in **Table 2**. Prior to imputation, a filter was employed which required that there were at least 2 measurements across all 42 samples for each peptide. Missing data points were imputed using a randomized intensity within the bottom 2% of the detectable signals. All peptides belonging to the same protein were then summed to the protein level and the datasets were subjected to a sample loading normalization. Following that initial normalization, the reference channel SPQC samples were used to normalize across the three individual TMT sets. Those normalized protein level data are shown in **Table 3**. Note that the total number of identified peptides within each TMT datasets is presented in the “peptides” column (number of peptides identified in set1, set 2, set 3).

Results

Depth of Coverage of LC-MS/MS analysis. Following database searching and peptide scoring using Proteome Discoverer validation, the data was annotated at a 1% protein false discovery rate. The overall coverage was **29,034 total peptides corresponding to 1,996 unique proteins**.

Measures of analytical versus biological variability. In order to assess technical reproducibility, we calculated % coefficient of variation (%CV) for each protein across the 6 SPQC pool channels that were interspersed throughout the study (**Table 3**). The mean %CV of the SPQC pools was **27.1%**. To assess biological + processing + technical variability, %CVs were measured for each protein which averaged **27.0%**.

Initial Statistical Analysis. As an initial statistical analysis, we calculated fold-changes between various groups based on the protein expression values and calculated two-tailed heteroscedastic t-test on log2-transformed data for each of these comparisons (**Tables 4-11**). The column “x_FC” in those tables indicated fold change relative to the second group in the header. For example, if there was a protein that said 1.9 for tg vs wt, then it was 1.9-fold higher in tg relative to wt. If a protein was different

DPMSR –Report –Proj 6115 – Shonai & Soderling – TMT16 WT v Tg

between the groups with a p-value < 0.05 then we indicated that in the “Stats” column. If it says “up” this indicates the protein was higher in tg relative to wt with a p-value of < 0.05.

Conclusions. We achieved a much higher depth of coverage in this experiment (~ 20,000 peptides, ~2,000 proteins) relative to the previous attempt. There were no other analytical changes on our end (same fractionation, MS acquisition, data analysis pipeline) so hopefully it’s just from a different bait providing a more robust labeling compared to the previous bait. Note that we did not require a fold-change cutoff for stat significance (just pvalue <0.05). to increase stringency, you may want to consider a fold-change of 1.3 as a cutoff. There were many proteins differentially expressed between tg and wt in each of the fractions.

Supplemental Data. This document (6115_Report_020623.pdf) and an Excel spreadsheet containing Tables 1-11 (6115_SupplementalData_020623.xlsx) has been uploaded to the Project 6115 folder of Duke Box (<https://duke.app.box.com/folder/191981256371>)

DPMSR Proteome Discoverer Data Visualization Tool. To get a more interactive experience with these data, we have packaged the processed data into a Shiny web-based application. To use this application, you will need to download the supplemental “6115_020323.dpmsr_set” file locally and then go to: <https://dpmsr-pddp.cloud.duke.edu/> .

Brief instructions;

- Click “Browse” and upload the file above
- “load dpmsr_set”
- Setup Stats page by selecting comparison groups, p-value, fold-change, groups
- “set comparisons”
- “start analysis”

You can now navigate to various data visualizations under “stats”

- Graphs and Volcano plots can be modified and images downloaded
- Individual peptide distributions can be displayed by inputting protein accession number into protein plots tab

You can now navigate to various pathway tools under “pathway”

- Select organism and click “set pathway”
- You can now navigate through various geneontology, pathway, or protein interaction tools using your data (or a subset of your data).