thermo scientific

AI-driven TMT Data Analysis using CHIMERYS and INFERYS in Proteome Discoverer

Mark Sanders[1], Mick Greer[1], David M. Horn[1], Kai Fritzemeier[2]
[1]Thermo Fisher Scientific, San Jose, CA; [2]Thermo Fisher Scientific, Bremen, Germany

ABSTRACT

The most popular acquisition strategies for multiplexed quantitation using tandem mass tags (TMT) are broadly grouped into MS2 and MS3 based approaches. Here we demonstrate the benefit of adding machine learning (ML) algorithms to TMT data analysis tools. Addition of INFERYS and CHIMERYS[1] increase identified and quantified proteins. Furthermore, we recommend data analysis workflows that best suit MS2 and MS3 based methods.

INTRODUCTION

Multiplexed proteomics quantitation using Thermo Scientific™ Tandem Mass Tags™ (TMT™) is a powerful tool to measure differences in cellular states. Much effort has been applied to instrument development and development of smart acquisition strategies such as real time search (RTS) and synchronous precursor selection MS3 (SPS-MS3) to improve the quantitative performance of TMT workflows. At the same time, there has been significant progress leveraging deep-learning and AI to enhance the depth of proteomics data analysis. For instance, Here we demonstrate the application of deep learning algorithms (INFERYS and CHIMERYS) to increase identified and quantified TMT labeled peptides in MS2 and MS3 based acquisition methods.

MATERIALS AND METHODS

Sample preparation and data acquisition were carried out as described in Fürtwangler *et al*, MCP, **2022** (PXD029320)[2], Koenig *et al*, Proteomics,**2022** (PXD031277)[3], and Paulo *et al*, JASMS, **2021** (PXD020815)[4]. Data from these studies were used here.
CHIMERYS and INFERYS algorithms were implemented as nodes in the Proteome Discoverer 3.0 software, except for data from PXD031277 which was processed in Thermo Scientific™ Proteome Discoverer™ 3.1 using the updated INFERYS perdition model that now supports phosphorylation modifications. CHIMERYS was deployed in the cloud and received prepared data from a local instance of Proteome Discoverer 3.0. Once processed by the CHIMERYS service, results were transmitted to the local instance for final processing and results viewing.

CHIMERYS AND INFERYS WORKFLOWS

Figure 1. Representative TMT Processing Workflows in Proteome Discoverer

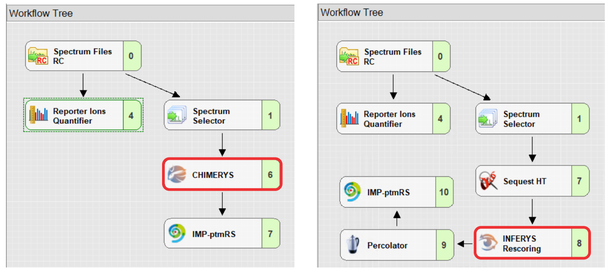


Figure 1 shows the processing workflows used for TMT data analysis. Note: ptmRS is included here because it was used for phosphorylation modification localization confidence (75% threshold), however this can be removed if modification localization is not required. CHIMERYS is a complete search engine which includes FDR validation, whereas INFERYS rescoring receives PSMs from SEQUEST and adds additional figures of merit (i.e. spectral angle) to each PSM which can be used by Percolator for the FDR calculation. An overview of both approaches can be seen in **Figure 2.**

Figure 2. Features of CHIMERYS and INFERYS

CHIMERYS

- FASTA Digest
- First Search
- Recalibration and Refinement
- Main Search + Deconvolution
- Percolator/ Mokapot

CHIMERYS

- Search Engine using ML based spectral prediction.
- Mokapot FDR validation introduced in PD3.1

INFERYS

- RT Prediction
- Fragment Prediction
- Rescoring (spectral angle etc.)

INFERYS

- Re-scores PSMs from SEQUEST using ML spectral prediction.
- Provides novel metrics to percolator (spectral angle etc.)

IMPROVEMENTS for MS2 and MS3 TMT DATA ANALYSIS

While CHIMERYS and INFERYS show improvements over SEQUEST in most cases **Figures 3&4** demonstrate that CHIMERYS offers the greatest improvement for MS2 based TMT data while INFERYS and CHIMERYS show comparable performance for MS3 based approaches. These results show the impact of data complexity on the results. For more complex data such as the MS2 data CHIMERYS can accurately deconvolute and predict complex chimeric spectra (**Figure 5**) while INFERYS is limited to the PSMs assigned by SEQUEST (via Precursor Detector node), which assign fewer chimeric spectra. However, for MS3 based approaches (**Figure 4**) INFERYS and CHIMERYS perform similarly.
Figure 3 results show the HYpro16 standard described by Paolo et al.[4] which consists of increasing amounts of yeast spiked into a constant background of human peptides. It is important to note in these data closeout was not used during real time search. Addition of closeout would restrict the number of peptides per protein selected for activation, therefore increasing the breadth of coverage for the targeted proteome. hrMS2 shows similar performance to RTS-MS3 however RTS-MS3 generates more accurate ratios (data not shown) and with closeout could yield more than hrMS2 in some cases.

Figure 3. INFERYS and CHIMERYS increase results vs. SEQUEST HT alone

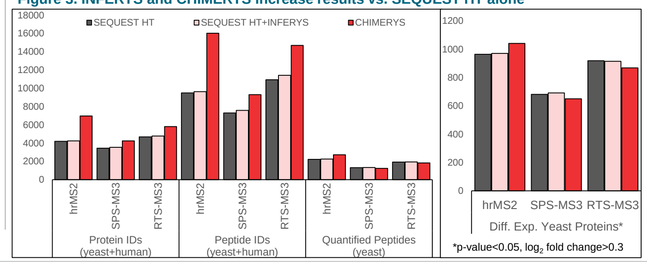
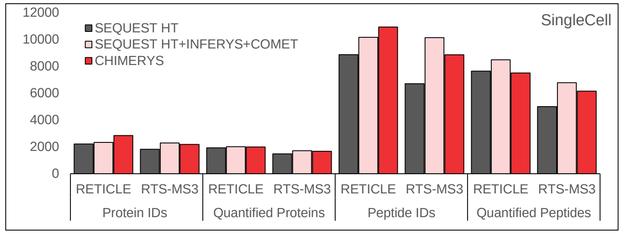


Figure 4. INFERYS and CHIMERYS increase results vs. SEQUEST HT for single cell data



PHOSPHOPEPTIDE PREDICTION

CHIMERYS and INFERYS models in PD 3.1 now support prediction of phosphorylation modifications and neutral losses. **Figure 5** demonstrates the capability of CHIMERYS to deconvolute and identify several PSMs from chimeric spectra. **Figure 6** shows a complex chimeric spectra which contains 7 TMTpro-labeled phosphopeptide PSMs. The upper half of the mirror plot shows the fragment ion matches assigned by SEQUEST for the 4 most abundant PSMs. The lower shows the matching INFERYS predicted fragment spectra. All PSMs which are assigned and predicted in Figure 6 are shown highlighted in the table below. For clarity only the first 4 assigned/predicted PSMs are displayed in the mirror plot.

Figure 5. CHIMERYS ID's up to 7 PSMs for a single MS2 scan (high-load phospho sample)

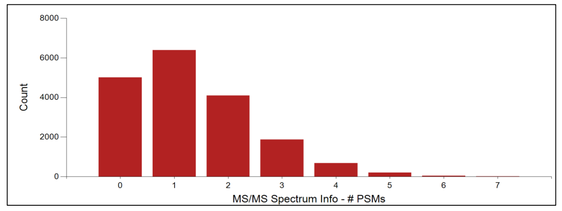
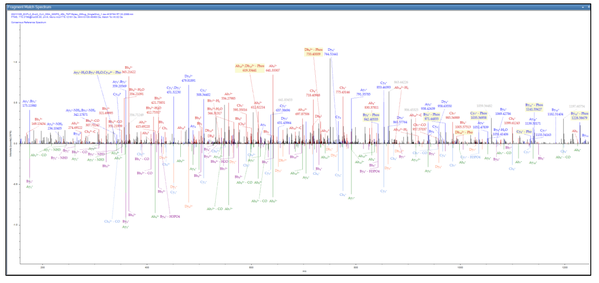


Figure 6. Predicted and Observed TMT labeled phosphopeptide PSMs



Annotated Sequence	Modifications	# Proteins	Master Protein Accessions	Protein Accessions	Normalized CHIMERYS Coefficient
[R].rSsDPALIQLSTSVSGVSNFSSEESPSR.[K]	N-Term(TMTpro); S3(Phospho)	1	Q8TEW0	Q8TEW0	0.27
[R].rSHSLYSQEQLDEsFSR.[N]	N-Term(TMTpro); S10(Phospho)	1	Q8TEW0	Q8TEW0	0.25
[R].eGVQQPLNVsLSEEQk.[S]	N-Term(TMTpro); S3(Phospho); K16(TMTpro)	1	P12270	P12270	0.22
[K].kAGsOCDIMDSSTEAPPSSk.[S]	N-Term(TMTpro); K1(TMTpro); S4(Phospho); K21(TMTpro)	1	Q9FSE8	Q9FSE8	0.13
[K].eEQSQITSQVKQQKSHR.[R]	N-Term(TMTpro); T11(Phospho)	1	Q96CW1	Q96CW1	0.08
[K].sVPVAVADEQEsEvEDDDLKPR.[G]	N-Term(TMTpro); S12(Phospho); S14(Phospho); K25(TMTpro)	1	Q9Y2K6	Q9Y2K6	0.04
[R].nNQVLCQQSQSRVAVQR.[I]	N-Term(TMTpro); T12(Phospho)	1	P49247	P49247	0.01

Figure 7. CHIMERYS improves analysis of phospho TMTpro data vs SEQUEST HT

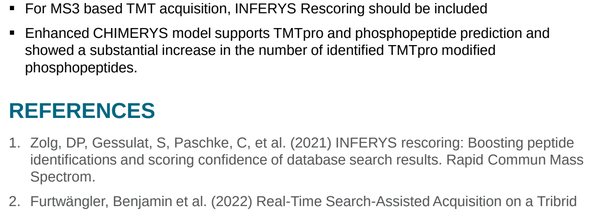


Figure 7 shows the increase in identification performance of TMT labeled phosphopeptides using CHIMERYS compared to SEQUEST HT. Larger gains in PSMs and IDs can be attributed to thorough assignment of chimeric spectra (**Figure 5**) from these complex data (200 µg starting material, before enrichment, 45 min gradient and 1.3 Th isolation.) [3]

CONCLUSIONS

INFERYS and CHIMERYS can be used to increase identified and quantified TMT peptides.

- INFERYS and CHIMERYS show greater performance than SEQUEST HT for complex data
- CHIMERYS should be included for MS2 data analysis
- For MS3 based TMT acquisition, INFERYS Rescoring should be included
- Enhanced CHIMERYS model supports TMTpro and phosphopeptide prediction and showed a substantial increase in the number of identified TMTpro modified phosphopeptides.

REFERENCES

1. Zolg, DP, Gessulat, S, Paschke, C, et al. (2021) INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results. Rapid Commun Mass Spectrom.
2. Furtwängler, Benjamin et al. (2022) Real-Time Search-Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics *MCP*, Volume 21, Issue 4
3. Koenig, C.,et al.(2022). Optimal analytical strategies for sensitive and quantitative phosphoproteomics using TMT-based multiplexing. *Proteomics*, 22, e2100245.
4. Paulo JA., et al.(2021) HYpro16: A Two-Proteome Mixture to Assess Interference in Isobaric Tag-Based Sample Multiplexing Experiments. *JASMS* 2021 6;32(1):247-254.

TRADEMARKS/LICENSING

PO2023-58EN

ThermoFisher SCIENTIFIC