

A Two-pass Informatics-driven Label-free Workflow For Discovery Of Neurovascular Mediators In PFO-Related Stroke

Michael Athanas¹; David Sarracino²; Taha Rezai²; Amol Prakash²; Jennifer Sutton²; Bryan Krastins²; Mingming Ning³; Mary F Lopez²

¹VAST Scientific, Cambridge, MA USA ; ²Thermo Fisher Scientific, BRIMS, Cambridge, MA, USA; ³Mass General Hospital/Harvard Medical School, Boston, MA

Thermo
SCIENTIFIC

Overview

Purpose: Discover potentially relevant markers of high and low abundance that distinguish stroke neurovascular mediators that are uniquely found in samples from individuals with a Patent Foramen Ovale (PFO).

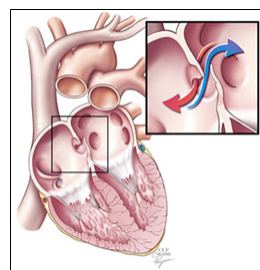
Methods: The workflow outlined involves two pass sample measurements: pass 1 full-scan measurements are used for quantitative analysis; an inclusion list derived from pass 1 drives targeted fragmentation acquisition.

Results: A preliminary panel of markers was determined confined to the analysis of samples taken from the right arterial chamber pre- and post-PFO closure.

Introduction

The product of a contemporary biomarker discovery experiment is a panel of illuminating marker candidates that distinguish the experiment classes such as normal and disease. Experiments that require samples obtained from a study involving significant cohorts are often time consuming and costly – especially if replicate measurements are required. In addition, such experiments are often long running and subject to various systematic effects such as sample degradation, instrumentation sensitivity drift, and so on. We have developed an approach to LC-MS discovery that optimizes quantitative sensitivity, monitors the impact of various systematic effects, and boosts ion-identification efficiency. We have devoted considerable effort towards experiment design, devised calibration methods, and developed configuration and operation procedures.

Patent Foramen Ovale (PFO) is a septal defect in the heart allowing blood to flow between the atrial chambers. The purpose of this study was to determine, through liquid chromatography and mass spectrometric analysis, whether there are differences in neurovascular mediators in plasma samples collected from the right and left atrial chambers before and after endovascular surgery to correct a PFO.



Methods

The experiment was designed for study of moderately large cohorts of plasma samples—39 from individuals who had suffered ischemic strokes and 4 from individuals who had not suffered strokes: 172 samples in all.

The plasma samples were enzymatically digested followed by online desalting using a trap column [Step 2]. Samples were spiked with a well-studied synthetic protein internal standard [1] prior to monitor enzymatic digestion. In addition, Michrom angiotensins standards [2] were spiked at into samples prior to LC injection to assess acquisition quantitation robustness. Glocogon was added to deter well-plate adhesion.

We found that the chromatography and instrument methods for optimal full-scan quantitative measurements conflicted with methods of optimal fragmentation scans [Step 1]. We exploited the mass spectrometer's mass accuracy and broad dynamic range by dividing the data measurement into two distinct passes. The first pass focused on acquiring uncompromised and optimized full-scan (MS) data for highly reproducible quantitation. This first full-scan quantitative pass was used to generate an inclusion list of potentially interesting features. This inclusion list was then used for targeted fragmentation scan acquisition on a second pass of a subset of the data samples.

Results

During experiment design, the acquisition robustness was determined from repeatedly measuring a designated standard plasma sample under various conditions. CV's of the integrated intensity of the spiked standards were used as a metric of robustness and reproducibility and was found to be about 8% in the designated standard plasma. We determined that replicate sample measurements were excessive with such high reproducibility. In order to obtain confidence that our measurement was consistently robust, an acquisition cycle was developed where triplicate measurements of the designated standard plasma were interspersed amongst every six real sample measurements (Figure 1). In addition, the cycle contained two blank runs to flush the column and acquisition of one of the samples with a method to produce top-10 fragmentation scans.

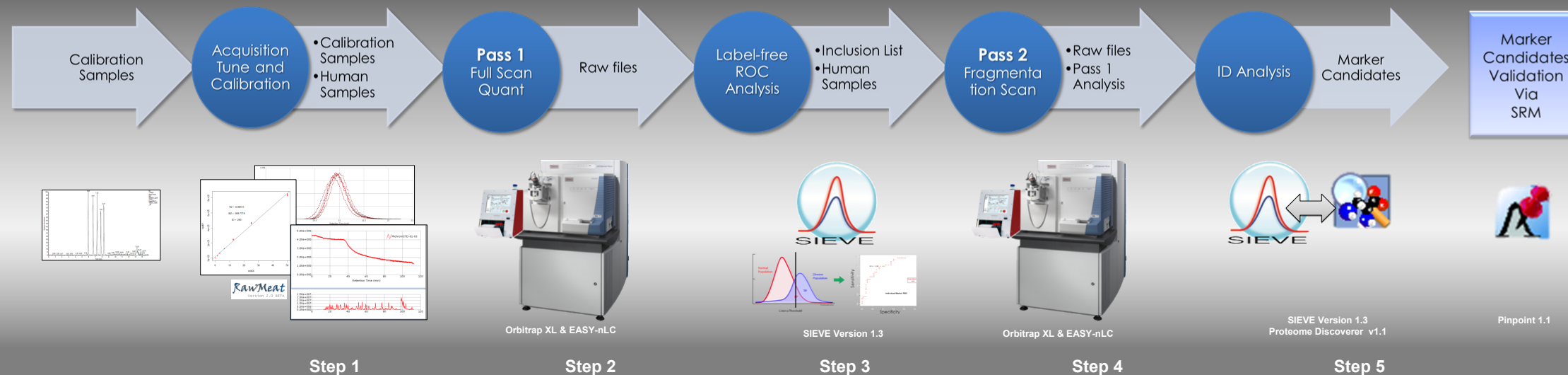
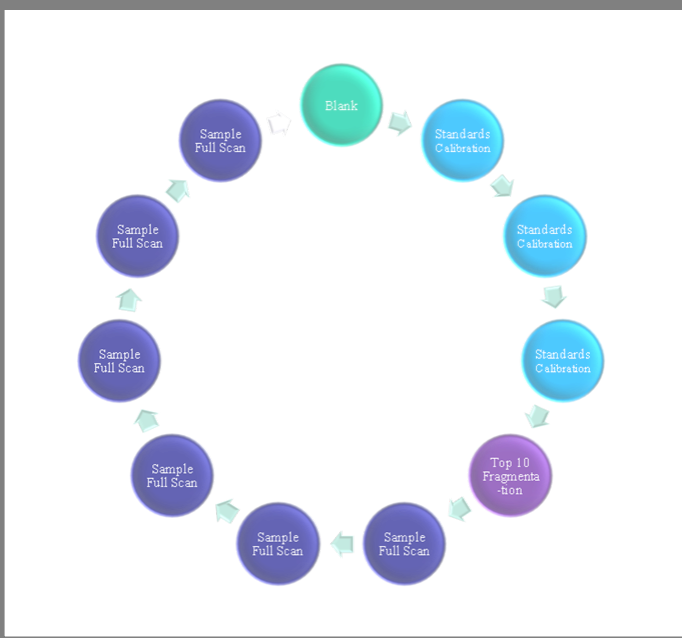
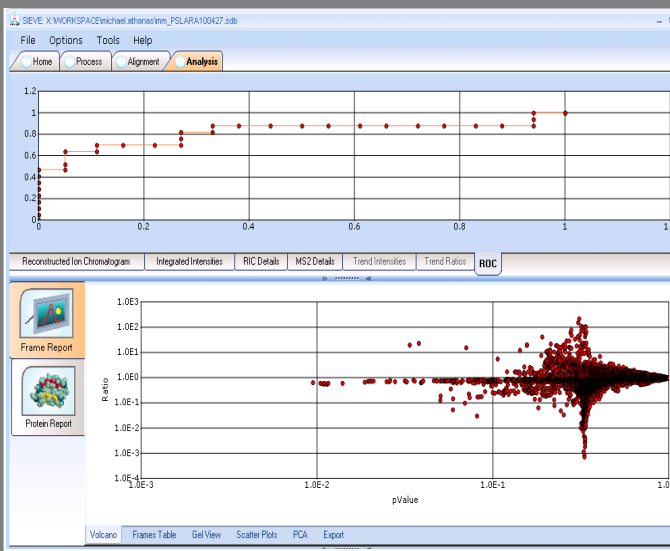


FIGURE 1. Sample measurements for Pass 1 were measured without technical replicates. Other sample acquisitions were interspersed for instrument maintenance and quality monitoring. The cycle of measurements began with a blank run to cleanse the system, followed by triplicate standard plasma measurements for quality monitoring (full-scan only), followed by a run from a random sample in the cycle with combined full-scan and MS2 fragmentation scan. The cycle was concluded with six measurements (full-scan only) of human samples.



The preliminary analysis in this paper was confined to the study of plasmas taken from the right arterial chamber prior to PFO closure and shortly after closure. Data were analyzed using Thermo Scientific SIEVE software version 1.3 with chromatographic alignment followed by feature extraction using unsupervised statistical techniques that included isotope deconvolution [Step 3]. ROC curves (Figure 2) were constructed for ratios of the top marker candidates determined by their ROC Area Under the Curve (AUC). An

FIGURE 2. A Screenshot from SIEVE v1.3. Data from Pass 1 are chromatographically aligned and framed to ascertain potentially interesting features in the collective data set. A ROC analysis is performed on each frame. Information derived from the ROC analysis and other information is used to compile an inclusion list for Pass 2 LC-MS acquisition.



inclusion list was constructed for the best candidates based upon various criteria including ROC AUC, low ratios, high ratios, high abundance, and low abundance. Fragmentation scans from the second pass were analyzed for identification using SEQUEST and Percolator [3][Step 4].

SIEVE was used again to combine the fragmentation search results from Pass 2 with the quantitative information from Pass 1 [Step 5]. Fragmentation scan information was assigned to SIEVE frames based upon the precursor *m/z* and retention time.

FIGURE 3. LC-MS data acquisition quality was monitored throughout Pass 1 by performing triplicate measurements of a plasma standard with the Michrom Medium Molecule Test Mixture [2]. The SIEVE frames corresponding to these peaks were constructed. Peak integrated intensity from SIEVE was determined by integrating the peak area confined to a frame. The CV's for five of the peaks were averaged across a four-week period corresponding to Pass 1 acquisition.

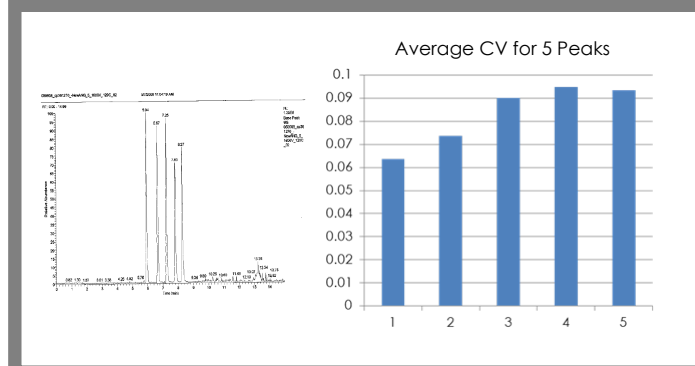


FIGURE 4. The top 50 AUC single-marker candidates were reanalyzed by combining every pair and determining their expression ratio. The ROC analysis of the top-10 ratio pairs is shown below. Identification of each of the marker candidates was determined from an identification analysis (SEQUEST and Percolator[3]) from data derived from Pass 2.

Marker Pair	AUC	Descriptions
1	0.85	nuclear receptor binding SET domain protein 1 isoform a [Homo sapiens] [MASS=267339]
2	0.85	calcium binding protein P22 [Homo sapiens] [MASS=22456] [MASS=22456] golgin 97 [Homo sapiens] [MASS=88170] [MASS=88170]
3	0.83	WD repeat domain 89 [Homo sapiens] [MASS=43214] [MASS=43214] nuclear receptor binding SET domain protein 1 isoform a [Homo sapiens] [MASS=267339]
4	0.82	ankyrin 3 isoform 1 [Homo sapiens] [MASS=480407] [MASS=480407] nuclear receptor binding SET domain protein 1 isoform a [Homo sapiens] [MASS=267339]
5	0.82	killer cell immunoglobulin like receptor two domains long cytoplasmic tail 5B precursor [Homo sapiens] [MASS=40577] [MASS=40577] nuclear receptor binding SET domain protein 1 isoform a [Homo sapiens] [MASS=267339]
6	0.82	family with sequence similarity 71 member B [Homo sapiens] [MASS=64755] [MASS=64755] nuclear receptor binding SET domain protein 1 isoform a [Homo sapiens] [MASS=267339]
7	0.82	RAB6 interacting protein 1 [Homo sapiens] [MASS=147095] [MASS=147095] nuclear receptor binding SET domain protein 1 isoform a [Homo sapiens] [MASS=267339]
8	0.81	WD repeat domain 89 [Homo sapiens] [MASS=43214] [MASS=43214] ankyrin 3 isoform 1 [Homo sapiens] [MASS=480407] [MASS=480407]
9	0.81	keratin hair basic 1 [Homo sapiens] [MASS=54928] [MASS=54928] nuclear receptor binding SET domain protein 1 isoform a [Homo sapiens] [MASS=267339]
10	0.81	upstream binding transcription factor RNA polymerase I isoform a [Homo sapiens] [MASS=89405] [MASS=89405] golgin 97 [Homo sapiens] [MASS=88170] [MASS=88170] upstream binding transcription factor RNA polymerase I isoform a [Homo sapiens] [MASS=89405] [MASS=89405]

Conclusions

We developed a robust approach to marker discovery in which we exploit the mass spectrometer's accurate mass and broad dynamic range capabilities by taking two distinct passes of data measurement. From Pass 1 quantitative data, marker candidates are determined using ROC analysis, expression ratio, and signal abundance. An inclusion list is compiled and used for a subsequent acquisition pass of a subset of the data. Data from the two passes are combined to provide an informative candidate marker list that can be used for SRM validation. The most significant marker candidates are determined by a ROC analysis of ratios of markers. That is, the sample-to-sample ratio of two individual candidate markers (Figure 4).

References

- (1) Bryan Krastins; Amol Prakash¹; Scott Peterman¹; David Sarracino¹; Michael Athanas¹; Taha Rezai¹; Mary Lopez¹. Development of a Synthetic Protein Quality Control (QC) Standard for the Assessment of Sample Proteolysis Reproducibility. 2010, ASMS 2010 Poster
- (2) Medium Molecule Test Mixture (CTM/0001/02)(5 Angiotensins), Michrom Biosciences
- (3) Lukas Käll, John D. Storey, Michael J. MacCoss and William Stafford Noble, Assigning confidence measures to peptides identified by tandem mass spectrometry, *Journal of Proteome Research*, 7(1):29-34, January 2008

SEQUEST is a registered trademarks of the University of Washington. RawMeat is a trademark of VAST Scientific. 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 manner that might infringe the intellectual property rights of others. For Research Use Only. Not intended for diagnostic procedures.