# Application of Label-Free Quantitative Differential Analysis for Biomarker Discovery in *Tuberculosis mycobacterium*

Michael Athanas[1], Bryan Krastins[2] , David Sarracino[2] , Amol Prakash[2], Taha Rezai[2], Mary F Lopez[2], Alejandra Garces[3], Sarah Fortuane[3]
[1]VAST Scientific, Cambridge, MA ,[2]Thermo Fisher Scientific, Cambridge, MA; [3] Harvard University, Boston, MA

## Overview

**Purpose: Mass spectrometric and bioinformatic analysis of *Mycobacterium tuberculosis* strains lacking the ESX1 locus to identify differentially expressed proteins. Discovery data were used to develop SRM assays for differentially expressed proteins.**

**Methods: High-resolution mass spectrometry coupled with label-free differential analysis using SIEVE software. MS spectra from discovery experiments were mined to facilitate development of SRM assays using novel SRM workflow software.**

**Results: SIEVE analysis confirmed the identification of the five previously identified secreted proteins as well as other differentially expressed proteins across the mutant strains. SRM assays for two proteins were developed.**

## Introduction

*Mycobacterium tuberculosis* (*Mtb*) has evolved specialized secretion systems for the transport of proteins across complex cell walls (1). One of these, the ESX1 secretion system, is critically required for the virulence of *Mtb.* Indeed, the primary attenuating deletion in BCG, the live attenuated vaccine used against tuberculosis, is loss of the genes encoding the ESX1 system. The ESX1 structural components are thought to form a multi-subunit cell-envelope spanning structure. However, the molecular function of the five known ESX1 substrates is unclear. It has been speculated that individual ESX1 substrates independently act as effectors to disrupt host cell functions and alternatively, that they interact to form an extracellular virulence machine. In order to comprehensively identify substrates of the ESX1 secretion system, we coupled high resolution LC-MS/MS with novel label-free differential analysis software to analyze *Mtb* mutant strains lacking the ESX1 locus(2). Bioinformatic analysis was carried out using SIEVE software. Results from the SIEVE analysis confirmed the identification of the five previously identified secreted proteins as well as other differentially expressed proteins across the mutant strains . High-resolution MS data from these discovery experiments were used in the development of selective reaction monitoring (SRM) based assays for target peptides in the secreted proteins.

## Methods

**Sample preparation**
*Mtb* culture supernatant s were prepared as described in (2) . Secreted proteins were precipitated from the supernatants with TCA and the protein pellet was resuspended in SDS PAGE loading buffer. Samples were run approximately 1 cm into10% SDS PAGE gels and the entire protein containing band was excised and subjected to in-gel enzymatic digestion before loading onto the mass spectrometer.

**High resolution LC-MS/MS**
High resolution LC-MS/MS analysis was carried out on a Thermo Scientific LTQ Orbitrap XL mass spectrometer. Samples in 5% (v/v) acetonitrile 0.1% (v/v) formic acid were injected onto a 75 µm x 25 cm fused silica capillary column packed with Thermo Scientific Hypersil Gold C18AQ 5µm media, in a 250 µL/min gradient of 5% (v/v) acetonitrile, 0.1% (v/v) formic acid to 30% (v/v) acetonitrile, 0.1% (v/v) formic acid over the course of 180 minutes with a total run length of 240 minutes. The LTQ Orbitrap XL$^{TM}$ mass spectrometer was run in a top 5 configuration at 60K resolution for a full scan, with monoisotopic precursor selection enabled, and +1, and unassigned charge state rejected. The analysis was carried out with CID fragmentation modes.

**Label free Differential analysis using SIEVE**
The SIEVE iterative workflow included chromatographic alignment, a global intensity-based feature extraction and aggregate protein identification assignment. Chromatographic alignment was based upon the pair-wise MS full scan comparison of all experimental MS runs with respect to a chosen reference MS run (Figure 1). Overlapping correlation sub-matrices (tiles) were computed using a novel scalable adaptive tile algorithm. An optimal path through each tile was determined using dynamic programming and a final alignment score was calculated (Figure 2). Subsequent to chromatographic alignment, potentially interesting features were exposed based upon high-intensity peaks found in the aligned collective data set. Individually, these peaks defined frames i.e. well defined rectangular regions in the full scan (*m/z* versus retention time) plane (Figure 3). Reconstructed ion chromatograms were calculated for each frame to assess relative expression ratio and supporting statistics. After framing, MS2 fragment scans associated with each frame were processed with SEQUEST. A consensus protein report was constructed by statistically aggregating frame information to construct peptides and then peptide information to build proteins (Figure 4).

**SRM assays**
SRM assays were developed on a Thermo Scientific TSQ Vantage triple quadrupole mass spectrometer, Surveyor MS pump, Micro autosampler and an IonMax Source equipped with a low flow metal needle. Reverse phase separations were carried out on a 1mm X 50mm Thermo Scientific Hypersil Gold 1.9µm C18 particle. Solvent A was LC-MS grade water with 0.2% (v/v) formic acid, and solvent B was LC-MS grade 30% (v/v) acetonitrile with 0.2% (v/v) formic acid (Optima grade reagents, Thermo Fisher Scientific). Thermo Scientific SRM Workflow prototype software (available upon request) was used for targeted protein quantification. This software algorithm is used to predict candidate peptides and for choosing multiple fragment ions for SRM assay design, building an instrument method and a sequence file, and also for automatic peptide identity confirmation and quantitative data processing. For the workflow described herein, the differentially expressed peptides that were identified in the LC-MS/MS discovery MS data were imported directly into SRM builder software. Transitions were chosen based on the predominant fragments observed in the discovery data (>5 transitions per peptide). Peptides were identified by co-eluting light and heavy transitions derived from synthetic peptide standards.

## Results

Label-free differential analysis of technical and biological replicates of *M. tuberculosis* knockout strains. Selected proteins were subsequently targeted for SRM analysis (Figure 7).

**FIGURE 1. Chromatographic alignment was based upon the pairwise MS full scan comparison of all experimental MS runs with respect to a chosen reference MS run.**

**FIGURE 2. Overlapping correlation sub-matrices (tiles) were computed using a novel scalable adaptive tile algorithm. An optimal path through each tile was determined using dynamic programming and a final alignment score was calculated.**

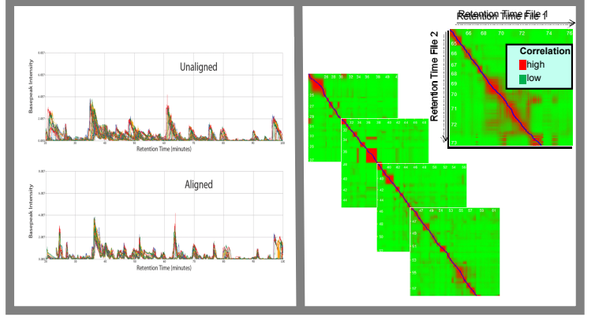



**FIGURE 3. Potentially interesting features were exposed based upon high-intensity peaks found in the aligned collective data set. Individually, these peaks defined frames i.e. well defined rectangular regions in the full scan (*m/z* versus retention time) plane. Reconstructed ion chromatograms were calculated for each frame to assess relative expression ratio and supporting statistics**

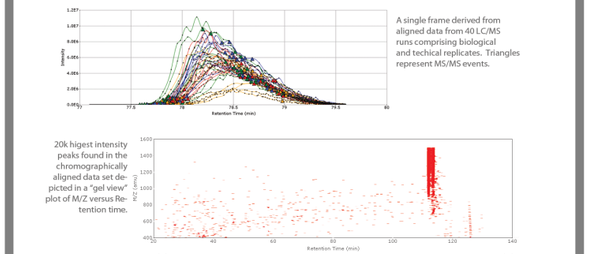



**FIGURE 4. Protein candidates were identified using SEQUEST on a frame-by-frame basis. Selection criteria were based upon a 2% or better false discovery rate calculated by Percolator (3) machine learning algorithm. No Xcorr or other SEQUEST-specific selection criteria were imposed.**

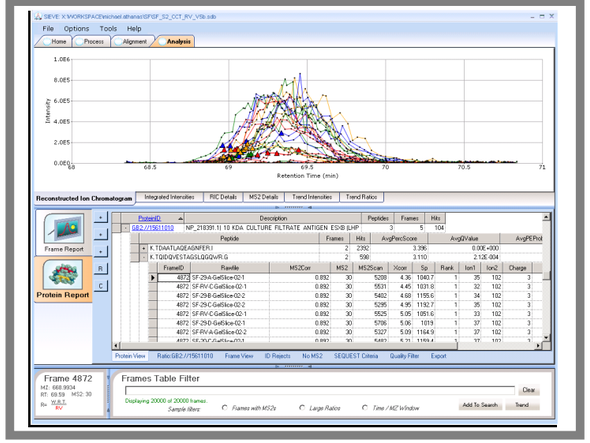



**FIGURE 5. Ratios of wild type to an RD1 restricted strain were calculated. Protein ratios were calculated using variance weighted averaging of each individual peptide measurement.**

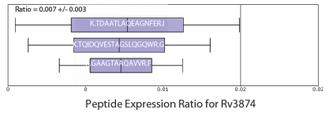
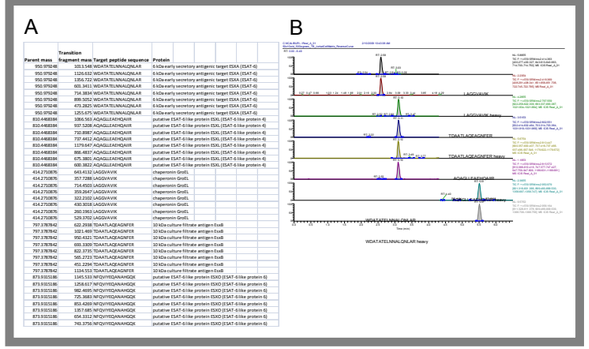



**FIGURE 6. Suppression of several RD1 proteins responsible for attenuated virulence of BCG strain of *Mtb* were confirmed in the SIEVE analysis.**

| Locus | GID | Measured Ratio |
|---|---|---|
| Rv3875 | 57117165 | 0.003 +/- 0.001 |
| Rv3874 | 15611010 | 0.007 +/- 0.003 |
| Rv3616c | 15610752 | 0.001 +/- 0.001 |
| Rv3615c | 15610751 | 0.001 +/- 0.001 |
| Rv3865 | 15611001 | 0.409 +/- 0.060 |
| Rv3870 | 15611006 | 0.001 +/- 0.001 |
| Rv3871 | 15611007 | 0.001 +/- 0.001 |
| Rv3877 | 15611013 | 0.295 +/- 0.064 |

**FIGURE 7. SRM assay of proteotypic peptides from differentially expressed proteins in *M. tuberculosis* mutant strains.**

**The LC-MS/MS data from the SIEVE discovery experiments were mined using novel SRM Workflow software to create a multiplexed SRM method (A) for 5 of the differentially expressed peptides. Heavy labeled internal quantitative standard peptides were synthesized for each target and spiked into each sample before analysis on the triple quadrupole mass spectrometer. Four *Mtb* culture supernatant samples were analyzed in triplicate; an example of the resulting data is shown in (B). Using these methods, the level of detection (LOD) for the peptides was 500 attomoles and level of accurate quantification (LOQ) was 1 femtomole on column. CV's for replicate samples were less than 20% for all target peptides.**




## Conclusions

- High-resolution LC-MS/MS was performed on several strains of *Mycobacterium tuberculosis.*
- SIEVE label-free analysis of the raw mass spectra confirmed the identification of five previously identified secreted proteins as well as other differentially expressed proteins (2, 4) across *Mtb* mutant strains lacking the ESX1 secretion system.
- LC-MS/MS spectra from the discovery experiments were mined using novel SRM Workflow software to develop SRM assays for several of the differentially expressed proteins.
- Quantitative SRM data were obtained for the targeted peptides in cell culture supernatants.

## References

1. Abdallah M. *et. al.*, Type VII secretion – mycobacteria show the way Nature Reviews, 2007 5:883.
2. Fortune SM, Jaeger A, Sarracino DA, Chase MR, Sassetti CM, Sherman DR, Bloom BR, Rubin EJ. Mutually dependent secretion of proteins required for mycobacterial virulence. PNAS, 2005 102:10676.
3. Käll L, Canterbury J, Weston J, Noble WS, MacCoss M. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nature Methods 2007 4:923.
4. Atmakuri K, Fortune SM . Regulation of protein secretion by ... protein secretion? Cell Host Microbe. 2008 4:190.