

# Rapid Discovery of Differentially Expressed Proteins in T2D Plasma Samples Using Improved UHPLC Chromatography and pSMART Data Acquisition

Scott Peterman,<sup>1</sup> Mary Lopez,<sup>1</sup> David Sarracino,<sup>1</sup> Amol Prakash,<sup>2</sup> Maryann Vogelsang,<sup>1</sup> Bryan Krastins,<sup>1</sup> Gregory Byram,<sup>1</sup> Gouri Vadali,<sup>1</sup> Mazi Mohiuddin,<sup>1</sup> Patrick Muraca,<sup>3</sup> Allison B. Goldfine,<sup>4</sup> Mary Elizabeth Patti<sup>4</sup>

<sup>1</sup>Thermo Fisher Scientific BRIMS, Cambridge, MA, <sup>2</sup>Optys Technologies, Inc., Philadelphia, PA, <sup>3</sup>Nuclea Biotechnologies, Cambridge, MA,

<sup>4</sup>Joslin Diabetes Center and Harvard Medical School, Boston, MA

## Overview

**Purpose:** Develop a robust, comprehensive workflow to perform large-scale global protein differential expression analysis facilitating putative groups of protein biomarkers.

**Methods:** Combine high-flow UHPLC chromatography with pSMART data acquisition and Pinnacle software for robust and efficient workflow support.

**Results:** Over 500 proteins (>5000 peptides) were quantified in 134 samples originating from 4 groups generated distinct sets of biologically and statistically relevant proteins translated into more targeted studies.

## Introduction

Translational proteomics links global protein discovery to targeted quantification with the ultimate goal of identifying and verifying relevant groups of proteins that collectively form potential biomarkers. To increase the effectiveness of the assay, two changes are made to traditional translational proteomics, increased number of proteins/peptides quantified across larger numbers of samples per biological group. These two changes facilitate robust determination of biological and statistical relevancy. These requirements introduce tremendous challenges for the experimental methods to maintain thorough identification and quantification without sacrificing throughput. We introduce a workflow to handle the increased analytical demands by incorporating ultra-high performance liquid chromatography (UHPLC)<sup>1</sup> with pSMART<sup>2</sup> to generate comprehensive LC-MS and MS/MS records for plasma peptides. In addition, we utilize the Pinnacle software for qualitative and quantitative data processing.

## Methods

### Sample Preparation

A set of 134 plasma samples were collected at the Joslin Diabetes Center representing five different biological groups. A total of 49 samples were classified as controls, 12 as type 2 diabetes (T2D), and 29, 25, and 23 samples for three other groups. The samples were digested with trypsin in the BRIMS Center using standard protocols and spiked with the PRTC kit prior to analysis.

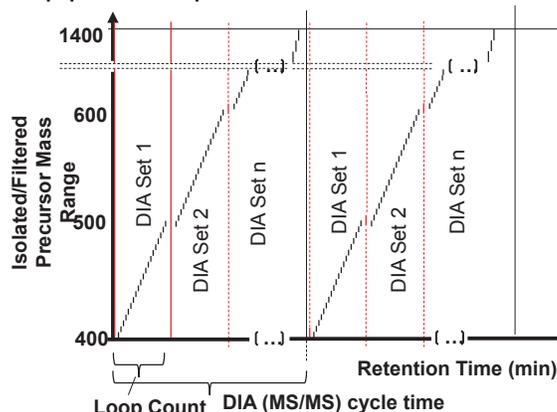
### Liquid Chromatography (or more generically Separations)

A Thermo Scientific™ Vanquish™ UHPLC autosampler and pump delivered an analytical flow rate of 200  $\mu\text{L min}^{-1}$  for a linear gradient of 0.49%  $\text{min}^{-1}$  in 85 minutes using A) 0.2% formic acid and B) MeCN (0.2% formic acid). A UHPLC column comprised of three individual 250 x 2.1 mm Thermo Scientific™ Acclaim™ 120 C18 columns aligned in series and heated to 55 °C. A cycle time of 110 minutes was used from injection to injection.

### Mass Spectrometry

All experiments were performed on the Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer. The first analysis was performed using data dependent acquisition (DDA) scheme using a top 10 experimental method with MS resolving power of 120k and 15k for MS/MS data acquisition. A maximum ion fill time of 100 msec was used for all scan events. The resulting data was used to build a consolidated spectral library. The second data acquisition method used pSMART, which is comprised of two independent acquisition cycles for MS and narrow DIA spectral acquisition events. See Figure 1 for more details. The resolution settings and maximum ion fill times used for pSMART were the same as those listed for the DDA experiments. A loop count of 10 was used to meter the DIA acquisition events in between the MS scans.

**FIGURE 1.** The pSMART acquisition scheme used for global protein profiling shows the interspersed MS and narrow DIA acquisition scheme. Ten narrow DIA spectra (Loop Count) are acquired after every MS spectra. The total acquisition cycle acquires at least 9 full scan MS spectra and 80 narrow DIA events were acquired covering a precursor m/z range of 400-1400 Da. The goal of pSMART is to acquire 9 HR/AM MS data points and at least 1 narrow DIA spectrum under each peptide elution profile.

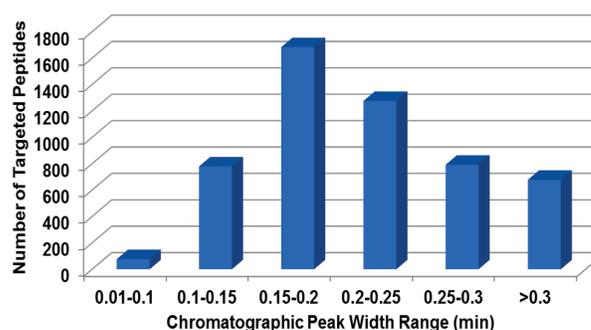


### Data Analysis

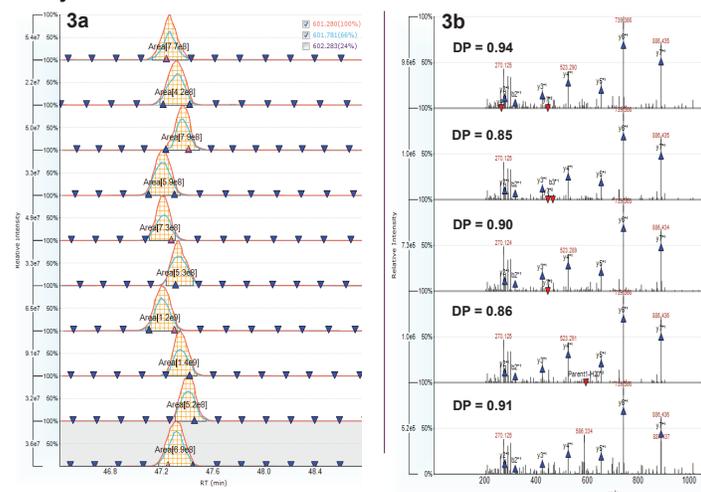
An unbiased search was performed for all DDA data using Thermo Scientific™ Proteome Discoverer™ 1.4 software. The search results were consolidated to form a spectral library. All translational proteomics studies were performed using the Pinnacle software loaded on a Velocity data processing computer. (both products from Optys Technologies, Philadelphia, PA) The DDA spectral library was used to determine all targeted proteins and peptides as well as all LC and MS information. This reference information was imported into Pinnacle for post-acquisition data extraction, verification, and quantification. In addition, the Pinnacle software performed statistical relevancy analysis using receiver operator curve (ROC) analysis and p- and q-value determination. Biological relevancy was determined using Ingenuity Pathway Analysis software (Qiagen, Redwood City, CA).

## Results

**FIGURE 2.** Incorporation of UHPLC pumps, columns, and high flow rates significantly increases the chromatographic performance for peak capacity, sensitivity, selectivity, and reproducibility. Figure 2 shows the reproducibility of repetitively measured retention times for all peptides across the month-long study.



**FIGURE 3** The results of UHPLC analysis is narrower peak widths which presents challenges for global profiling experiments. Narrow peak widths reduces cycle times and inhibits standard DIA experiments. Incorporation of pSMART enables routine qualitative and quantitative analysis of a majority of observed peptides. Figure 3a shows representative LC analysis of the target peptide EWFSETFQK from Apo C-1 on ten different samples randomly acquired. The two most abundant precursor isotopic XICs demonstrate chromatographic reproducibility. The blue triangles represent the narrow DIA spectra acquired that contain the precursor m/z value for the target peptide. The “up” triangles which represent matched product ion spectra. Those triangles in red are presented in Figure 3b and the order of DIA product ion spectra are linked to the XIC trace from top to bottom. The dot-product correlation coefficient displayed in each figure is determined based on the spectral library.



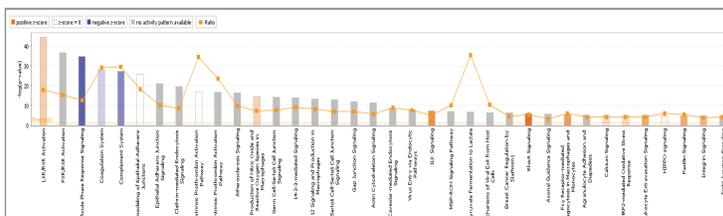
**FIGURE 4** Screen capture from Pinnacle showing the comprehensive protein analysis based on multiple peptide identification and processing. The bands displayed represent AUC values for each RAW file as well as the average AUC per group for rapid evaluation. In addition to the AUC display, the statistical analysis is presented for each peptide. The measured values for each peptide are used to determine the protein response.

Score	Name	PeptideCount	GroupRatioVisual	Differential Expression pValue	Differential Expression qValue	Differential Expression ROC
33 items	44232e8f9f_961074_2:spoliprotein M isoform 1	14		0.04402515e+402	0.22632038932	0.5915207552
Name	RatioVisual	GroupRatioVisual	Differential Expression pValue	Differential Expression qValue	Differential Expression ROC	
NQ44CCarboxymethylglutamate			0.18134351e+43207	0.157488575029		
SPNPPCCarboxymethylglutamate			0.154325039502	0.163558555048		
MK05CCarboxymethyllysine			0.210100119624	0.13015515453		
STP7CCarboxymethyllysine			0.02022934e+3622	0.02050771032		
KVH8L:K20722.2			0.230100615e+4192	0.150462751067		
TEP53CCarboxymethylphenylalanine			2.9e+30.02610460506	0.089740488565		
ELIATPVDVDFRMAAGAPMQLR			0.2501291010025	0.1637074765		
MFIQKMAALVITLMLKIQCCarboxymethyllysine			0.0959239304035	0.169359630028		
ARL3TR			0.02022934e+3622	0.02050771032		
FLVYR			0.25010760301002	0.163558555048		
QGLCarboxymethyllysine			0.04025150e+402	0.198010176504		
MK05CCarboxymethyllysine			0.2501049502625	0.155440185455		
VFHATVQSTDA			0.1402510950007	0.02050771032		

**FIGURE 5** Representative tabular output for global protein expression analysis from Pinnacle. The column arrangement displays average AUC values per group (GroupRatioVisual), number of peptides measured per protein, and p- and q-values for expression levels and the ROC scores for all proteins. The band view represents the average AUC group as labeled. The p-value, q-value, and ROC scores are in the order of group 1:0, group 2:0, 3:0, and 4:0.

Score	Name	GroupRatioVisual	PeptideCount	MaxPeptideArea	Differential Expression pValue	Differential Expression qValue	Differential Expression ROC
17.79	17.79		1	0.1112521941712	0.22020189328	0.1710212518	
3.546	0.02022934e+3622		3	0.02022934e+3622	0.22020189328	0.16601671046	
3.546	5.0e-43.8e-10.001915e+4		1	0.0011716280083	0.0011716280083	0.71046617208	
8.466	0.25039510e+4010		1	0.25039510e+4010	0.25039510e+4010	0.54062674064	
3.449	0.01501501040510e+4		34	0.25039510e+4010	0.25039510e+4010	0.05019048810	
2.649	4.0e-30.02114.0e-35.0e+4		30	0.170320170083	0.170320170083	0.70461761079	
1.249	6.5e-30.0724.1e+35.0e+4		8	0.170320170083	0.170320170083	0.67059078077	
2.449	0.04402515e+4025		14	0.32032038932	0.32032038932	0.59053207552	
9.348	1.2e-30.1210e-31.0e+4		19	0.110320190083	0.110320190083	0.740137076077	
1.040	1.0e-40.0471.5e+43.0e+4		44	0.00132010010083	0.00132010010083	0.79046618077	
1.811	1.8e-40.0717.0e+43.0e+4		48	0.110320190083	0.110320190083	0.710190176178	
1.340	0.0103090010053e+4		25	0.270320320083	0.270320320083	0.67042045077	
2.740	5.0e-40.01001010e+4		15	0.080320320083	0.080320320083	0.770410143077	
3.841	3.8e-30.0313.7e+31.0e+4		223	0.160320170083	0.160320170083	0.68046576077	
1.840	9.6e-30.016101.5e+4		37	0.240320160083	0.240320160083	0.67054077074	
2.641	2.1e-30.07601015.5e+4		21	0.110320190083	0.110320190083	0.75041066074	
1.747	5.5e-40.04510.07710e+4		1	0.080320320083	0.080320320083	0.66041015907	
4.648	7.7e-40.251088012		1	0.110320190083	0.110320190083	0.710130185008	
5.546	0.121102171.8e+40.01		1	0.32032038932	0.32032038932	0.580410188077	
2.248	0.01012401688e+4		1	0.280320320083	0.280320320083	0.660560501078	
2.848	8.6e-40.0121.5e+31.8e+3		1	0.110320190083	0.110320190083	0.7604101761078	
1.240	1.2e-30.10251.2e+31.0e+4		40	0.14032015011	0.14032015011	0.6605101761078	
1.548	0.030306801010e+4		2	0.280320320083	0.280320320083	0.6204101761078	
1.4411	2.4e-30.0910101010e+4		35	0.110320190083	0.110320190083	0.750410188077	
2.848	1.0e-30.12012.010		2	0.110320190083	0.110320190083	0.69046037078	

**FIGURE 6** Exporting the protein names, measured ratios, and confidence factors to Ingenuity Pathway Analysis. Results for Canonical pathway analysis from Ingenuity Pathway Analysis. The report consolidates proteins and measured expression ratios into biological groups. Figure 7b shows the Complement System pathway and expression ratios imported from the Pinnacle software.



## Conclusion

The presented research workflow demonstrates a high-throughput analytic method for global protein profiling. The incorporation of high-flow UHPLC provides significant advantages for handling complex biological samples in a highly reproducible manner. Future studies will evaluate smaller bore UHPLC columns and lower flow rates while maintaining the similar chromatographic performance. Additional benefits of the workflow include:

- UHPLC generated excellent separation facilitating MS analysis of non-depleted digest samples
- The UHPLC set up facilitated extremely reproducible separations over the month-long study
- Data collection using pSMART data acquisition facilitated reproducible analyte sampling by HR/AM MS and MS/MS
- The Pinnacle software automated ca. 92 MB of data processing in 1.5 hrs resulting in comprehensive qualitative, quantitative, and statistical analyses for over 500 proteins and 5000 peptides

## References

1. Gilar, M, Daly, A. E., Kele, M., Neue, U. D., Gebler, J. C. Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high performance liquid chromatography *J. Chrom. A.*, **2004**, *1061*, 183-92
2. Prakash, A., Peterman, S., Ahmad, S., Sarracino, D., Frewen, B., Vogelsang, M., Byram, G., Krastins, B., Vadali, G., Lopez, M. Hybrid data acquisition and processing strategies with increased throughput and selectivity: pSMART analysis for global qualitative and quantitative analysis *J. Proteome Res.*, **2014**, *13(12)*, 5415-30

For research use only. Not for use in diagnostic procedures.

[www.thermoscientific.com](http://www.thermoscientific.com)

©2015 Thermo Fisher Scientific Inc. All rights reserved. ISO is a trademark of the International Standards Organization. iRC PRO is a trademark of 2Next srl. All other trademarks are the property of Thermo Fisher Scientific and its subsidiaries. This information is presented as an example of the capabilities of Thermo Fisher Scientific products. It is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others. Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details.



**Africa** +43 1 333 50 34 0  
**Australia** +61 3 9757 4300  
**Austria** +43 810 282 206  
**Belgium** +32 53 73 42 41  
**Brazil** +55 11 3731 5140  
**Canada** +1 800 530 8447  
**China** 800 810 5118 (free call domestic)  
400 650 5118

**Denmark** +45 70 23 62 60  
**Europe-Other** +43 1 333 50 34 0  
**Finland** +358 9 3291 0200  
**France** +33 1 60 92 48 00  
**Germany** +49 6103 408 1014  
**India** +91 22 6742 9494  
**Italy** +39 02 950 591

**Japan** +81 6 6885 1213  
**Korea** +82 2 3420 8600  
**Latin America** +1 561 688 8700  
**Middle East** +43 1 333 50 34 0  
**Netherlands** +31 76 579 55 55  
**New Zealand** +64 9 980 6700  
**Norway** +46 8 556 468 00

**Russia/CIS** +43 1 333 50 34 0  
**Singapore** +65 6289 1190  
**Sweden** +46 8 556 468 00  
**Switzerland** +41 61 716 77 00  
**Taiwan** +886 2 8751 6655  
**UK/Ireland** +44 1442 233555  
**USA** +1 800 532 4752

**Thermo**  
SCIENTIFIC

A Thermo Fisher Scientific Brand