Proteomics-Based Biomarker Discovery Study in Human and non-Human Plasma Using a Two-Pass Workflow

Maryann S. Vogelsang¹, David A. Sarracino¹, Janin Schulte², Bryan Krastins¹, Michael Athanas¹, Amol Prakash¹, Alejandra Garces¹, Gouri Vadali¹,³, Shadab Ahmad¹, Gaiané Demirdjian², Jennifer N. Sutton¹, Heinz Redl⁴, Joachim Struck², Bruno Darbouret², and Mary F. Lopez¹

¹Thermo Fisher Scientific, Cambridge, MA, USA; ²Thermo Fisher Scientific, Hennigsdorf, Germany; ³VAST Scientific, Cambridge, MA, USA; ⁴Ludwig Boltzmann Institute, Vienna, Austria
Overview

**Purpose:** Sample preparation relying on fractionation to simplify the complexity and large dynamic range of plasma or serum samples does so at the potential cost of inaccurate or unreliable abundance measurements. Here we describe a simplified approach to biomarker discovery using a “Two-Pass” workflow that reduces the need for physical sample fractionation. The workflow encompasses robust, reproducible sample preparation, chromatography and strong informatics-driven data analysis.

**Methods:** Two-Pass discovery workflow using high-resolution LC-MS/MS coupled to ROC and differential expression analyses from stratified human sample cohorts and non-human animal models.

**Results:** The Two-Pass workflow reduced the number of replicates needed. Compared to a single-pass experiment, the workflow provided 20–50% more quantitatively associated protein identifications in a shorter time period (2–10X).

Introduction

Proteomic-based biomarker discovery approaches have primarily focused on direct profiling of serum or plasma to determine abundance changes that can be used to discriminate between populations. Due to the complexity and large dynamic range of serum and plasma, samples are often processed by depletion and/or fractionation in order to reduce their complexity. However, these manipulations can result in inaccurate or unreliable abundance measurements. Here we describe a simplified approach to biomarker discovery using a Two-Pass workflow to reduce the need for physical sample fractionation. The workflow encompasses robust, reproducible sample preparation, chromatography and strong informatics-driven data analysis.

Previously, we introduced a two-pass workflow exploiting the mass spectrometer’s accurate mass and broad dynamic range, which investigated the uncompromised quantitative data in Pass 1, and then targeted differentially expressed MS features in Pass 2. In Pass 1, we take advantage of the speed of the hybrid mass spectrometer to measure quantitative MS1 frames while concurrently measuring and identifying the Top-10 data-dependent MS2. In Pass 2 we identify additional differentially expressed peptides from the inclusion list built from Pass-1 data analysis.

Methods

**Sample Preparation**

This poster presents results from both primate and human plasma samples. The Two-Pass workflow was applied to 64 plasma samples of microbial-infected primates (monitoring eight subjects over a 24 hr time period, two of which were controls). Additionally, a small cohort of IRB approved human plasma samples were analyzed. The preliminary results in this report are intended to demonstrate the Two-Pass workflow, not biomarker discovery results.

**Liquid Chromatography and High-Resolution Mass Spectrometry**

As shown in Figure 1, the Two-Pass workflow strategy consists of the separate optimization of MS parameters and configuration for protein quantification and identification.

**Pass 1.** Plasma samples (300 ng of primate samples or 500 ng of human samples), were digested with trypsin and injected onto a Thermo Scientific EASY-nLC system configured with a 10 cm x 100 μm trap column and a 25 cm x 100 μm ID resolving column. The sample load was optimized for optimum quantification. Full scan data. Buffer A was 98% water, 2% methanol, and 0.2% formic acid. Buffer B was 10% water, 10% isopropanol, 80% acetonitrile, and 0.2% formic acid. Samples were loaded at 4 μL/min for 10 min, and a gradient from 0–45% B at 375 nL/min was run over 130 min, for a total run time of 150 min (including regeneration and sample loading). The Thermo Scientific LTQ Orbitrap Velos hybrid ion trap-Orbitrap mass spectrometer was run in a standard Top-10 data-dependent configuration, except that a higher trigger-threshold (20K) was used to ensure that the MS2 did not interfere with the full-scan duty cycle. This ensured optimal full-scan data for quantification. MS2 fragmentation and analysis were performed in the ion trap mass analyzer.
Proteomics data analysis was performed using Thermo Scientific SIEVE software version 2.0 that features chromatographic alignment, framing, differential ROC and ratio analyses. Both Top-10 data-dependent scans and full-scan data were analyzed with the SIEVE™ software using chromatographic alignment followed by feature extraction using unsupervised statistical techniques including isotope deconvolution. ROC curves were constructed for the top marker candidates as determined by their ROC Area-Under-the-Curve (AUC). Based upon various criteria including ROC AUC, low ratios, high ratios, high abundance, and low abundance, an inclusion list was created for the best candidates. This inclusion list was used for MS2 acquisition in Pass 2.

Data Analysis

Pass 2. Masses that represent differentially expressed MS1 features were built into an inclusion list and used in Pass 2 analysis. This ensured that the instrument only acquired MS2 data for the masses in the inclusion list (not the highest intensity masses). A larger sample load was used in the Pass 2 runs (600 ng to 800 ng), allowing for higher quality MS2 spectra. Because these full-scan spectra would not be used for quantification, peak shape and intensity reproducibility were not crucial. All fragmentation analysis was done in the Orbitrap mass analyzer, using both HCD and CID. Fragmentation scans from Pass 2 were analyzed using SEQUEST® and FDR analysis to make identifications. The fragmentation search results from Pass 2, along with the quantitative information obtained from Pass 1, were also analyzed using the SIEVE software. Fragmentation scan information was assigned to frames based upon the precursor m/z and retention time.

Results

Time-Saving Two-Pass Workflow

The Two-Pass workflow was applied to both sample sets. Our previous findings have demonstrated that the Two-Pass workflow can accurately detect, quantify and identify unlabeled differentially-expressed proteins within plasma samples. Given the robust chromatography, high-resolution mass spectrometers and computing power of SIEVE 2.0 software, the workflow enables identification of differentially expressed proteins in a single-pass experiment that is two- to ten-times faster than other approaches (Table 1).

FIGURE 1. Two-Pass discovery workflow. With robust LC-MS/MS methods and powerful SIEVE software, we are able to identify differentially-expressed proteins in unlabeled plasma samples.
TABLE 1. High-throughput is an important advantage of the Two-Pass workflow. For the sample preparation and LC-MS/MS steps, the workflow in this example has a ten-fold time advantage over fractionation sample preparation approaches. In comparable timeframe we can investigate ten times the number of human samples. Thus, the Two-Pass workflow offers significantly improved sample statistics.

<table>
<thead>
<tr>
<th></th>
<th>Two-Pass Workflow</th>
<th>Classic Fractionation</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Depletion</td>
<td>NO</td>
<td>1 hr/sample</td>
</tr>
<tr>
<td>Fractionate</td>
<td>NO</td>
<td>10 fractions</td>
</tr>
<tr>
<td>Digest</td>
<td>50 samples</td>
<td>500 samples</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>100 hrs</td>
<td>1000 hrs</td>
</tr>
</tbody>
</table>

Figure 2 shows the method for assessing systematic errors without technical replicates. Systematic errors are assessed from truplicate acquisitions of standardized peptide samples. Internal standards are spiked in all samples. This approach eliminates the need for sample replicates and conserves valuable specimens. All samples are acquired in high-resolution full-scan and Top-10 data dependent mode, on the LTQ Orbitrap Velos mass spectrometer. Biological variance and outliers are assessed with CV, PCA and other statistical methods.

FIGURE 2. Pass 1 acquisition cycle.

Figure 3 shows the gel view of the frames obtained from nine LC-MS runs. In this example, we have 20,000 potentially useful frames in the gel view. A frame represents a potentially interesting feature (peptide) found in a collective data set. Using frame filters, we can separate the differentially expressed frames between the two groups (see Table 2).

FIGURE 3. Gel view of frames from 9 LC-MS runs.
SIEVE Software Analysis

The following data are representative of a proteomics differential case study with ROC analysis from an animal model for sepsis.

Table 2 shows that the Two-Pass workflow allows for confident identification of total protein as well as unlabeled differentially expressed peptides. Initial analysis steps include the generation of 20,000 frames (peptides). Using frame filters based upon ROC AUC or ratios, we easily identified frames that separate the infected from non-infected groups. Given that our Pass-1 experiments are simultaneously collecting Top-10 data-dependent MS2 spectra, we can easily identify the abundant proteins (FDR=1%). The unidentified frames can then be exported to an inclusion list for Pass-2 analysis.

TABLE 2. Number of proteins and peptides that meet filtering criteria in Pass 1 within a ROC analysis experiment.

<table>
<thead>
<tr>
<th>Condition/ Filters</th>
<th>Proteins</th>
<th>Peptides (frames)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL frames</td>
<td>269</td>
<td>20,000</td>
</tr>
<tr>
<td>AUC &gt; 0.8</td>
<td>127</td>
<td>8562</td>
</tr>
<tr>
<td>Fold Change &gt;1.5</td>
<td>65</td>
<td>3944</td>
</tr>
<tr>
<td>Identified in Pass 1</td>
<td>65</td>
<td>112</td>
</tr>
<tr>
<td>Unidentified in Pass 1 (for Pass-2 inclusion list)</td>
<td>0</td>
<td>3832</td>
</tr>
</tbody>
</table>

Figure 4 shows a frequency histogram of scheduled fragment events per minute for Pass-2 analysis (Pass-2 Inclusion List). After the given frame filter conditions [AUC>0.8 and (NRatio>1.5 or NRatio<0.6) and PRElement<1 and goodid<1 and Charge=1 and Charge=5], 1228 frames were exported as an inclusion list for Pass-2 analysis. NOTE: These frames represent differentially expressed peptides.

FIGURE 4. Frequency histogram of scheduled fragment events per minute for Pass-2 analysis

Figure 5 shows an example of proteins identified in Pass 1 and Pass 2 that have a minimum of two peptides with FDR of 1%, from ROC analysis of human samples. The identified plasma proteins from our mass spectrometry analysis were searched against a database of known plasma protein concentrations. In Pass 1, we identify proteins that span at least nine orders of magnitude.

FIGURE 5. Plasma protein concentrations of identified proteins.
Pass-2 Results

On average, we achieved an 82% success rate in MS2 acquisitions from the inclusion list. Frame parameters pending, we have had even higher success rates in other experiments.

An example whisker plot of peptide coverage of a potential biomarker is shown in Figure 6. Pass 2 often results in an increased peptide coverage of the proteins identified in Pass 1. With respect to roughly 10% of frames representing one peptide of a given protein in Pass 1, we identified a second peptide for that corresponding protein in Pass 2. This result strengthens confidence in identification and quantification.

FIGURE 6. Example whisker plot of peptide coverage of a potential biomarker.

Conclusion

- The Two-Pass workflow was successfully applied to multiple cohorts of both primate and human plasma research samples.
- The Two-Pass workflow accelerates the time to targeted assays and validation of potential biomarkers.
- The Two-Pass workflow delivers increased confidence in protein biomarkers through increased peptide coverage of differentially expressed peptides spanning nine orders of magnitude of protein abundance.

References


SEQQUEST is a registered 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 manners that might infringe the intellectual property rights of others.