Large-Scale Quantitative Proteomic Analysis Identifies Multiple Pathways In COPD-Associated Lung Cancer

B. Sandri, PhD[1], A. Limper, MD[2], P. Jagtap, PhD[4], Y. Ping, MD[2], C. Murie, PhD[3], O. Larsson, PhD[3], P. Bitterman, MD[1], T. Griffin, PhD[4], L. Higgins, PhD[4], T. Markowski, BS[4], C. Wendt, MD[1]
[1]University of Minnesota Minneapolis, MN/US, [2]Mayo Clinic Rochester, MN/US, [3]Karolinska Institute Solna/SE, [4]Center for Mass Spectrometry and Proteomics, St. Paul, MN/US

Introduction

- Chronic Obstructive Pulmonary Disease (COPD) and Lung Cancer affect 25 million Americans.
- Clinical reports support COPD as a risk factor for lung cancer, independent of smoking.
- The underlying mechanisms that predispose COPD patients to lung cancer remains unknown.
- Developed comprehensive analytical workflows capable of data normalization across 13 iTRAQ runs.

Goal: Identify molecular pathways in the etiology of lung cancer associated with COPD using wide-scale 8-plex iTRAQ with MS/MS along with a highly-customized workflow and data processing using the GalaxyP platform.

Methods

Samples:

80 peripheral lung tissue samples (non-malignant) spanning eight disease categories (Table 1) snap frozen at the time of lung resection obtained from the NIH Lung Tissue Consortium Clinical Center and Mayo Clinic

Tissue processing:

Lung tissue (100 mg) washed in a PBS solution were placed in Eppendorf SafeLock™ tubes containing approximately 75µL 2.0mm zirconium oxide beads, 50µL 1.0mm zirconium oxide beads, 50µL 0.5mm zirconium oxide beads with 6X sample volume lysis buffer consisting of 0.5M TEAB, 7M urea, 2M thiourea, 20% methanol and 4mM TCEP. The Bullet Blender Storm™ bead mill homogenizer (Next Advance Averill Park, NY) was then utilized to homogenize the sample at 4º Celsius at maximum agitation for 10 minutes. Thorough lysis and membrane disruption was achieved through use of a Barocycler® NEP2320 (Pressure Biosciences South Easton, MA) capable of applying a maximum hydrostatic pressure of 35,000 psi. The samples underwent thirty cycles at 36º Celsius 35,000 psi for 30 seconds and ~0 psi for 10 seconds. MMTS was added to a final concentration of 8mM

Labeling and Detection:

- Randomized sample list was used to determine run order with two of the eight possible iTRAQ labels reserved for pooled mastermix.
- All processed samples were labeled with iTRAQ 8-plex reagent according to manufacturer's protocol (ABSciex Framingham, MA.)
- Raw files obtained directly from the Orbitrap Velos Mass Spectrometer were imported into **GalaxyP** for further processing. All raw files were converted to .mgf files and a Protein Pilot 4.5 search was performed with a custom database generated from the human UniProt database and ABSciex contaminant database.
- False discovery rate (FDR) analysis employed the **GalaxyP** framework using highly-reproducible, robust, and easily-shared workflows. Proteins with a 1% FDR for each iTRAQ sample were calculated.
- Statistical workflows were developed entirely within the **GalaxyP** environment **http://usegalaxyp.org.**
- Quantitative protein expression values with corresponding p-values conforming to 1% FDR or lower were analyzed using Ingenuity Pathways Analyses (IPA) Ingenuity® Systems, Qiagen, Valencia, CA

Methods Continued

Disease Classification	Smoker vs. Never-Smoker	Patient Sex	Average Age (Age Range)	FEV1 Average (Range)
GOLD 0-1, No Cancer	Never Smoker	2M/8F	63.2 ± 14.6 (37-85)	102.6 ± 15.2 (84-132)
GOLD 0-1, Cancer	Never Smoker	7M/3F	61 ± 7.8 (44-68)	96.5 ± 15.4 (77-120)
GOLD 0-1, No Cancer	Smoker	4M/6F	61.7 ± 10.9 (43-79)	93.9 ± 10.1 (82-116)
GOLD 0-1, Cancer	Smoker	8M/2F	56.3 ± 5.9 (47-64)	96.3 ± 15.2 (79-127)
GOLD 2, No Cancer	Smoker	3M/7F	68.4 ± 12.8 (45-84)	63.2 ± 8.7 (52-77)
GOLD 2, Cancer	Smoker	7M/3F	70.3 ± 9.6 (52-83)	54.2 ± 9.0 (39-73)
GOLD 3-4, No Cancer	Smoker	7M/3F	72.2 ± 10.5 (48-84)	15.2 ± 3.7 (10-22)
GOLD 3-4, Cancer	Smoker	4M/6F	67.7 ± 8.6 (53-82)	33.2 ± 9.9 (19-47)

Table 1: Disease categories and patient demographics. Ten lung tissue specimens were enrolled from each disease category.

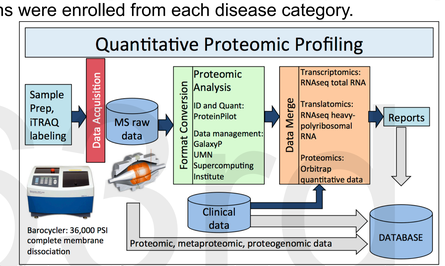

Figure 1: Custom workflow designed to interrogate quantitative proteomic data in a highly repeatable and transparent manner.

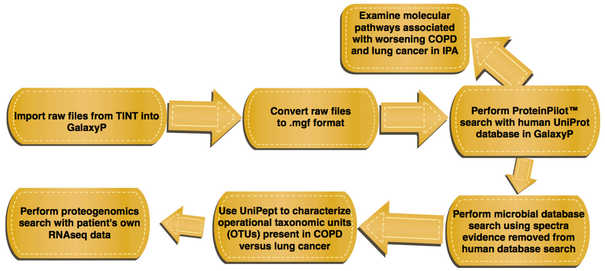

Figure 2: Workflow diagram examining proteomic, metaproteomic, and proteogenomic data derived from patient's RNAseq within the **GalaxyP** environment.

Results

Summary of Alignment Quality

Protein Set	Master	1	2	3	4	5	6	7	8	9	10
Proteins in Set	2812	1448	2266	1999	2055	2113	1989	1805	1762	1682	1682
Set Proteins Aligned to Master		1361	2138	1908	1882	2007	1885	1658	1624	1548	1555
Set Proteins Not Aligned to Master		87	128	91	173	106	104	147	138	134	127
% of Set Proteins Aligned		*94.0%*	*94.4%*	*95.4%*	*91.6%*	*95.0%*	*94.8%*	*91.9%*	*92.2%*	*92.0%*	*92.4%*
% of Set Proteins Not Aligned		*6.0%*	*5.6%*	*4.6%*	*8.4%*	*5.0%*	*5.2%*	*8.1%*	*7.8%*	*8.0%*	*7.6%*

Table 2: Number of proteins, per iTRAQ run, that were consistently detected at a quantitative level across multiple experiments.

Results Continued

Gene Symbol	Protein	CancerFev1 Qval
CAV1	Caveolin 1	0.067
PTRF	Polymerase I and transcript release factor	0.067
GNAS	G-protein alpha subunit	0.092
MMRN2	Multimerin 2	0.092
TLN1	Talin-1	0.092
VCL	Vinculin	0.092
WFS1	Wolframin	0.092
RPL17	60S ribosomal protein L17	0.108
SF3B2	Splicing factor 3B subunit 2	0.108
SORBS1	Sorbin	0.168

Table 4: Top 10 proteins that were statistically different between lung cancer and COPD samples without lung cancer in smokers. Plotted against FEV1 values these proteins demonstrated a positive coefficient of correlation with lung cancer positive compared to lung cancer negative samples.

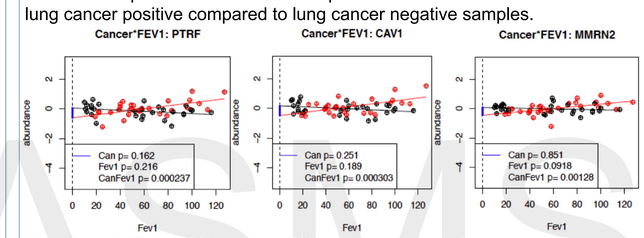

Figure 2: PTRF, CAV1, and MMRN2 demonstrate statistically significant positive coefficient expression profiles between lung cancer positive (red) and lung cancer negative (black) samples. The custom built NOMAD R module was used to generate normalized quantitative values (graphed as abundance, y-axis) against patients FEV1 score (x-axis).

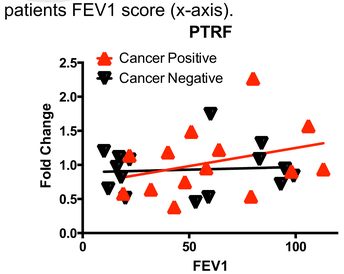

Figure 3: Independent validation of NOMAD derived quantitation through iTRAQ protein ratios. PTRF demonstrates a similar expression pattern using ABSciex determined protein ratios.

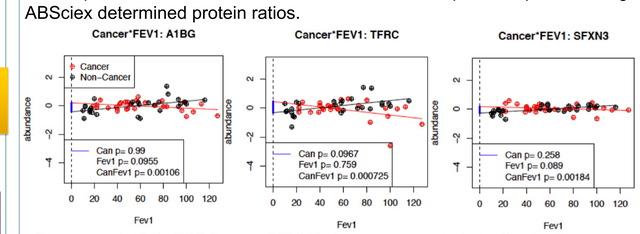

Figure 4: A1BG, TFRC, and SFXN3 demonstrate statistically significant negative coefficient expression profiles between lung cancer positive (red) and lung cancer negative (black) samples.

Results Continued

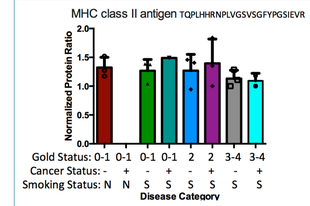

Figure 5: A novel proteoform with an alternate start site was discovered in three separate iTRAQ runs searching against a patient-derived RNA-seq translated database.

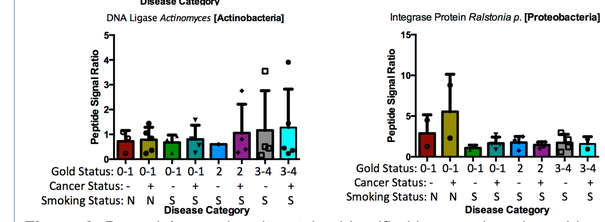

Figure 6: Bacterial-associated proteins identified in unassigned peptides remaining after human protein database searches.

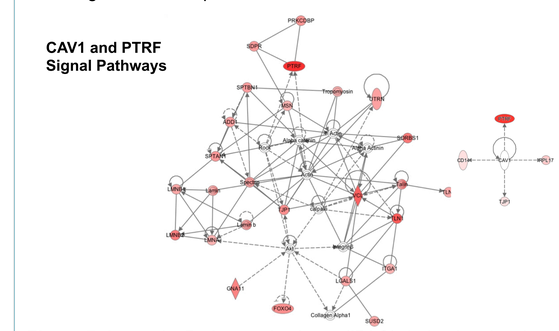

Figure 7: Ingenuity Pathways Analyses (IPA®) derived molecular pathways perturbed in lung cancer patients with severe COPD. Genes in red demonstrated direct involvement with darker shades having a higher z-score.

Conclusions and Future Directions

- Approximately 2,000 proteins were quantitatively detected in multiple iTRAQ runs characterizing COPD and lung cancer sample proteomes
- In a parsimonious approach we are able to profile the microbial proteome that is consistent with our previously published data. PubMed PMID:23071781; PubMed Central PMCID:PMC3460539.
- The number of identifications from both proteogenomic and metaproteomic databases is limited yet consistent across multiple replicates.
- Gene ontology demonstrates differentially expressed proteins represent aberrant cell movement, cell signaling and cell morphology.
- Cav1 and PTRF are involved in caveolae formation and are highly dysregulated in patients with severe COPD and lung cancer.
- Putative upstream regulators include ERBB2, ERG, MYCN, Tropomyosin and TGFβ1 in lung cancer with advanced COPD.

Acknowledgement: This work was supported in part by: NIH T32 HL07741 (BS, CW), **NIH R01HL107612** (CW, TG, AL, YP, OL, PB) and DBI-ABI award NSF 1147079 (TG). Computational support provided by Bart Gottschalk and James Johnson of Minnesota Supercomputing Institute (MSI).

Department of Medicine
University of Minnesota
Driven to Discover℠
Division of Pulmonary, Allergy, Critical Care and Sleep