Integration of MSstatsTMT into Proteome Discoverer Using the Scripting Node

David M. Horn, 1 Ting Huang, 1 Meena Choi, 1 Olga Vitek, 2 Rosa I. Viner, 1 Frank Berg, 1 Kai Fritzemeier, 1 and Carmen Paschke 2
1Thermo Fisher Scientific, San Jose, CA, 95134, 2Northeastern University, Boston, MA, 3Thermo Scientific GmbH, Bremen, Germany

ABSTRACT
Purpose: Integration of the MSstatsTMT tools directly into Thermo Scientific™ Proteome Discoverer™ 2.5 software using the Scripting Node.

Methods: An R script was created to send the PSM and study information to the MSstatsTMT R library for pairwise ratio generation. The Scripting Node subsequently re-imports the results for presentation and visualization in the Proteome Discoverer software.

Results: This script was evaluated with 3 datasets: the yeast triple knockout Thermo Scientific™ Pierce™ TMT11plex standard, two replicates of the same yeast triple knockout standard, and a three replicates TMT1plex experiment showing the effect of SARS-CoV-2 infection of human cells. We also demonstrate the integration of the MSstatsTMT results with the enrichment analysis function, aiding the downstream bioinformatic analysis of each of these datasets.

INTRODUCTION
The MSstatsTMT software package has become a standard tool for DDA, SRM and DIA label-free quantification and recently the MSstatsTMT™ libraries have been released for multiplexing TMT experiments. These R libraries require direct input of data from Proteome Discoverer, but they require a manual step to export the list of peptide-spectral matches and a separate file that maps quan channels and the various study factors. Here we show how to use the Proteome Discoverer 2.5 Scripting Node to automatically call an R script to send study factor annotations and peptide-spectral match tables to the MSstatsTMT R Bioconductor library. The Proteome Discoverer software subsequently reimports the results for visualization and bioinformatic analysis. We will demonstrate this workflow on multiple TMT datasets of increasing complexity.

MATERIALS AND METHODS
Creation of MSstatsTMT script
TMT datasets were analyzed by the Proteome Discoverer software using the standard analysis template for TMT SPS MS3 data, adding a Scripting Node to the Post Processing section in the Consensus workflow (see Figure 1).

RESULTS
Example 1: Yeast triple knockout – single replicate
The yeast triple knockout standard is a TMT1plex dataset with 3 replicates each of the MET6, HIS4, and URA2 gene knockouts plus 2 replicates of the parental strain as the final two channels. For each of the gene knockout samples, the protein expressed by that gene should have little to no abundance in the results relative to the parental strains and other proteins in the amino acid biosynthetic pathways will be upregulated to synthesize the amino acids in alternative pathways. This specific dataset was acquired on a Fusion Lumos instrument using the SPS MS® approach. In Proteome Discoverer, the default workflows for TMT SPS MS® were used to analyze the dataset and a Scripting Node with the customized script was used to produce the MSstatsTMT information.

Example 2: Yeast triple knockout – 2 replicates
For this example, two technical replicate datasets of the triple knockout were run in the same analysis. The increased number of replicates should lead to an increase in the number of significantly changed proteins for each gene knockout species compared to the parental strain. However, with multiple files, there are a subset of peptides with missing quantification values where the peptide was identified in only one of the two datasets. For the MET6 parental ratio, the new –log2 adjusted p-value is 14.7 versus 5.4 for the single dataset, showing the improvement in confidence in significantly changed proteins for each gene knockout species compared to the parental strain.

Example 3: SARS-CoV-2 host cell proteomics
Bojkova et al recently released a preprint that used TMT quantification to study the effect of the SARS-CoV-2 virus on human host cell proteins over a 24 hour period. The raw data for this study was published in the PRIDE archive with accession PXD017710. These data were analyzed in the Proteome Discoverer 2.5 using a similar data analysis strategy as described in the preprint. These data included three biological replicates and uses a common bridge channel to scale abundances across samples. For this poster, only the TMT quantification of the pulsed heavy-labeled peptides were quantified and the ratios for the virus-infected versus control samples were compared.

CONCLUSIONS
• The MSstatsTMT algorithms were successfully integrated into the Proteome Discoverer software using the Scripting Node
• MSstatsTMT can handle simple to complex datasets and adeptly handles datasets with missing values
• MSstatsTMT results integrate well with other features in the platform, including the new enrichment service to be released in the Proteome Discoverer 2.5 software

REFERENCES

TRADEMARKS/LICENSEING
© 2020 Thermo Fisher Scientific Inc. All rights reserved. Tandem Mass Tag and TMT are trademarks of Proteome Sciences, LLC. SEQUEST is a trademark of the University of Washington. All other trademarks are the property of Thermo Fisher Scientific and its subsidiaries. This information is not intended to encourage use of these products in any manner that might infringe the intellectual property rights of others.

PO65774-EN 0428