Poster Reprint

ASMS 2018
MP-090

Urine Proteomics Workflow for the Discovery of Biomarkers in Children with Asthma

Paulos Chumala[1]; Carol Haney-Ball[2]; Tess Kelly[1]; Brooke Thompson[1]; Oluwafemi Oluwole[1]; Anna Afanasieva[1]; Donna Rennie[1]; Joshua Lawson[1]; George S. Katselis[1]

[1]CCHSA/Medicine, University of Saskatchewan, Saskatoon, SK, Canada; [2]Agilent Technologies, Santa Clara, CA

Introduction

- Asthma can be difficult to diagnose as there is no gold standard and presenting phenotypes differ. A number of illnesses result in asthma-like symptoms and because asthma can affect very young children non-invasive sampling is desirable.
- This research is focused on the identification of proteins which may be useful for class prediction of asthma phenotypes. Urine has been shown to be an appropriate matrix for studying metabolites and proteins linked to disease (1-4) while other matrices have also been used to study proteins affected by asthma (5).
- In this study, we use MS-based urine proteomics as a tool in phenotyping asthma, asthma-like and control groups. The *specific objectives* are: 1) to use MS-based proteomics to identify urine protein markers; 2) to compare urine protein marker profiles among groups; 3) to use multivariate statistical analysis to build a predictive model.

Experimental Workflow

Data & Sample Collection (University of Saskatchewan)

- 155 participants, 6-14 years of age
- Surveys were completed by the child's parent and urine samples were collected in the field by trained U of S CCHSA research personnel
- Clinical assessment (spirometry, prick test, and exercise challenge testing)
- Children are classified into a diagnostic group based on their survey response: a) previous diagnosis of asthma; b) asthma-like symptoms; c) no asthma

Sample Processing (University of Saskatchewan)

- Urine samples were collected and stored at -20°C
- Samples were processed in Dr. Katselis' lab by centrifugation, along with the addition of protease and phosphatase inhibitor cocktail, aliquoted, and stored at -80°C
- Creatinine concentration determination by QuantiChrom™ Creatinine Assay Kit
- Samples were concentrated using a 1K MWCO device
- Total protein concentration determination by BCA assay
- Protein concentration normalized to 5 mg/mL prior to in-solution digestion

Experimental

Mass Spectrometry Conditions

Analysis was performed using an Agilent 6550 QTOF mass spectrometer with a Jet Stream source. All samples were analyzed in MS only mode and random samples were additionally analyzed in a data dependent MS/MS mode.

Table 1: MS Conditions		
Gas Temp	250°C	
Gas Flow	14 L/min	
Nebulizer	35 psi	
Sheath Gas Temp	250°C	
Sheath Gas Flow	11 L/min	
Nozzle Voltage	0 V	
Vcap	3700 V	
MS mass range	100-1700 m/z	
MS spectra/sec	8	
MS/MS mass range	50-1700 m/z	
MS/MS spectra/sec	3	
Precursors per spectrum	20	
Ramped CE per 100 Dalton	slope	offset
Collision Energy z=2	3.1	1
Collision Energy z=3 & >3	3.6	-4.8
Isolation width	Narrow	

Chromatographic Conditions:

Column: Poroshell, 120 EC-C18, 2.1x100 mm, 2.7 μm
Injection Volume: 8 μL
Flowrate: 200 μL/min
Mobile Phase A: 0.1% formic acid in water
Mobile Phase B: 0.1% formic acid in acetonitrile
Stop time: 63 min; Post time: 10 min

Table 2: Timetable (minutes)	%B
0	3
45	30
55	60
59	95
62	95
63	3

MS Data Analysis

- 45 samples were only analyzed (15 per each group)
- Total spectral intensity of proteins normalized based on protein : creatinine ratio
- Spectrum Mill, MPP and Pathways Architect Software

Results and Discussion

- Reproducibility is key to being able to differentiate biological variation versus instrument or sample preparation.
- To help achieve this, LCMS data was acquired at standard flow rates in MS only. Every 5th injection was a pooled QC sample (technical replicates) and every 10th injection one of the samples was analyzed in MS/MS mode.
- Figure 1 shows the PCA of unfiltered entities. The pooled QC samples cluster tightly indicating excellent instrument reproducibility. Without filtering of entities, no clustering of biological groups is observed and several outlier samples are indicated.
- Even with quality control filtering for frequency and removing the low abundance and high abundance entities no significant entities are found running an ANOVA with multiple testing hypothesis correction and a p value <0.05. This is due to the high biological variability in this set.
- However, on average it is easy to see that there are entities whose profiles are different between groups as shown in the profile plot in Figure 2.
- An alternate strategy was to filter based on fold change >2. 124 entities are found and of these 45 are identified by their MS/MS spectrum. The hierarchical cluster analysis result is shown in Figure 3.
- Notably protein S100A8, which has been shown (5) to be down in sputum from asthma samples is also down in urine from asthma. Closely clustered with protein S100A8 are several Ig kappa peptides, hemopexin; all down compared to the control.

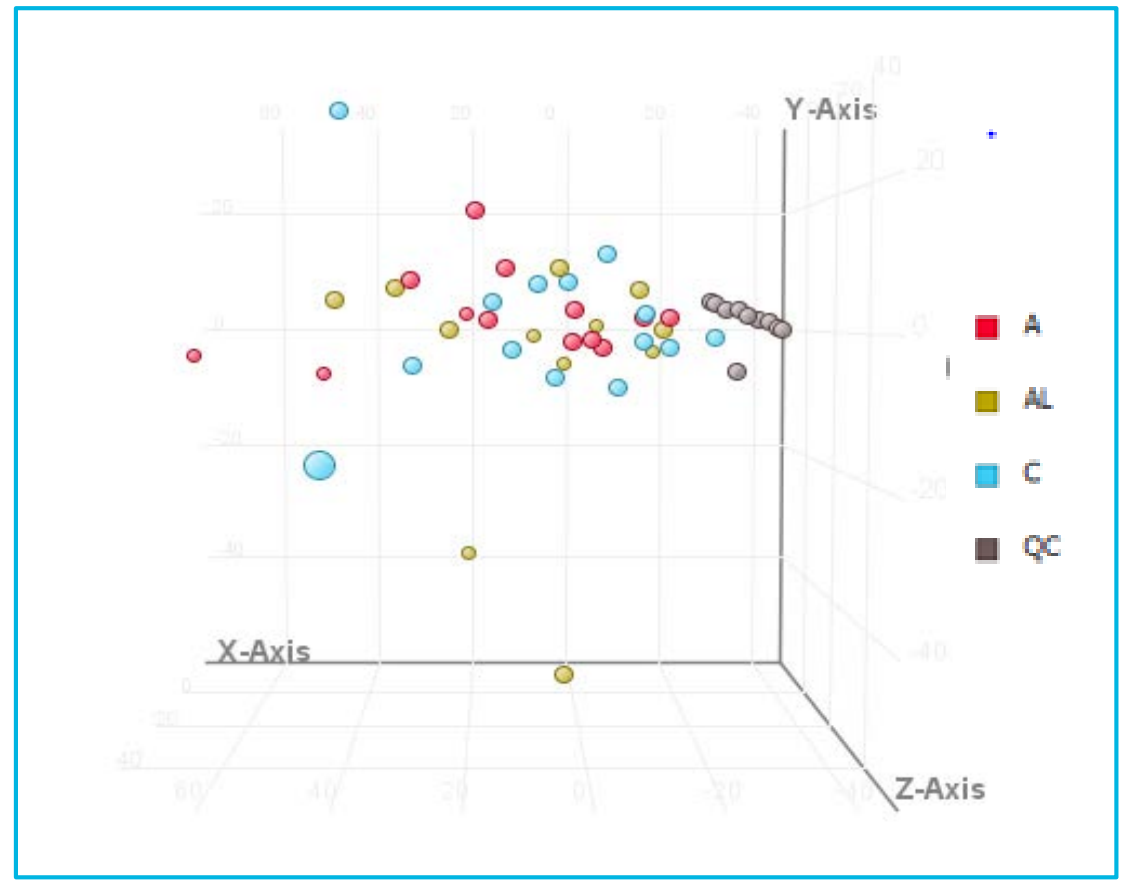


Figure 1: PCA plot of top 2000 entities to evaluate reproducibility of instrument performance. No clustering of biological groups is observed but excellent grouping of the pooled QC is seen.

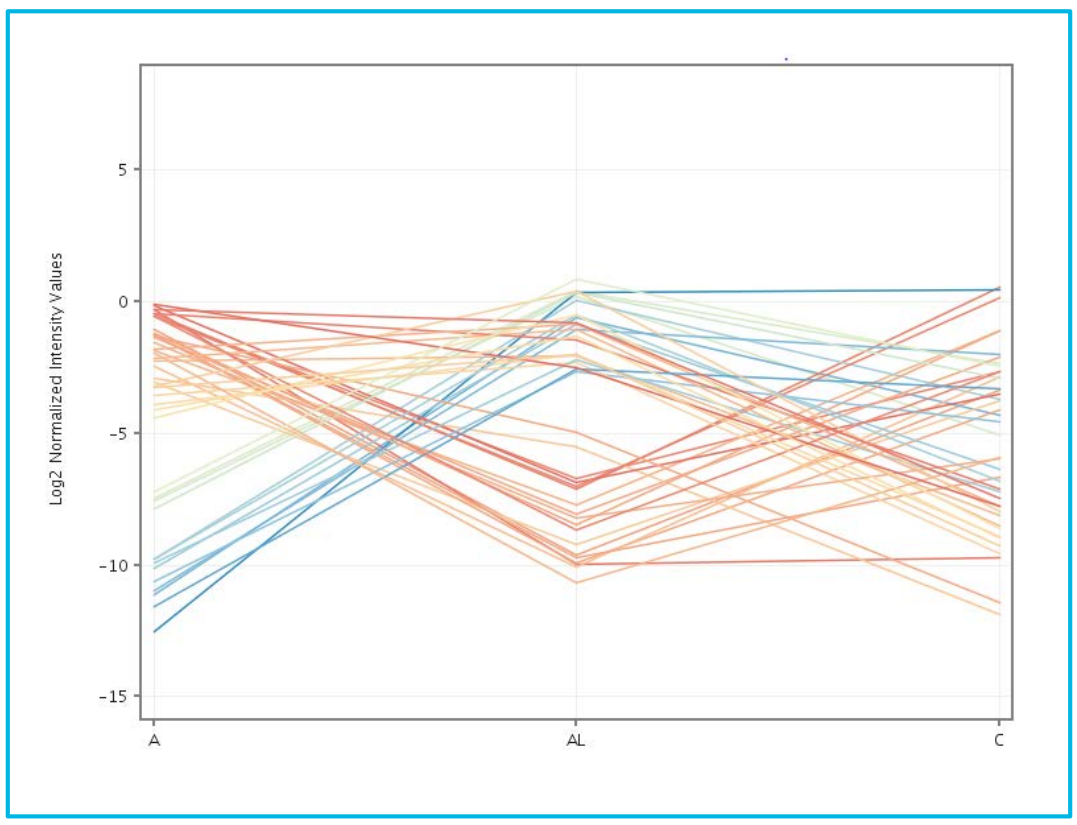


Figure 2: Profile plot of 43 entities filtered for Frequency of >80% any one group, mass >800, high pCor and pCoVar on the S-Plot of the PCA. On average it is clear that between control and asthma and asthma-like there are trends. These entities will be used for classification testing.

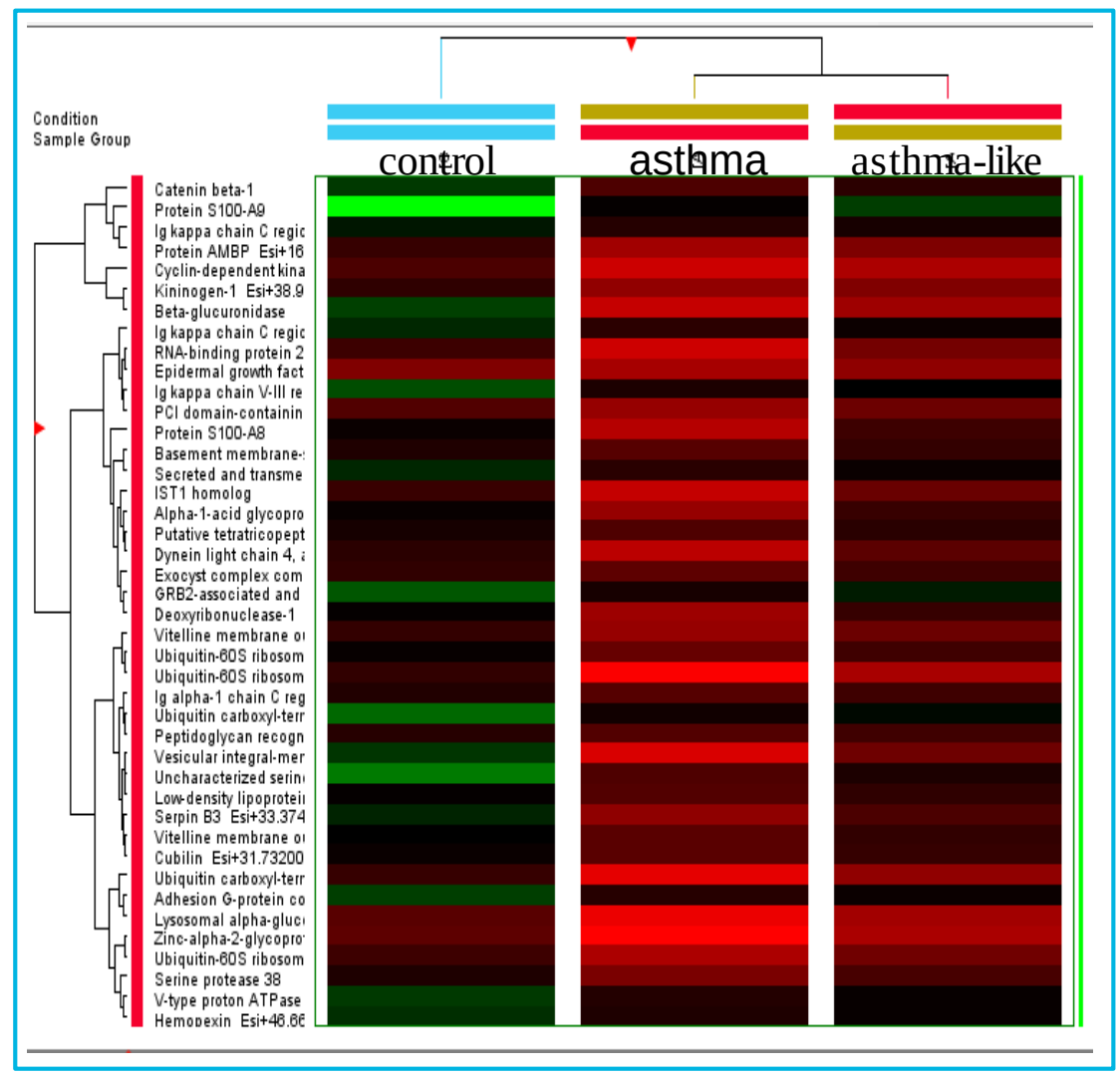


Figure 3: Hierarchical cluster of the 45 entities which are identified in Spectrum Mill Proteomics search engine of the MS/MS data and matched based on RT, mass, and MS/MS spectra. These identified entities were in the cohort of 124 entities exhibiting >2 fold change down in asthma or asthma-like compared to control.

Results and Discussion

Multivariate Analysis

- Ultimately the goal is to be able to differentiate all three groups, so multivariate methods were performed to see if an effective model could be built from this small data set.
- Because there were no statistically significant entities to rely on, a correlation/covariance plot (S-plot) was used to manually select the entities most responsible for driving the variation. These 43 entities are plotted in Figure 2 and used for PLS-DA (Figure 4).
- The model has reasonable accuracy for such a small data set which has large variation within groups and this is reflected in the goodness of prediction Q^2 of 0.28 on component 2 and a sum of two component R^2 of 0.63.

Pathways Analysis

- Pathway architect revealed a number of pathways for the differentially expressed proteins. One example is the inflammatory response pathway, where a network of proteins associated with asthma is shown in Figure 5. Fibronectin has been found to differentiate asthmatic individuals undergoing allergen inhalation challenge (6) whereas IL-4 is known to have a strong influence in childhood asthma (7).

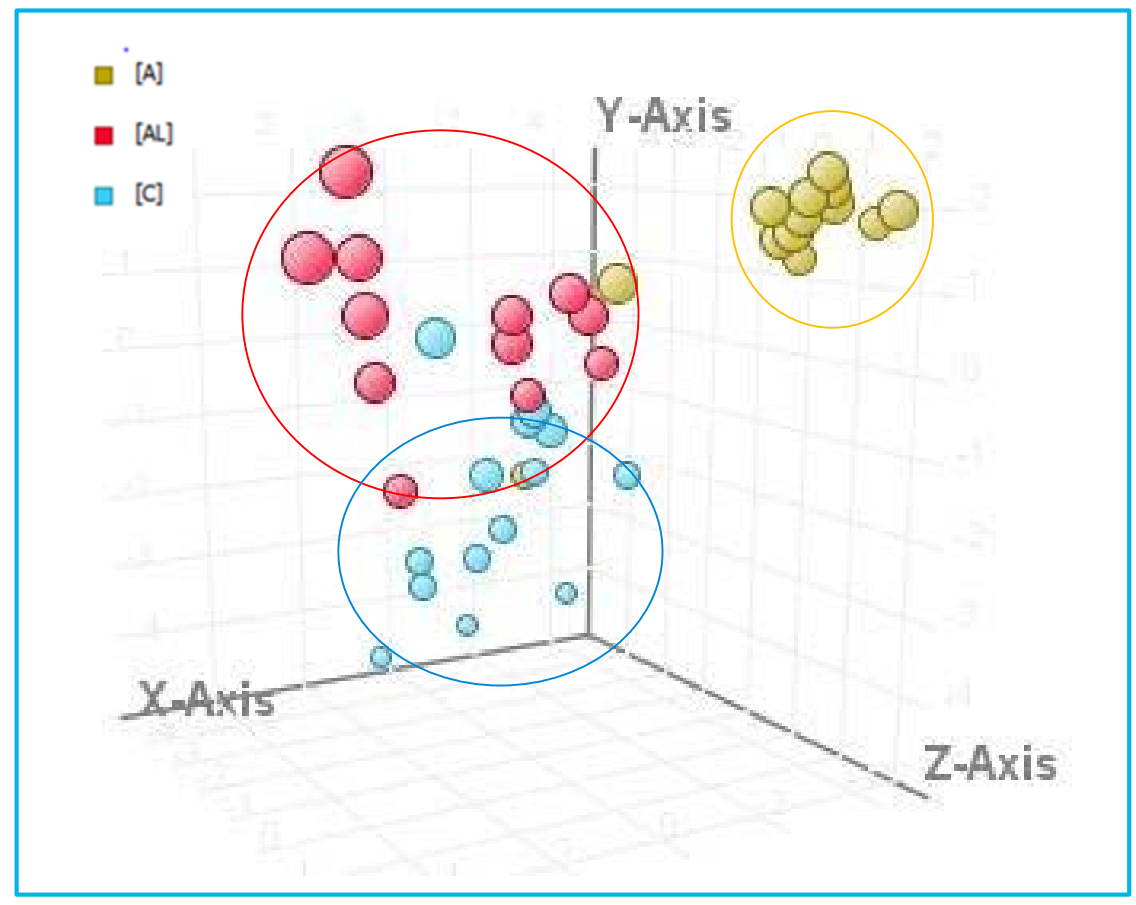
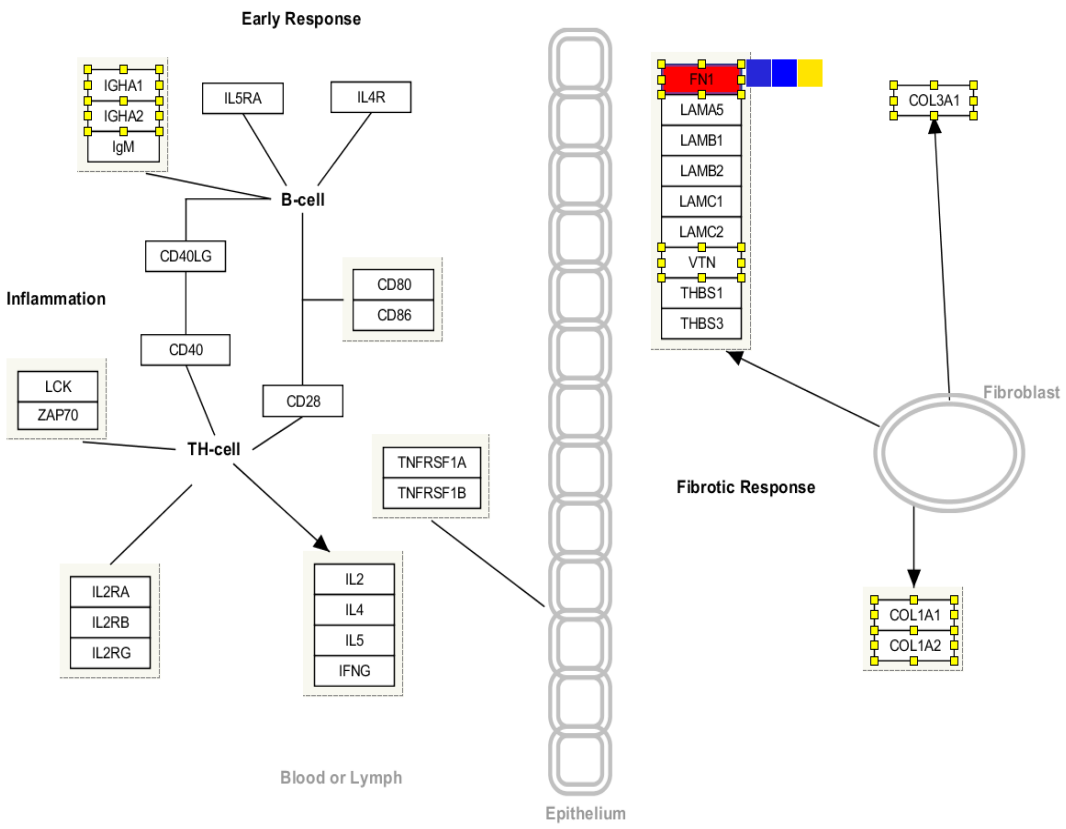


Figure 4: PLS-DA model shows good classification of the three groups. Overall accuracy of the model is 89.7% with classification accuracy of asthma and control at 92.3% and 92.9% respectively. Asthma-like remains the most challenging with an accuracy of 83.3 on the test set.



Name	DB	DB ID	A: Log2	AL: Log2	NA: Log2
COL1A1	Swiss-Prot ID	P02452	13.748	12.188	15.985
COL1A2	Swiss-Prot ID	P08123	14.133	12.365	16.267
COL3A1	Swiss-Prot ID	P02461	14.449	12.018	16.853
FN1	Swiss-Prot ID	P02751	11.245	10.783	14.14
VTN	Swiss-Prot ID	P04004	12.288	11.59	13.359
IGHA2	Swiss-Prot ID	P01877	12.046	11.695	13.851
IGHA1	Swiss-Prot ID	P01876	12.046	11.695	13.851

Figure 5: HS Inflammatory Response Pathway WP244_85199, as revealed by Pathways Architect analysis.

Conclusions

- ✓ Standard flowrate HPLC combined with the Jet Stream source shows excellent reproducibility and robustness for proteomics profiling studies.
- ✓ Urine is a challenging matrix and normalization to total protein and creatinine concentrations is necessary when working with biological replicates.
- ✓ Using clustering algorithms on the averaged group data we were able to find 43 proteins which show >2 fold change in the asthma and asthma-like groups compared to control.
- ✓ Using entities which have high pCor and pCoVar we were able to build a class prediction model; this will be tested against a much larger dataset, a work which is in progress.
- ✓ Pathways analysis will enable us to create a list of target proteins for future quantitative studies.

References

1) Mattarucchi, E. et al., Biomed. Chromatogr., 2012, 26, 89-94.
2) Park, Y.H., et al., J. Allergy Clin Immunol., 2017, 139(5), 1518-1523.
3) Kentsis, A., et al., Proteomics Clin Appl., 2009, 3(9) 1052-1061.
4)Pang, J.X., et al., J. Proteome Res., 2002, 1(2), 161-169.
5) Gharib, S.A., et al., J. Allergy Clin. Immunol., 2011, 128, 1176-1184.
6) Viglio, S., et al., Proteomes, 2014, 2, 18-52.
7) Kabesch, M., et al., J. Allergy Clin. Immunol., 2006, 117, 269-274.

Agilent
Trusted Answers