The Large-scale Analysis of Protein Expression in Human Ovarian Tumor Samples by LC-MS/MS/MS

Anthony Possemato, Sean Beausoleil, Klarisa Rikova, Hong Ren, Kimberly Lee, Michael Comb
Cell Signaling Technology, Inc., Danvers, MA

Introduction

The ability to evaluate protein expression from human tissue in a high-throughput, quantitative fashion is an invaluable resource and has eluded researchers for years. In this report, we describe a multiplexed approach using TMT (Tandem Mass Tag) labeling followed by SCX (Strong Cation Exchange) chromatography and LC-MS/MS/MS analysis to identify and quantify the expression of over 5200 proteins from 16 ovarian tumors compared to normal ovarian tissue. While the majority of the quantified proteins evaluated in this study represent common proteins seen in most tissues, we were able to identify several potential disease markers in the tumor samples based on their differences in expression. The apparent value from these initial experiments leaves us optimistic in the combination of these technologies.

Methods

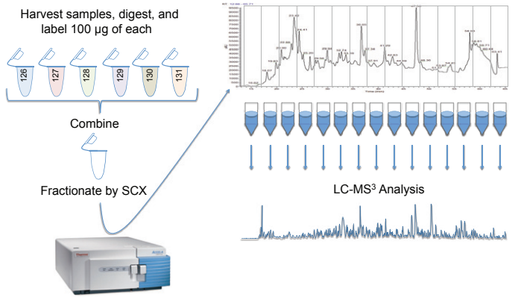
Ovarian tissue, both tumor and normal, were collected and immediately placed in liquid nitrogen. Samples were later weighed and processed in the appropriate amount of a 9 M urea lysis buffer. Samples were then reduced and alkylated before being digested with trypsin. After digestion, samples were purified over SepPak C18 columns and eluted peptides were lyophilized. After peptide quantitation 100 µg of each sample was labeled using TMT reagents. For the 126 channel, designated as the "control" or "normal" channel, a pool of peptides from six different normal ovarian tissues was combined in equal amounts and aliquoted for all subsequent runs (Table Six-plex Design). Labeled samples were then combined and subjected to Strong Cation Exchange (SCX) chromatography. Forty SCX fractions were collected, combined, and again purified over SepPak C18 columns before being processed by LC-MS/MS/MS analysis. Data generated from the MS analysis were processed through several modules in CORE, ending with quantitated data.

Six-plex Design

Channel	Tumor Set A	Tumor Set B	Tumor Set C	Tumor Set D
126	Normal Pool	Normal Pool	Normal Pool	Normal Pool
127	OC 03	OC 01	MGH	OC 01
128	OC 12	OC 07	OC 11	OC 08
129	OC 13	OC 16	OC 20	A 11
130	OC 26	OC 19	OC 26	A 15
131	OC 30	OC 26	A 22	A 19

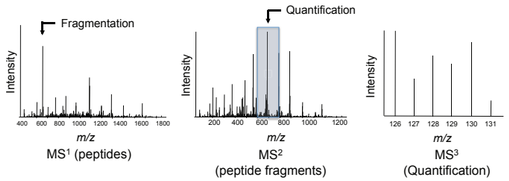
Normal Pool: B20, B26, B27, B28, B29, B30

The 126 channel for each six-plex was designed using a pool of six "normal" tissue samples, with each providing 1/6th of the total peptide amount. Tumor samples were duplicated across different six-plexes to provide a method of internal control.



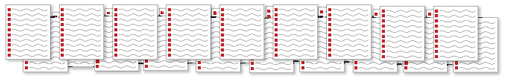
Samples are harvested in 8M Urea, homogenized if necessary, reduced, and alkylated before being digested. Samples were purified over SepPak C18 columns before being subjected to peptide quantification. 100 µg of each sample was then labeled with the appropriate TMT reagent according to the optimized protocol, combined at a 1:1:1:1:1:1 ratio, and again purified over a SepPak C18 column. Samples were then fractionated over a ACCELA® HPLC using a 4.6X200 mm SCX column packed with polysulfoethyl aspartamide material (5 µm particle size with 200-A pore). A 1ml fraction was then collected at each minute over the course of the 60 minute gradient. The first 40 fractions were then dried down and purified over another SepPak C18 column. Each purified fraction was then subjected to MS/MS/MS analysis. Data processing was performed using CORE. (Protein Sieve, Protein Assembler, Protein Quant.)

MS³ Quantification

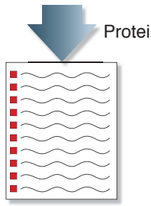


The incorporation of the MS/MS/MS methodology (Ting, et., 2011) dramatically decreases contaminating signal often seen in other MS/MS type approaches with TMT reagents.

CORE Processing



Unique peptides	Total peptides	Total proteins	Protein FDR	Peptide FDR
32,913	52,554	5,995	15.88% (476)	1.96% (514)



Unique peptides	Total peptides	Total proteins	Protein FDR	Peptide FDR
31,569	51,134	4,665	1.97% (46)	0.25% (63)

The preprocessed data from the MS analysis is brought into CORE containing a high level of error, especially within the protein level. These large lists of peptides are then filtered through the Protein Sieve module. This module will condense the peptide information from all associated fractions within the six-plex and create a consolidated list of proteins based on a probability filter of 2%. This stringency will eliminate a large majority of the low quality peptides, or peptides seen only once in the dataset which may be miss assigned. The post processed data above reflects how these peptides can effect the total protein number.

Results

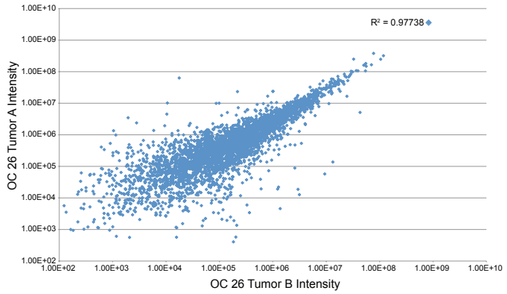
From these initial experiments, we were able to obtain ~450,000 MS/MS spectra correlating to ~50,000 total peptides (~35,000 unique) mapping to ~4000 quantified proteins per six-plex (Table A). In total we observed over 5200 unique quantified proteins. Raw data from the MS analysis was processed using in-house software. Taking advantage of the MS/MS/MS quantification methodology, we were able to dramatically reduce contaminating signal from our data (Ting et al., 2011). After examining the data in depth it was very apparent there were varying levels of contamination from blood proteins (serum albumin, hemoglobin, etc) seen throughout the six-plexes. We took into consideration normalizing the data based on common ribosomal protein intensities. After normalization, a majority of the proteins seen across the different six-plexes exhibited relatively equal abundance, which was expected. We were then able to easily identify several outliers from the data.

Table A

Six-Plex	Unique Quantified Proteins
Ovarian Tumor Set A	4,155
Ovarian Tumor Set B	4,340
Ovarian Tumor Set C	4,375
Ovarian Tumor Set D	3,640

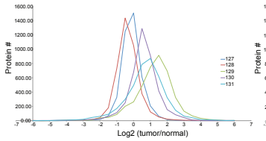
Post-processing results from CORE gave on average ~4100 unique quantified proteins per six-plex.

R^2 OC 26 Tumor Set A vs. B

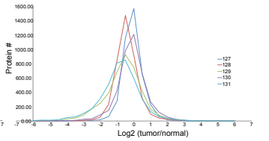


Allowing for sample duplicates within the different six-plexes enabled us to evaluate the consistency of the data. As you can see by these R^2 values, duplicate samples within the different six-plexes maintained a very high level of consistency. Thus, providing us with added confidence in the data and the method.

Raw Data

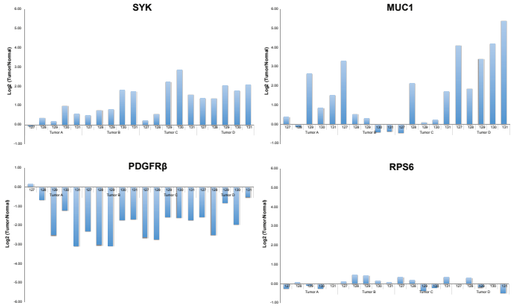


Normalized Data



Normalization was performed within each six-plex using abundant ribosomal proteins seen within the different samples. Based on the summed intensities of these common proteins, a ratio was created and then applied to the remaining data. As you can see with these plots, a majority of the data has shifted toward 0, meaning a majority of the data is unchanged, which is the expected result.

Interesting Findings



Here are a few examples of the kinds of data available from these such experiments. We can easily view cases where protein abundance is elevated in tumor vs. normal (SYK and MUC1), and then where protein abundance is down regulated in tumor vs normal (PDGFRβ). The last example RPS6, showing very little change in abundance from tumor to normal, is a good example of the type of data seen from the majority of proteins in the data sets.

Summary

Using the approach described here, we were able to accurately identify and quantify thousands of proteins observed in normal and tumor tissue in a single experiment. The ability to quantitatively profile disease states directly in human tissue represents a major advance in proteomics toward better understanding of the underlying mechanisms in human diseases such as cancer.

Reference: Ting L. et. al. (2011) *Nat. Methods* 8: 937-940

Cell Signaling TECHNOLOGY®

Contact Information
Anthony Possemato, Cell Signaling Technology, Inc. **Email:** apossemato@cellsignal.com