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

B. Sandri, PhD1, A. Limper, MD2, P. Jagtap, PhD3, Y. Peng, MD2, C. Murie, PhD3, O. Larsson, PhD3, P. Bitterman, MD1, T. Griffin, PhD4, L. Higgins, PhD4, T. Markowski, BS4, C. Wendt, MD1
1University of Minnesota Minneapolis, MN/US, 2Mayo Clinic Rochester, MN/US, 3Karolinska Institute Solna/SE, 4Center 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 MiMS 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 with 6X sample volume lysis buffer consisting of 0.5M TEAB, 7M urea, 2M thiourea, 20% methanol and 4mM TCEP. The Bullet Blender SafeLock™ tubes containing approximately 75 µL 2.0mm 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) at a quantitative level across multiple experiments.

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 were performed directly from the Orbitrap Velos Mass Spectrometer and 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://usegalaxy.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

Results

- 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. Published PNAS Early Edition.
- 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.
- Protein-signature regulators include ERBB2, ERG, MYCN, Tropomyosin, and TGFβ1 in lung cancer with advanced COPD.

Acknowledgement: This work was supported in part by: NH T32 HL07741 (CW, TG), NIH R01HL107612 (CW, TG, AL, YP, OL, PB) and DBI-ABI award NSF 1147079 (TG).

Conclusions and Future Directions

- Top 10 proteins that were statistically different between all 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.
- Top 10 proteins that were statistically different between all 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.
- Top 10 proteins that were statistically different between all 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 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.

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

Figure 5: A novel proteoform with an alternative start site was identified in the lung cancer positive samples.

Figure 6: iTRAQ-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.

Figure 7: Ingenuity Pathways Analyses (IPA®) derived molecular pathways in COPD-associated lung cancer.

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

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

Table 3: Number of proteins, per iTRAQ run, that were consistently detected at a quantitative level across multiple experiments.

Table 4: Disease categories and patient demographics. Ten lung tissue specimens were enrolled from each disease category.

Table 5: Number of proteins, per iTRAQ run, that were consistently detected at a quantitative level across multiple experiments.

Table 6: Number of proteins, per iTRAQ run, that were consistently detected at a quantitative level across multiple experiments.

Table 7: Number of proteins, per iTRAQ run, that were consistently detected at a quantitative level across multiple experiments.

Table 8: Number of proteins, per iTRAQ run, that were consistently detected at a quantitative level across multiple experiments.