The Power of Multiplexing- Combining TMT Discovery and Targeted Label Free Quantitation for Biomarker Analysis

[1]Xiaoyue Jiang, [2]Sergei Snovida, [1]David Horn, [3]Vic Spicer, [3]Oleg Krokhin, [1]Rosa Viner, [1]Andreas Huhmer
[1]Thermo Fisher Scientific, San Jose, CA; [2]Thermo Fisher Scientific, Rockford, IL; [3]University of Manitoba, Winnipeg, Canada

ABSTRACT

Purpose: To develop a high accuracy, sensitivity, robustness and speed biomarker discovery and verification workflow based on a Tandem Mass Tag™ (TMT™) quantitation approach and transfer of retention time information into label-free quantitation using peptide retention prediction.

Methods: The depleted plasma from three normal and three diabetic patients was digested and analyzed in parallel using TMT-based and label-free methods. TMT-labeled peptides were fractionated and analyzed on a Thermo Scientific™ Orbitrap Fusion™ mass spectrometer (MS). Analysis of non-labeled peptide mixtures was performed using 1D LC-MS on a Thermo Scientific™ Q Exactive™ HF-X hybrid quadrupole-Orbitrap™ mass spectrometer.

Results: 48 potential biomarker proteins were found in diabetic patients using a TMT-based discovery workflow and validated by targeted label free peptide quantification.

INTRODUCTION

LC-MS/MS has been a tool for biomarker discovery and validation for more than a decade. However, the current FDA approved biomarkers were almost exclusively developed by immunoassay. While LC-MS/MS can analyze thousands of analytes simultaneously, it is also prone to generating false positive results. In the discovery phase, achieving high quantitation accuracy can be a challenge when using the laborious label-free quantitation method. Insufficient sensitivity represents another bottleneck for the wide dynamic range of protein concentrations in plasma samples. Following the discovery phase, 10-100 potential biomarker proteins need to be targeted and verified on a smaller scale (10-50 patients) before validating a few true candidates on hundreds or thousands of samples. This verification step is also quite challenging due to the delicate balance between the number of targets versus instrument throughput. Here, we propose a workflow for plasma proteomics using the multiplexed TMT approach for biomarker discovery followed by rapid and robust verification with capillary flow LC on a novel Orbitrap™ platform with up to 40Hz scan speed (Figure 1).

MATERIALS AND METHODS

Sample Preparation

Twenty µL of plasma from three normal and three diabetic patients were depleted using the Thermo Scientific™ Pierce™ Top 12 Abundant Protein Depletion Spin Column. For the discovery experiment, the depleted plasma samples were digested, labeled with Thermo Scientific™ TMT6-plex™ reagents, mixed in equimolar amounts, and fractionated using the Pierce™ High pH Reversed-Phase Peptide Fractionation Kit. For the label free targeted MS experiment, the unlabeled depleted samples were analyzed using 1D LC-MS. All samples were spiked with the Pierce™ Peptide Retention Time Calibration (PRTC) kit as an internal standard to monitor and align elution time values.

Liquid Chromatography and Mass Spectrometry

Eight fractions of TMT6 labeled digest after high pH fractionation were separated using a Thermo Scientific™ 50cm Easy-Spray™ Column and a Thermo Scientific™ EASY-nLC™ 1200 UPLC system with acetonitrile 8% to 30% over 100 min, 30% to 50% over 40 min, at a flow rate of 300 nL/min. Each fraction was analyzed on an Orbitrap Fusion MS using synchronous precursor selection (SPS) MS^3 quantitation. The unlabeled samples (2µg injection) were separated on an Thermo Scientific™ UltiMate™ 3000 LC system at a flowrate of 5 ul/min and analyzed using loop 10 parallel reaction monitoring (PRM) and a full scan at 7.5K resolution on Q Exactive HF-XMS.

Data Analysis

LC-MS data were analyzed using the Byonic™ search engine in Thermo Scientific™ Proteome Discoverer™ 2.2 software using static carbamidomethylation (C), TMT 6plex (K,N-term) and dynamic oxidation (M) and deamidation (N, Q) modifications. Data were searched against the human protein subset of UniProt and results were filtered using a 1% protein FDR threshold. 3-4 peptides were selected from each protein that exhibited a significant (p-value <0.05) difference between the control and diabetic groups for targeted analysis. Retention time prediction was performed for these peptides using the Sequence-Specific Retention Calculator algorithm (SSRCal), which provided an output for both TMT labeled and unlabeled peptides (hydrophobicity index (HI)). The elution time window for each target was ±3 minutes. Skyline (3.6.0.10493) software was used to process the PRM data[1].

RESULTS

Biomarker discovery with TMT quantitation

Label free quantitation is widely used in biomarker discovery but it has limitations such as low throughput and poor reproducibility. The protein concentration in plasma spans 9 orders of magnitude which makes the detection of low abundance proteins extremely difficult even after depletion. Fractionation helps alleviate the sample complexity to improve protein/peptide identification, but requires a complicated normalization procedure for accurate quantitation and also significantly increases analysis time.

Recently, isobaric labeling techniques such as TMT and iTRAQ have become popular due to higher throughput and better precision and accuracy, because quantitation across samples is achieved in a single scan. The multiplexing capability of TMT significantly saves instrument time and provides possibilities to perform extensive fractionation without worrying about the normalization across fractions. In the first series of experiments, we performed a TMT-based discovery analysis using SPS MS^3 quantitation and reversed-phase fractionation of depleted plasma sample digests at high pH (Figure 1A). We identified 879 protein groups and 12,865 unique peptides with 95% of peptide groups quantifiable across 6 samples (Figure 2).

The new statistical analysis tools in Proteome Discoverer 2.2 software were used to detect proteins with significantly different abundances between the diabetic and normal samples in the discovery experiment (p value <0.05, log2 (ratio of two groups) >0.5, Figure 3). Some of the differentially expressed proteins are known to be diabetic markers such as Talin, insulin-like growth factor-binding protein 4 and hepatocyte growth factor activator [2-4]. In total, we observed 48 proteins with p-values less than 0.05 and peptides from these proteins were subsequently selected for targeted analysis.

Figure 1. Scheme of the discovery and targeted verification Workflows

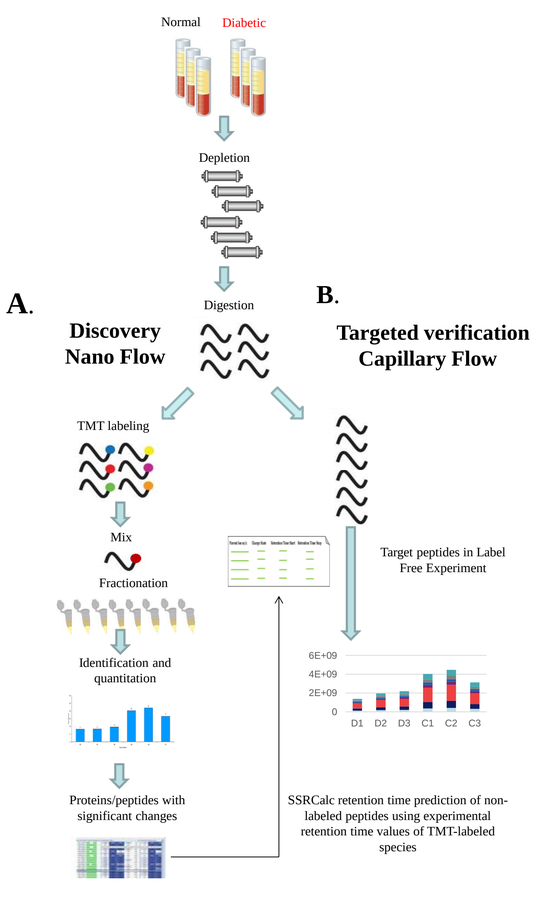


Figure 2. Numbers of (a) protein groups and (b) unique peptides identified from depleted plasma from 1ug of each fraction as well as the combined results.

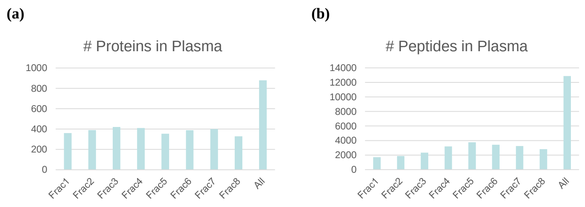
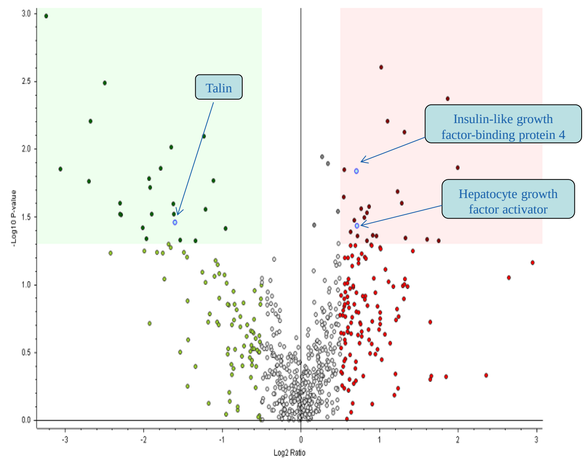


Figure 3. Volcano plot of plasma proteins which demonstrate significant difference between diabetic (x3 biological replicates) and normal (x3 biological replicates) patients (p value <0.05, log2 (ratio of two groups) >0.5).



Targeted verification of discovery quantitation

The major challenge for biomarker verification is to appropriately balance speed, sensitivity, and robustness. As dozens or even hundreds of samples are required to verify all the potential biomarker proteins found in the discovery experiment, the verification assay has to be fast. On the other hand, the assay also has to be sensitive enough to detect the proteins of interest which are generally of low abundance. We propose to use a sensitive targeted PRM quantitation method, coupled with capillary flow LC for improved robustness.

Nearly 180 peptide targets were selected for the 48 potential biomarker proteins from the discovery experiment. We applied this workflow on the highest speed Orbitrap™ analyzer with 40Hz speed. A full scan as short as 16ms was used to confirm peptide elution time followed by up to 10 targeted peptide MS/MS spectra at a resolution of 30K. This allowed us to save instrument time for collecting high resolution spectra on scheduled targets for more accurate quantitation. In this case, each targeted analysis can be accomplished within 1 hour leading to a throughput of 24 samples per day.

In scheduling experiments, the m/z, charge state, as well as the elution time are required to target peptides of interest (Figure 1B). All these parameters were obtained directly from TMT based discovery workflow using SSRCalc HI prediction as described [5] and using PRTC peptides as LC landmarks (Figure 4). Since the separation was achieved with reversed phase high performance LC, the elution time is proportional to the HI. As shown in Figure 4, the retention time and SSRCalc HI for the PRTC standard demonstrate a linear relationship for TMT labeled peptides. In order to get the elution time of interested peptides for the new LC separation conditions such as high flow, shorter gradient 1 hour vs 3 hours as for discovery phase, a new linear equation representing the fast gradient was generated from the PRTC run. SSRCalc HI of unlabeled peptides calculated from HI of TMT labeled peptides and the new linear equation enable us to predict elution time of the targets for the new LC set up (http://hs2.proteome.ca/SSRCalc/SSRCalcQ.html).

In the targeted experiment, we were able to quantify 30 out of the 48 proteins using the described strategy. The observed proteins showed consistent quantitation results compared to the TMT discovery experiment, as illustrated in Figure 5. The observed changes in protein expression were expected based on the literature. The missed proteins were either under the detection limit of targeted analysis or out of the scheduled time window, which can be confirmed through a fractionated analysis of unlabeled samples.

Figure 4. Generation of target peptide list by Sequence Specific Retention Calculator (SSRCalc) retention prediction tool.

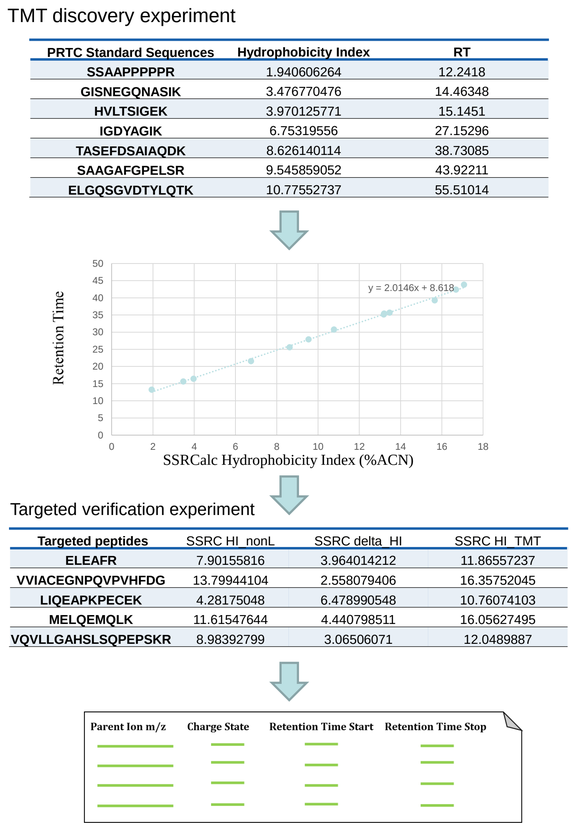
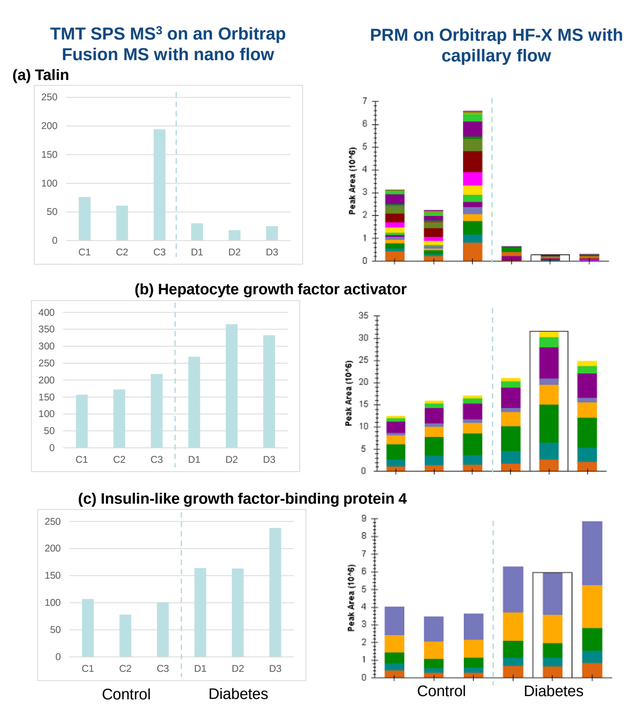


Figure 5. Comparison of TMT discovery quantitation and PRM targeted quantitation for potential diabetes biomarkers proteins identified in the study.



CONCLUSIONS

- A novel biomarker discovery workflow from TMT-based discovery to routine label free capillary flow quantitation using peptide retention prediction was developed.
- This workflow enables the biomarker discovery and validation in a highly multiplexed and rapid manner.
- Forty eight potential biomarkers for diabetes were identified and 30 were validated in targeted analysis of un-fractionated and unlabeled samples using 1 hour long robust LC separation at 5 ul/min flow rate.

REFERENCES

1. MacLean, B. et al. Bioinformatics. 26(7):966-8
2. Renno, W. et al. Medicina (Kannas) 42(2):147-63
3. Brismar, K. et al. J. of Clin Endocrinol and Metab 79:872-8
4. Fafalios, A. et al. Nature Med. 17(12):1577-84
5. Krokhin, O. et al. Anal Chem. 81(22):9522-30

ACKNOWLEDGEMENTS

We thank Susan Abbatiello for discussions about Skyline software and Marcus Kellmann for discussions of the PRM experiment.

TRADEMARKS/LICENSING