# Protein Expression Profile of Rat Type Two Alveolar Epithelial Cells During Hyperoxic Stress and Recovery

A THESIS
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Maneesh Bhargava

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Lynda BM Ellis, Vipin Kumar, David H Ingbar

May 2013

## Acknowledgements

I would like to express my gratitude to Drs. Ellis, Kumar and Ingbar for their supervision and guidance through out my Masters program. I also am extremely fortunate to have Dr. Ingbar and Dr. Wendt as my mentors. Their support has been critical in all my academic endeavors. It has been an honor to work with extremely dedicated collaborators and laboratory staff, without whom this work could not be completed.

I am most grateful to my family, my wife, Mukta, and my children, Divya and Rishi for their support and love.

## Dedication

This thesis is dedicated to my father whose memories and spirit of enterprise will always inspire me

## Abstract

Rationale: In rodent model systems, the sequential changes in lung morphology resulting from hyperoxic injury are well characterized, and are similar to changes in human acute respiratory distress syndrome (ARDS). In the injured lung, alveolar type two (AT2) epithelial cells play a critical role restoring the normal alveolar structure. Thus characterizing the changes in AT2 cells will provide insights into the mechanisms underpinning the recovery from lung injury.

Methods: We applied an unbiased systems level proteomics approach to elucidate molecular mechanisms contributing to lung repair in a rat hyperoxic lung injury model. AT2 cells were isolated from rat lungs at predetermined intervals during hyperoxic injury and recovery. Protein expression profiles were determined by using iTRAQ® with tandem mass spectrometry.

Results: Of 959 distinct proteins identified, 183 significantly changed in abundance during the injury-recovery cycle. Gene Ontology enrichment analysis identified cell cycle, cell differentiation, cell metabolism, ion homeostasis, programmed cell death, ubiquitination, and cell migration to be significantly enriched by these proteins. Gene Set Enrichment Analysis of data acquired during lung repair revealed differential expression of gene sets that control multicellular organismal development, systems development, organ development, and chemical homeostasis. More detailed analysis identified activity in two regulatory pathways, JNK and miR 374. A Short Time-series

Expression Miner (STEM) algorithm identified protein clusters with coherent changes during injury and repair.

Conclusion: Coherent changes occur in the AT2 cell proteome in response to hyperoxic stress. These findings offer guidance regarding the specific molecular mechanisms governing repair of the injured lung.

## Table of Contents

## List of Tables

# List of Figures

## Introduction

Healthy rats and mice exposed to 100% oxygen develop respiratory failure and death within 2-3 days (1, 2). There is evidence that hyperoxia is damaging to the human lung as well (3-5). Histological examination of lungs from these animals shows pathological changes similar to those that occur in human Acute Respiratory Distress Syndrome (ARDS), with evidence of apoptosis and necrosis of the alveolar endothelium and epithelium (1, 6).

Rodent lung morphology has been well characterized during acute oxygen-lung toxicity and subsequent recovery (1, 7, 8). In animals, exposure to normobaric hyperoxia for 72 hours results in an exudative phase that is characterized by alveolar type I (AT1) epithelial cell death; swelling and necrosis of endothelial cells; interstitial edema; and filling of alveoli with an exudative fluid (4, 7). Cessation of oxygen exposure after 60 hours leads to a rapid decrease in the number of neutrophils within the first three days and an increase in the number of monocytes and lymphocytes, resulting in a normal differential of lung white blood cells by day seven (8). These sequential changes in lung morphology parallel those observed in human acute lung injury (ALI), making this a good animal model for studying mechanisms of lung repair following ALI (8, 9).

We utilized this model to elucidate the underlying molecular mechanisms governing the role of alveolar type 2 epithelial (AT2) cells in lung repair. *In vivo* exposure of adult animals to hyperoxia selectively kills AT1 cells, leaving a denuded basement membrane (7, 10). In response, AT2 cells function as

regenerative cells (1, 2, 11); to renew the alveolar epithelium by proliferating and differentiating into AT1 cells (12). However, AT2 cells are also targets of oxygen-induced injury. Decreased surfactant production and other changes reflecting a compromised metabolic state are seen in AT2 cells after 64 hours of *in vivo* oxygen exposure (13). How AT2 cells initiate repair in the setting of hyperoxic injury remains unknown.

Broad changes in gene expression have been documented in adult rat AT2 cells during *in vivo* hyperoxia injury. Specifically, changes were seen in genes annotated to biological processes such as cell cycle, differentiation, proliferation, and telomerase (14). These findings suggest active alveolar repair during hyperoxic exposure and recovery. Still unknown, however, is how these broad changes in gene expression translate to changes in specific protein expression patterns.

We hypothesize that lung repair is governed by highly organized changes in the alveolar epithelial cell proteome. To test this, we used mass spectrometry based proteomics to characterize changes in the rat AT2 cell proteome across multiple time points of injury and recovery in an experimental model with exposure to > 95% oxygen.

## Methods

### *Hyperoxia exposure and recovery protocols*

The University of Minnesota Institutional Animal Care and Use committee approved all experimental protocols using animals. Specific pathogen free male Sprague Dawley rats between 175 and 199 grams were exposed to > 95% $O_2$ at 1 atmosphere for 60 hours with *ad libitum* access to food and water as previously described (1, 8, 15). An exposure time of 60 hours provides a nonlethal dose in rats that damages the alveolar epithelium, with well-characterized morphometric changes (8). Rats were allowed to recover in room air. AT2 cells were isolated from these rats at defined time points during hyperoxia exposure (24, 48, and 60 hours of exposure to > 95% oxygen) and recovery (12, 36, 96, and 168 hours in room air following 60 hours of exposure >95% oxygen). AT2 cells isolated from rats continuously housed in room air were used as controls and protein abundance was measure relative to these cells.

### *Isolation of AT2 cells*

We used a standard elastase digestion technique to isolate AT2 cells (15-17). To obtain an adequate protein concentration we pooled samples from two rats at each time point. Freshly isolated AT2 cells were stored at -80°C until cells from all time points were obtained. To account for biological and technical variability, all experiments were performed in triplicate.

### *Preparation of soluble and insoluble protein fractions*

To increase the number of proteins identified, we fractioned the protein lysate into soluble and insoluble fractions for separate protein analysis. Cell pellets were thawed and incubated on ice for 30 minutes and in hypotonic cell lysis buffer containing 10mM HEPES (pH 7.5), 10mM NaCl, 1.5mM $MgCl_2$, and EDTA-free protease inhibitor (Calbiochem Protease Inhibitor Cocktail Set III, catalogue number 539134) at 1:500 dilution. Cells were lysed with 20 strokes in a Dounce homogenizer. Unbroken cells and nuclear material were removed by centrifugation at 500 x g for 10 minutes. The supernatant was centrifuged at 110,000 x g for 45 minutes to separate the soluble (supernatant) from the insoluble fraction (pellet), then stored at -80°C prior to proteomic studies. The soluble proteins were precipitated with 100% ethanol by incubating overnight at -20°C followed by centrifugation at 14000 x g. Ethanol precipitation was not needed for the insoluble fraction. Soluble and insoluble protein fractions were suspended in 0.5M TEAB, and the total amount of protein was quantitated by using a Bradford protein assay (Thermo Fisher Scientific, Waltham, MA).

***iTRAQ® labeling and 2D LC-MALDI-MS***

The workflow for protein identification and downstream analysis is outlined in Figure 1. Equal amount of proteins (30 µg) from each time point were processed for trypsin digestion and iTRAQ® labeling with supplied reagents, according to the manufacturer's protocol (Applied Bio systems), with the exception of one sample for which 15 µg of protein was processed for both soluble and insoluble fractions (18). The total peptide mixture was purified with

an MCX Oasis cartridge (Waters, Milford, MA) before separation via two-dimensional liquid chromatography-mass spectrometry (2D LC-MS). LC and MS experimental details were previously reported (19). In brief, proteins were separated offline in the 1st dimension into 14 peptide-containing fractions, collected in 3-minute intervals, on a strong-cation exchange column. Each fraction was separated in the 2nd dimension by C18 reversed phase capillary LC with a Tempo LC-MALDI robotic spotter (AB Sciex). Data dependent acquisition of the top 30 most intense peaks per LC fraction was performed on a 4800 MALDI-TOF/TOF analyzer (AB Sciex).

***Protein database searching and relative quantitation***

We analyzed the tandem mass spectra with Protein Pilot™ version 2.01 software (AB Sciex), which uses the Paragon™ scoring algorithm (20). A list of inferred proteins and relative protein abundances were reported. The ratios of reporter ions provided relative protein abundance among sample types for select proteins. Bias-corrected average protein ratios were calculated from peptide-level ratios. The bias factor is derived from the median average protein ratio for each pair of reporter ions. In ProteinPilot™, the median ratio is adjusted to '1' and all protein ratios are adjusted by the same factor, which corrects the quantitation results for systematic errors in pipetting, mixing, or measuring protein concentration. The protein database used for searches was a concatenated "target-decoy" Rat IPI database with 79970 sequences. Search parameters included the following: 8-plex peptide label; quantitation; bias correction; cysteine

methyl methanethiosulfonate; trypsin enzyme; thorough search mode; biological modifications (includes >220 post-translational and artifactual modifications); and minimum detected protein threshold 95% confidence. All relative quantitation was done using proteins from AT2 cells isolated from uninjured rat lungs as reference. In order to allow for comparison across different iTRAQ ® experiments, the pooled control cells from six uninjured animals were used.

***Computational analysis***

Gene Ontology Enrichment Analysis. Functional annotation clustering analysis was performed by using the online software Database for Annotation, Visualization, and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov) (21). DAVID generates an enrichment score by testing the relatedness of different combinations of genes according to common biologic function, chromosomal location, or regulation. A high enrichment score for a group of genes indicates that annotation term members are playing important roles in a given experiment. An enrichment score of 1.3 is equivalent to a non-log scale p-value of 0.05. The p-value, also termed the EASE score, represents a modified Fisher exact test; the smaller the p-value, the more significant the gene association.

Gene Set Enrichment Analysis (GSEA). GSEA is a computational method that determines whether an *a priori* defined set of genes shows statistically significant, concordant differences between two biological states. Because of multiple hypothesis corrections, very few genes are found to be statistically

significant when investigated individually. GSEA directly tests each pathway and gene set (known from previous biological knowledge) for associations; whereas the constituent genes in the pathway may be weakly associated individually, together they may be more significant.

The Broad Institute provides an easy-to-use Java implementation of the GSEA method on their website (22). The GSEA analysis report highlights enrichment gene sets with an FDR of less than 0.25. We permutated the class labels 100 times and kept the original protein profiles unperturbed to avoid major distortion in the data. The current version of the gene set data base (MSigDB) contains 6,769 genes divided into 5 major collections: C1, positional gene sets; C2, curated gene sets; C3, motif gene sets; C4, computational gene sets; and C5, the GO biological processes.

We performed GSEA by treating the two stages of our experiments – hyperoxia (first 3 time points) and recovery (last 4 time points) – as binary class labels. Therefore, the gene sets/pathways that are reported here are those that changed their profile significantly between hyperoxia exposure and recovery.

Short time-series expression miner (STEM). Using STEM analysis, we investigated how the protein profiles change continuously over time during injury and recovery. The STEM algorithm identifies genes that have similar behavior over a short time series (3-8 time-points) with all genes clustered into one of a set of pre-defined patterns based on transformation of gene profiles into "units of change"(23).

For our analysis, the STEM clustering algorithm was used to create 50 model profiles. STEM uses the hyper-geometric distribution to compute the significance of overlap between genes from the experiment and model profiles. Since the model profiles are defined independent of the data, the boundaries in expression space that they induce remain the same between experiments. Thus, the data from the three separate iTRAQ® LC-MS experiments were analyzed by using the 'repeat data' functionality in STEM.

Default filtering parameters were used except for minimum correlation between repeats, which was set to -1, and minimum absolute expression changes, which were set to 1. The clustering algorithm then assigns each gene passing the filtering criteria to the model profile that most closely matches the gene's expression profile as determined by the correlation coefficient. Since the model profiles are selected independent of the data, the algorithm can then use a permutation test to determine which profiles have a statistically significant higher number of genes assigned to them. This test determines assignments of genes to model profiles by using a large number of permutations of the time points (or columns). It then uses standard hypothesis testing to determine which model profiles have significantly more genes assigned under the true ordering of time points compared to the average number assigned to the model profile in the permutation runs. Significant model profiles can either be analyzed independently or grouped together, based on similarity to form clusters of significant profiles.

***Statistical analysis***

To identify if individual proteins demonstrate a significant change over time, we analyzed the log intensity ratios by applying a linear mixed effects model, which efficiently takes into account the dependence of observations over time for the same animal. Specifically, a one-way ANOVA with random intercept model was fitted to study longitudinal differences in the protein profile over the examined time points (24). After the p-value for each protein was computed on the linear mixed effects model we controlled for an overall false discovery rate (FDR) of ≤10% (25) to identify significantly changed proteins. In addition to the model-based inference of log intensity ratios we explored an alternative novel Short Time-series Expression Miner algorithm that clusters the time series profiles to identify common expression patterns.

Multiple hypothesis correction for controlling for global significance level was done by using FDR (25) that measures expected proportion of false positives among the positive findings. The FDR cutoff was set at 1% (global) for the protein identification in Protein Pilot. As multiple compressions were made in linear mixed effects model, multiple hypothesis correction was performed controlling the FDR at 10%. For the GSEA, an FDR cutoff of 25% was used to identify the differentially expressed gene sets as described in the original publication (22).

## Results

### *Protein identification from iTRAQ® and 2D LC-MALDI-MS experiments*

Specific proteins identified in rat AT2 cells over the experimental time course are listed in Table S1. In the soluble cell fraction, the number of proteins identified at a global FDR of 1% from the three independent iTRAQ experiments were 126, 202, and 204; the total number of distinct proteins for the three experiments combined was 308; and the number of overlapping proteins present in all three separate experiments was 79. In the insoluble fraction, the total numbers of proteins identified at a global FDR of 1% were 202, 342, and 448; the total number of distinct proteins for the three experiments combined was 651; and the number of overlapping proteins present in all three separate iTRAQ experiments was 99. This gave a combined total for the soluble and insoluble fractions for the three-iTRAQ experiments of 328, 546 and 652 proteins and 959 distinct proteins.

These proteins included several of the expected AT2 cell proteins, including surfactant proteins and alkaline phosphatase. Similar to previous reports, levels of surfactant proteins A and B were higher during recovery from lung injury than during injury (13) (data not shown) and the $\beta_1$ subunit of Na,K-ATPase increased in abundance after cessation of hyperoxic stress (Figure 2). Other proteins that participate in maintaining the redox status in the cell, such as glutathione peroxidase, glutathione-s-transferease, thioredoxin and

peroxiredoxin, also showed changes in their abundance during hyperoxic stress and recovery (data not shown).

***Longitudinal differences in the protein profile***

To determine whether the changes in protein abundance over time were similar between separate iTRAQ experiments, a Pearson correlation was calculated for the three time series of a protein. The average Pearson correlation was 0.3 for the soluble fractions and 0.65 for insoluble fractions. Although the absolute degree of fold change for a protein varied between iTRAQ experiments, the direction of change was similar for the majority of proteins. Therefore, we used all unique proteins from all three iTRAQ experiments in further analysis.

The linear mixed effects model identified proteins that demonstrated a significant difference between any time point during injury or recovery. A total of 183 proteins demonstrated statistically significant changes during the injury-recovery time course when controlling for an FDR of ≤ 10% (Table S2). Table 1 shows the ten proteins with the most significant changes (based on FDR) during injury and recovery.

***Gene ontology and gene set enrichment analyses***

To determine the biological relevance of these 183 proteins, we performed gene ontology enrichment analysis. Because one protein could be annotated to multiple biological processes, we used the functional annotation-clustering tool available in DAVID. This reduces the burden of associating similar terms and

focuses the biological interpretation. We found that these proteins were most commonly associated with the following processes: programmed cell death, regulation of macromolecular metabolic processes, ubiquitin-dependent catabolic process, protein complex assembly, cell migration, regulation of immune response, and regulation of cell development (Figure 3, Table S3).

In addition to testing individual proteins, we investigated gene sets that were differentially expressed during oxygen exposure and recovery in room air. For this we performed GSEA analysis, using gene sets from MSigDB and applying the filters described in Methods. This analysis identified two positional (C1), two motif (C3), and six GO gene sets (C5) with differential expression during recovery from hyperoxic stress. The two positional gene sets (C1) were for genes in cytogenetic band chr14q24 (systematic name: M1258) and in cytogenetic band chr9q21 (M17819). The first of the two motif gene sets (C3) was V$AP1_Q6 (M9431), which includes the proteins destrin, aldolase A, transgelin-2, lamin A/C, YWAHZ-ζ, Surfactant protein C, vimentin, and Na,K-ATPase. These genes have promoter regions [-2kb, 2kb] around a transcription start site containing the motif NNTGACTCANN that matches annotation for JUN: jun oncogene. The second motif gene set was TATTATA; MIR-374 (systematic name: M11229), which has three proteins – actinin alpha 4, actinin alpha 1, and Na,K-ATPase – that are targets for microRNA 374(mir-374). Lastly, six Gene Ontology gene sets (C5) were differentially expressed during recovery from

hyperoxic stress (Table 2). Individual GO biological processes in the differentially expressed gene set as included in Table S4.

***Short time-series expression miner analysis***

To address our hypothesis that lung injury and recovery are characterized by temporarily cohesive changes in the AT2 proteome, we used STEM analysis to identify protein groups that are sequentially related in the abundance of proteins in each group. Of the fifty randomly selected profiles investigated, seven showed a statistically significant higher number of proteins than expected (shown as colored profiles in Figure 4). Of these seven, we investigated four profiles (numbered 43- Figure 5A, 7- Figure 5B, 46 Figure 5C, and 47-Fiugre 5D) that exhibited physiologically relevant patterns of change in their protein abundance over the injury- recovery time course .To determine the biological relevance of the proteins in these four profiles we performed GO enrichment analysis of the proteins with the AT2 cell proteome identified in our experiment as the background. Each profile (cluster) enriches distinct biological processes (Table 3). Individual proteins in the clusters and detailed GO enrichment data for each of the 4 profiles are in Tables S5 and S6, respectively.

Profile 43 (Figure 5A) consists of a group of 59 proteins that are decreased during hyperoxia exposure and start increasing during recovery (p-value = $9.7 \times 10^{-18}$). For the majority of these proteins, the trend was a high expression by 36 hours of recovery in room air, with levels close to control values by 96 hours of recovery after hyperoxic stress. GO analysis (Table 3) indicated

that some of these proteins represent a cellular response to stress. More interesting is the contribution of several of these proteins to alveolar epithelial repair mechanism, such as nucleic acid metabolism, regulation of transcription, protein complex assembly and cell proliferation (Table S6). Profile 7 (Figure 5B) contains 51 proteins that increase during the period of oxygen exposure when histologic destruction of the lung occurs (p-value = $4.7 \times 10^{-15}$. As this is during hyperoxic stress, it is not surprising that some of the protein enrichment clusters involve oxido-reduction metabolism and programmed cell death. Profile 46 (Figure 5C) consists of 46 proteins that demonstrate a variable expression during oxygen exposure but are elevated during recovery in room air (p-value $1.7 \times 10^{-10}$). Several of these proteins are involved in the homeostatic response to hypoxia/oxygen, cellular ion homeostasis and response to nutrient levels (Table S6). These proteins also participate in cell motility and adhesion. Profile 47 (Figure 5D) contains 42 proteins that demonstrate a very coherent decrease at 24 hours and then a significant increase at 48 hours of hyperoxic stress (p-value $2.0 \times 10^{-7}$). These proteins represent processes for cellular organization by actin filaments (Table S6).

## Discussion

Overall, our results document sequential changes in the protein expression profile of AT2 cells during hyperoxic stress and subsequent recovery in a rodent model of ALI. State-of-the-art proteomics in combination with novel STEM analysis revealed coherent changes in the protein expression profile during both the injury and recovery phases. Similarly, our GSEA results demonstrate differentially expressed gene sets during recovery from lung injury. Through these approaches – useful for investigating large number of proteins over several time-points during injury and recovery – we identified protein modules that are likely involved in the reparative process in damaged lung. Identifying these proteins in clinical lung injury samples could provide biomarkers of repair in ARDS and new insights into our understating of lung repair mechanisms.

### *Concordance of AT2 proteomics results with previous single-molecule studies*

Previous studies focused on specific genes and/or molecules to understand the mechanism by which oxidative stress results in diffuse alveolar damage (14, 17, 26-28). Our proteomics results confirm previous published reports that specific AT2 cell proteins, such as surfactant proteins (29), respond to hyperoxic stress. Our group previously found that although the mRNA expression of $\beta_1$- Na,K-ATPase subunit increased during hyperoxic stress, the protein level was stable during hyperoxia (17). Similarly, our current study found

that $\beta_1$- Na,K-ATPase subunit protein level did not change during hyperoxic stress, although we were able to demonstrate that higher protein levels occurred during recovery.

***Gene Ontology biological processes involved in hyperoxic injury and recovery***

Repair of the damaged alveolar epithelium is a complex process that requires a coherent sequence of events such as secretion of the extracellular matrix; cytoskeletal reorganization; cell spreading, migration, proliferation, and differentiation (30). Our Gene Ontology enrichment analyses identified numerous biological processes that are consistent with this current paradigm of lung injury and repair. For example, several reparative processes were enriched by proteins that changed during injury/recovery in our study, including processes involved in cell cycle regulation such as RNA splicing and nucleic acid transport, cell differentiation, regulation of cytoskeleton organization, and cell proliferation. In addition processes involved in cellular homeostasis such as protein complex assembly, regulation of endocytosis, and cellular metabolism were also enriched.

Since AT2 cells are immune modulatory, we were not surprised to observe proteins in the acute inflammatory response were regulated. As an example, proteins involved in T cell activation changed abundance during injury/recovery. T cell recruitment occurs in LPS-induced lung inflammation (31). Although T regulatory cells are recruited within 24 to 48 hours after intra-tracheal instillation of LPS, it is felt that lymphocytes do not play a role in development of ALI (32)

but rather participate in its resolution. In our study, proteins annotated to T cell regulatory processes had the highest levels during peak injury and lower levels during repair. The role of these alveolar proteins in regulating T lymphocyte function during hyperoxic stress needs further investigation, but conceivably could be similar to their role in LPS-induced injury.

Ubiquitination and proteolysis are also thought to play critical roles in ALI (33). These processes indirectly affect epithelial barrier function and alveolar edema clearance by targeting Na,K-ATPase (34) and epithelial sodium channels (35). Degradation of Na,K-ATPase through the ubiquitin-proteosome pathway has been studied in the context of hypoxia. Degradation can be prevented by ubiquitin-activating enzyme, lysosomal, and proteomsomal inhibitors (36, 37). While this suggests the participation of proteomsomal and lysosomal pathways, the specific ligases involved are unknown. Epithelial sodium channels are also ubiquinated in a complex manner that results in internalization of the protein followed by either inactivation (38) or activation (39). We identified two protein clusters (Table E3) with proteins annotated to ubiquitination that changed expression during injury/recovery cycle in AT2 cells. These proteins included small ubiquitin-related modifier 2; proteasome activator unit 1; proteasome subunit alpha 6; LOC 679794; LOC682626 (similar to cytochrome C); proteasome 26S subunit ATPase 2; heterogeneous nuclear ribonucleoprotein D; non-muscle myosin heavy chain 9; ER degradation enhancer; mannosidase

alpha-like 1; and RhoA. Further investigation into the role of these proteins could provide insight into the role of ubiquitination in lung injury. .

The question of whether apoptosis is central to epithelial damage in hyperoxic stress remains unanswered, but several mechanisms thought to be responsible for cell death in hyperoxic lung injury have been proposed. Exposure of cells to various stimuli may trigger activation of extrinsic (surface receptor pathways) or intrinsic (mitochondrial) pathways, both leading to caspase activated programmed cell death (40). Several pathways involving signaling though cell death receptors have been implicated during hyperoxic lung injury. However, cell surface receptor pathways such as TNFα (41), Fas (28), and CD40 (42) mediated signaling do not appear to play a central role in hyperoxic lung cell death. Mitochondria-dependent cell death could play a role through the Bcl-2 protein family (43, 44). Several of the proteins identified in our study (e.g., PTPRC, YWHAB, YWHAE, KRT18, CD59, RHOA, GSN) had GO biological process annotation of extrinsic apoptotic pathways. We also found proteins that participate in the intrinsic apoptosis pathway (GPX1, NQO1, LOC679794, LOC682626). Notably, HSPB1 had a GO annotation of negative regulation of apoptosis. Together, these data suggest that *in vivo* hyperoxic stress results in cell death by complex mechanisms. Targeting pathways involved in cell death remains an attractive target for therapy in ALI.

Cell migration is another understudied but potentially critical aspect of lung repair. We found that AT2 cell proteins annotated to cell migration demonstrated

significant changes in their expression level during hyperoxia and recovery. These proteins included those active in actin cytoskeleton assembly (coflin, moesin, vinculin, tropomyosin 1, MYH9, MYH10, vinculin, thymosin β4), integrin signaling (CD47), GTPase activity (rhoA), and signal transduction (S100A9). Although these proteins participate in cell migration in other organs, until now they have not been directly implicated in alveolar epithelial cells suggesting a new mechanism and potential therapeutic target.

***Gene Sets that participate in recover from hyperoxic stress***

The GO gene sets differentially expressed in our model were annotated to cell development, cell proliferation, and organogenesis. These GSEA findings complement our results from the GO enrichment analysis of our linear mixed effects models. We also identified two differentially expressed motif gene sets, V$AP1_Q6 and MIR-374. These motifs may represent a common mechanism by which gene sets are regulated during injury and repair. The motif V$AP1_Q6 contains genes with promoter regions [-2kb, 2kb] around transcription start site containing the motif NNTGACTCANN that matches annotation for JUN: Jun oncogene. Both c-jun and the JNK pathway have been previously investigated in hyperoxic lung injury and may play a protective role, as mice lacking jnk1 and jnk2 develop more rapid and severe lung injury in response to hyperoxia (45, 46). The microRNA Mir-374 was also identified by our GSEA analysis. Although several microRNAs are implicated in the acute inflammatory response, mir-374

has not been studied in ALI. However, mir-374 does have prognostic value in other lung conditions such as non-small cell lung cancer (47).

***Protein modules that are involved in hyperoxic injury and recovery***

Using a completely unsupervised approach (STEM), we identified protein modules that change coherently over injury and recovery. Profiles 43 and 7 mirror each other, exhibiting opposite expression patterns during injury and repair. Proteins in profile 7 increase during hyperoxic stress, decline to a minimum at time points when severe lung damage is present, and increase during late recovery. Because these proteins are responding to oxidative stress, it is not surprising they are annotated to oxido-reductase metabolic processes, apoptosis and metabolic processes. One could postulate that processes involved in redox status increase during oxidant stress, but they are either exhausted or the cells are unable to sustain expression resulting in cellular damage. In contrast, proteins in profile 43 display a decrease in expression with exposure to oxidative stress, but increase during the period of peak injury and early recovery from injury. These proteins participate in regulation of transcription, protein complex assembly and cell proliferation; thus, they may represent the cell's attempt to respond to structural damage. Further investigation into these expression patterns has the potential to identify novel mechanisms involved in recovery from hyperoxic stress and develop panel of proteins that measure the extent of epithelial damage or recovery.

***Study Limitations***

We focused on AT2 cells because they play a critical role in repair of an injured lung by proliferating, spreading, migrating, and differentiating to restore the alveolar-capillary barrier. However, our study does not capture the complex interactions between AT2 cells and other cells in the distal lung spaces. Our ability to identify proteins in the soluble cell fraction by MS may have been limited by the high dynamic range resulting from the use of bovine serum albumin during cell isolation. In addition, the identified proteome and measured fold change were not identical for the three different iTRAQ experiments. This could represent biological variations in the model of lung injury by oxidative stress and/or technical variability in the iTRAQ experiments. However, the Pearson correlation of 0.7 showed acceptable similarity in protein expression profile across the iTRAQ experiments. Thus, the number of proteins identified and the protein expression changes were similar across various experiments for the insoluble fraction. Although some of the bioinformatics tools we used were originally designed for microarray data analyses, other investigators have successfully used GSEA on proteomics datasets (48)

***Summary and conclusions***

We characterized AT2 protein expression changes during hyperoxic stress, a model of ALI. We identified both individual and groups of proteins that demonstrate significant change over time during progressive lung injury and subsequent recovery. Gene ontology enrichment analysis identified biological processes that are annotated to these proteins. Identification of these proteins in human clinical lung injury could provide biomarkers of lung injury and recovery, and also establish similarities between rodent hyperoxic stress and patients with respiratory failure on high levels of supplement oxygen.

***Bibliography***

1. Crapo JD, Barry BE, Foscue HA, Shelburne J. Structural and biochemical changes in rat lungs occurring during exposures to lethal and adaptive doses of oxygen. *Am Rev Respir Dis* 1980;122:123-143.

2. Adamson IY, Bowden DH, Wyatt JP. Oxygen poisoning in mice. Ultrastructural and surfactant studies during exposure and recovery. *Arch Pathol* 1970;90:463-472.

3. Barber RE, Hamilton WK. Oxygen toxicity in man. A a/barber re, lee j, hamilton wk: Oxygen toxicity in man. A prospective study in patients with irreversible brain damage. *N Engl J Med* 1970;283:1478-1484.

4. Kapanci Y, Tosco R, Eggermann J, Gould VE. Oxygen pneumonitis in man., light- and electron-microscopic morphometric studies. *Chest* 1972;62:162-169.

5. Davis WB, Rennard SI, Bitterman PB, Crystal RG. Pulmonary oxygen toxicity. Early reversible changes in human alveolar structures induced by hyperoxia. *N Engl J Med* 1983;309:878-883.

6. Ashbaugh DG, Bigelow DB, Petty TL, Levine BE. Acute respiratory distress in adults. *Lancet* 1967;2:319-323.

7. Crapo JD. Morphologic changes in pulmonary oxygen toxicity. *Annu Rev Physiol* 1986;48:721-731.

8. Thet LA, Parra SC, Shelburne JD. Sequential changes in lung morphology during the repair of acute oxygen-induced lung injury in adult rats. *Exp Lung Res* 1986;11:209-228.

9. Matute-Bello G, Frevert CW, Martin TR. Animal models of acute lung injury. *Am J Physiol Lung Cell Mol Physiol* 2008;295:L379-399.

10. Kapanci Y, Weibel ER, Kaplan HP, Robinson FR. Pathogenesis and reversibility of the pulmonary lesions of oxygen toxicity in monkeys. Ii. Ultrastructural and morphometric studies. *Lab Invest* 1969;20:101-118.

11. De Paepe ME, Mao Q, Chao Y, Powell JL, Rubin LP, Sharma S. Hyperoxia-induced apoptosis and fas/fasl expression in lung epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 2005;289:L647-659.

12. Adamson IY, Bowden DH. The type 2 cell as progenitor of alveolar epithelial regeneration. A cytodynamic study in mice after exposure to oxygen. *Lab Invest* 1974;30:35-42.

13. Holm BA, Matalon S, Finkelstein JN, Notter RH. Type ii pneumocyte changes during hyperoxic lung injury and recovery. *J Appl Physiol* 1988;65:2672-2678.

14. Chen Z, Chintagari NR, Guo Y, Bhaskaran M, Chen J, Gao L, Jin N, Weng T, Liu L. Gene expression of rat alveolar type ii cells during hyperoxia exposure and early recovery. *Free Radic Biol Med* 2007;43:628-642.

15. Bhargava M, Runyon MR, Smirnov D, Lei J, Groppoli TJ, Mariash CN, Wangensteen OD, Ingbar DH. Triiodo-l-thyronine rapidly stimulates alveolar fluid clearance in normal and hyperoxia-injured lungs. *Am J Respir Crit Care Med* 2008;178:506-512.

16. Dobbs LG. Isolation and culture of alveolar type ii cells. *Am J Physiol* 1990;258:L134-147.

17. Carter EP, Wangensteen OD, O'Grady SM, Ingbar DH. Effects of hyperoxia on type ii cell na-k-atpase function and expression. *Am J Physiol* 1997;272:L542-551.

18. Zhang Y, Wendt CH, Hertz MI, Nelsestuen. Dentification and validation of proteinase-3 and latent matrix-metalloproteinase 9 as potential biomarkers for chronic lung transplant rejection. *Clinical Proteomics* 2007;3:3-12.

19. Akkina SK, Zhang Y, Nelsestuen GL, Oetting WS, Ibrahlm HN. Temporal stability of the urinary proteome after kidney transplant: More sensitive than protein composition? *J Proteome Res* 2009;8:94-103.

20. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM, Schaeffer DA. The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics* 2007;6:1638-1655.

21. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nat Protoc* 2009;4:44-57.

22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-15550.

23. Ernst J, Bar-Joseph Z. Stem: A tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 2006;7:191.

24. McCulloch CE, Shayle, Searle R, Nehaus, John M. Generalized, linear, and mixed models. Wiley Inerscience; 2008.

25. Benjamini Y. Discovering the false discovery rate. *J R Statist Soc B* 2010;72:405-416.

26. O'Brien-Ladner AR, Nelson ME, Cowley BD, Jr., Bailey K, Wesselius LJ. Hyperoxia amplifies tnf-alpha production in lps-stimulated human alveolar macrophages. *Am J Respir Cell Mol Biol* 1995;12:275-279.

27. Shea LM, Beehler C, Schwartz M, Shenkar R, Tuder R, Abraham E. Hyperoxia activates nf-kappab and increases tnf-alpha and ifn-gamma gene expression in mouse pulmonary lymphocytes. *J Immunol* 1996;157:3902-3908.

28. Barazzone C, Horowitz S, Donati YR, Rodriguez I, Piguet PF. Oxygen toxicity in mouse lung: Pathways to cell death. *Am J Respir Cell Mol Biol* 1998;19:573-581.

29. Nogee LM, Wispe JR, Clark JC, Weaver TE, Whitsett JA. Increased expression of pulmonary surfactant proteins in oxygen-exposed rats. *Am J Respir Cell Mol Biol* 1991;4:102-107.

30. Bitterman PB, Polunovsky VA, Ingbar DH. Repair after acute lung injury. *Chest* 1994;105:118S-121S.

31. Morris PE, Glass J, Cross R, Cohen DA. Role of t-lymphocytes in the resolution of endotoxin-induced lung injury. *Inflammation* 1997;21:269-278.

32. D'Alessio FR, Tsushima K, Aggarwal NR, West EE, Willett MH, Britos MF, Pipeling MR, Brower RG, Tuder RM, McDyer JF, King LS. Cd4+cd25+foxp3+ tregs resolve experimental lung injury in mice and are present in humans with acute lung injury. *J Clin Invest* 2009;119:2898-2913.

33. Vadasz I, Weiss CH, Sznajder JI. Ubiquitination and proteolysis in acute lung injury. *Chest* 2012;141:763-771.

34. Helenius IT, Dada LA, Sznajder JI. Role of ubiquitination in na,k-atpase regulation during lung injury. *Proc Am Thorac Soc* 2010;7:65-70.

35. Staub O, Rotin D. Role of ubiquitylation in cellular membrane transport. *Physiol Rev* 2006;86:669-707.

36. Comellas AP, Dada LA, Lecuona E, Pesce LM, Chandel NS, Quesada N, Budinger GR, Strous GJ, Ciechanover A, Sznajder JI. Hypoxia-mediated degradation of na,k-atpase via mitochondrial reactive oxygen species and the ubiquitin-conjugating system. *Circ Res* 2006;98:1314-1322.

37. Lecuona E, Sun H, Vohwinkel C, Ciechanover A, Sznajder JI. Ubiquitination participates in the lysosomal degradation of na,k-atpase in steady-state conditions. *Am J Respir Cell Mol Biol* 2009;41:671-679.

38. Staub O, Dho S, Henry P, Correa J, Ishikawa T, McGlade J, Rotin D. Ww domains of nedd4 bind to the proline-rich py motifs in the epithelial na+ channel deleted in liddle's syndrome. *Embo J* 1996;15:2371-2380.

39. Rotin D, Staub O. Role of the ubiquitin system in regulating ion transport. *Pflugers Arch* 2011;461:1-21.

40. Danial NN, Korsmeyer SJ. Cell death: Critical control points. *Cell* 2004;116:205-219.

41. Barazzone C, Tacchini-Cottier F, Vesin C, Rochat AF, Piguet PF. Hyperoxia induces platelet activation and lung sequestration: An event dependent on tumor necrosis factor-alpha and cd11a. *Am J Respir Cell Mol Biol* 1996;15:107-114.

42. Barazzone C, Donati Y, Boccard J, Rochat A,Vestin, C, Kan C, Piguet, P. Cd40-cd40 ligand disruption does not prevent hyperoxia-induced injury. *America Journal of Pathology* 2002;160:67-71.

43. Chipuk JE, Moldoveanu T, Llambi F, Parsons MJ, Green DR. The bcl-2 family reunion. *Mol Cell* 2010;37:299-310.

44. Budinger GR, Mutlu GM, Urich D, Soberanes S, Buccellato LJ, Hawkins K, Chiarella SE, Radigan KA, Eisenbart J, Agrawal H, Berkelhamer S, Hekimi S, Zhang J, Perlman H, Schumacker PT, Jain M, Chandel NS. Epithelial cell death is an important contributor to oxidant-mediated acute lung injury. *Am J Respir Crit Care Med* 2011;183:1043-1054.

45. Morse D, Otterbein LE, Watkins S, Alber S, Zhou Z, Flavell RA, Davis RJ, Choi AM. Deficiency in the c-jun nh2-terminal kinase signaling pathway confers susceptibility to hyperoxic lung injury in mice. *Am J Physiol Lung Cell Mol Physiol* 2003;285:L250-257.

46. Li Z, Choo-Wing R, Sun H, Sureshbabu A, Sakurai R, Rehan VK, Bhandari V. A potential role of the jnk pathway in hyperoxia-induced cell death, myofibroblast transdifferentiation and tgf-beta1-mediated injury in the developing murine lung. *BMC Cell Biol* 2011;12:54.

47. Vosa U, Vooder T, Kolde R, Fischer K, Valk K, Tonisson N, Roosipuu R, Vilo J, Metspalu A, Annilo T. Identification of mir-374a as a prognostic marker for survival in patients with early-stage nonsmall cell lung cancer. *Genes Chromosomes Cancer* 2011;50:812-822.

48. Cha S, Imielinski MB, Rejtar T, Richardson EA, Thakur D, Sgroi DC, Karger BL. In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: Annotation by protein set enrichment analysis and gene ontology. *Mol Cell Proteomics* 2010;9:2529-2544.

***Appendix***

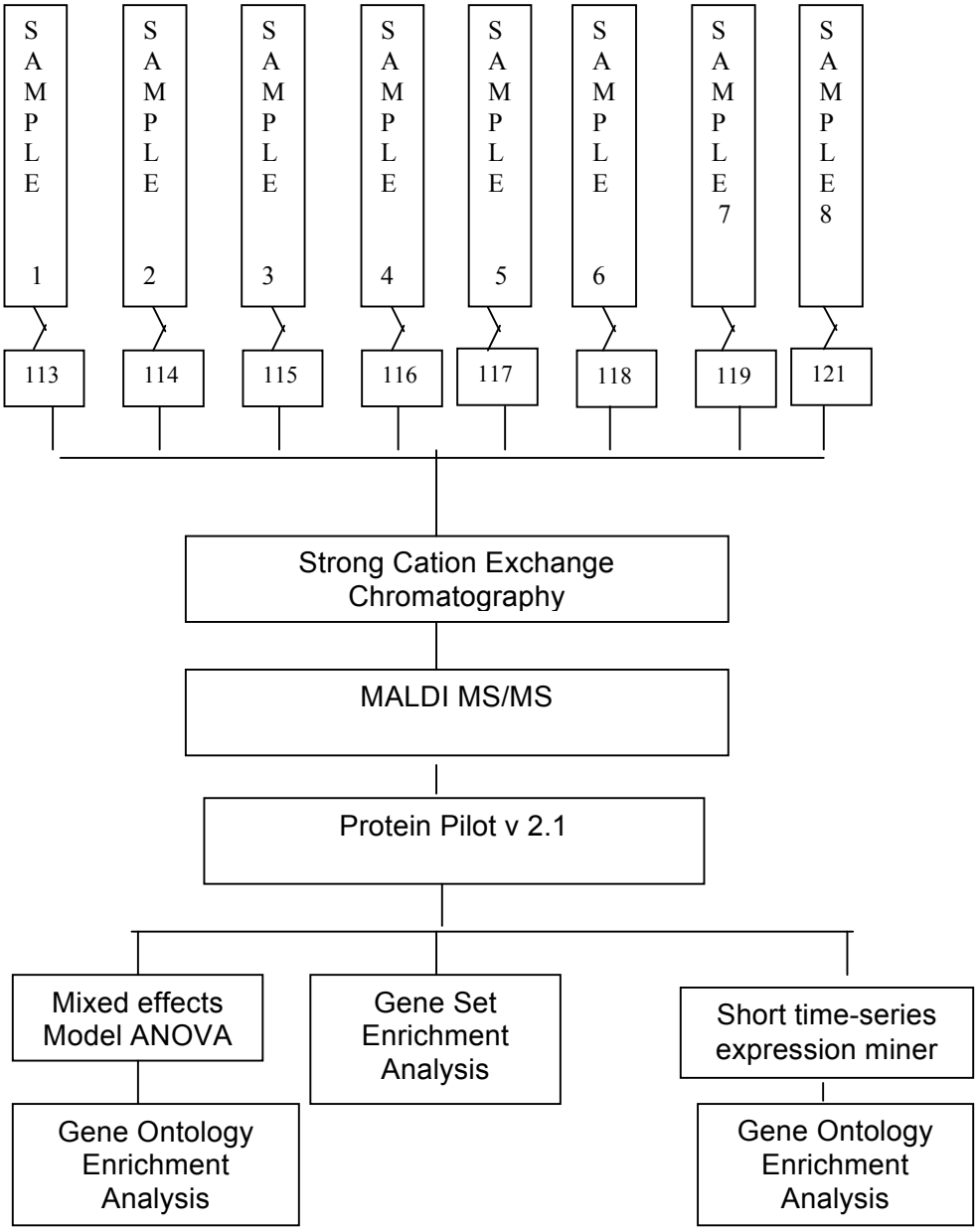


**Figure 1: Workflow for AT2 cell processing, MS data acquisition, and data analysis by bioinformatics tools**. Protein fractions were isolated and analyzed using iTRAQ® in combination with 2D LC-MALDI-MS. Trypsin digested peptides from different experimental conditions were labeled with eightplex iTRAQ® reagent and subjected to 2D LC MS. MS spectra were analyzed with Protein pilot to identify and quantify proteins. Computational analysis was performed by using DAVID for Gene Ontology enrichment analysis, Gene Set Enrichment Analysis, and Short Time-series Expression Miner (STEM) analysis followed by GO enrichment analysis.

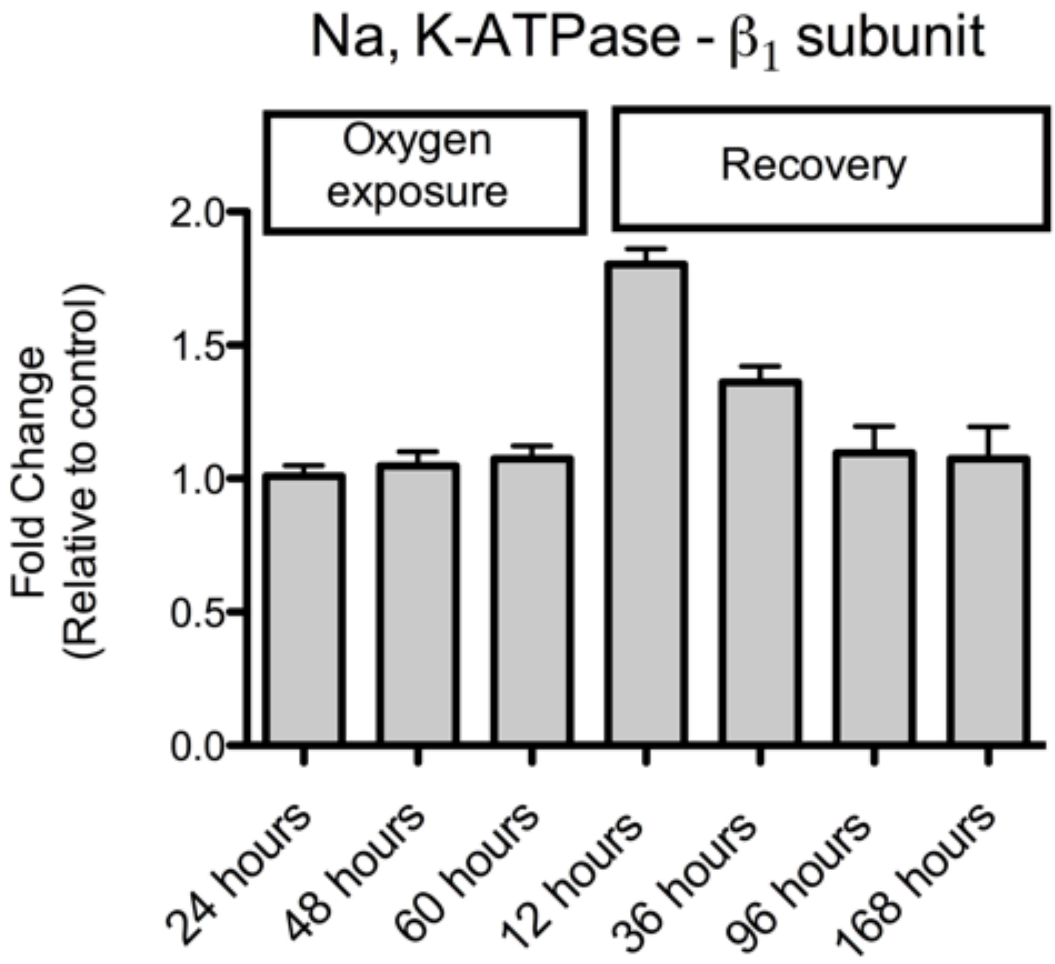


**Figure 2: Mass spectrometric changes in Na,K-ATPase protein abundance during injury and recovery.** The highest levels of $\beta_1$ Na,K-ATPase were present at 12 hours after cessation of $O_2$ exposure and remained elevated at 36 hours of recovery. $\beta_1$ Na,K-ATPase subunit was present in only two of the three iTRAQ® MS/MS experiments and was found to be one of the proteins that demonstrated a significant change in the linear mixed effect model (FDR ≤ 10%).

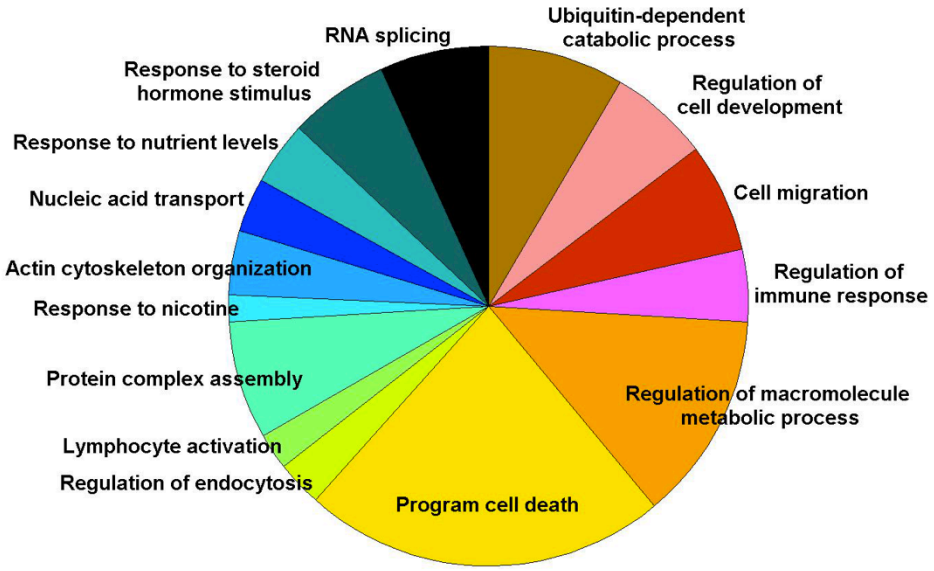


**Figure 3: Biologic processes enriched by proteins that demonstrate a significant change during hyperoxic injury-recovery.** A linear mixed effects model was used to identify significantly changed proteins during injury and recovery. GO enrichment analysis using DAVID was performed on the 183 proteins that demonstrated a significant change controlling for a FDR ≤ 10%. Functional Annotation Clustering tool was used to group related biological processes. In the functional annotation clustering tool, an enrichment score of 1.3 that corresponds to a non-log scale p-value of 0.05 was used as the cutoff for significance. The pie chart shows the number of proteins that belong to each functional cluster obtained from DAVID functional annotation clustering tool.

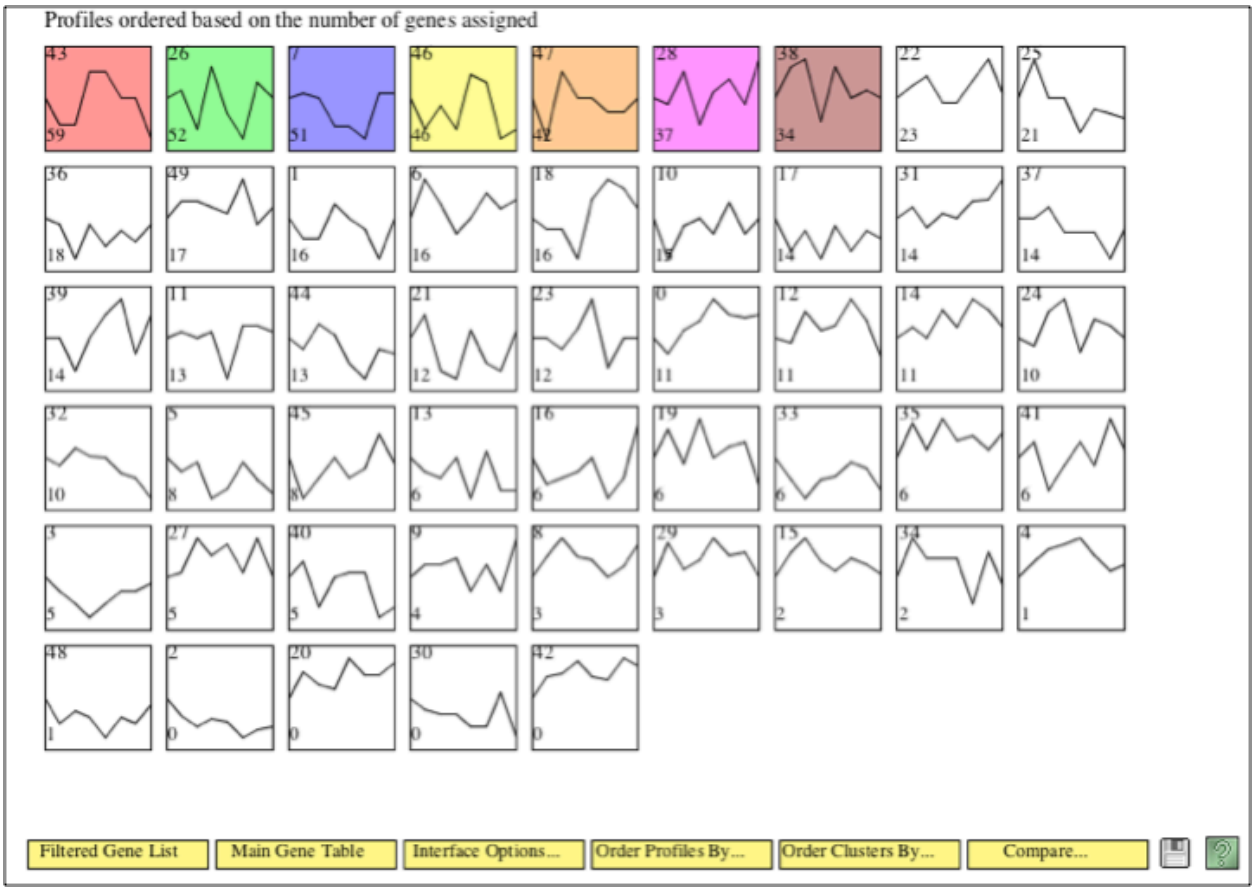


**Figure 4: Identification of significant protein groups that demonstrate coherent changes during hyperoxic injury and recovery by STEM algorithm.** Data from the three separate iTRAQ® experiments was analyzed using 'repeat data' functionality in STEM algorithm. 50 distinct randomly generated model profile depicted in each small square were tested. For each profile, represented in the profile box, x-axis represents the 7 time points of our experimental sampling during oxygen exposure and recovery. The y-axis represents expression change (fold change relative to control). The number on the top left-hand corner of a profile box is the profile ID number and the bottom left-hand corner is the number of proteins assigned to that profile. Colored profile boxes had a statistically significant number of genes assigned. Seven of the 50 protein profiles (in color) had a statistically significant number of proteins assigned to them by determining which model profiles have significantly more genes assigned under the true ordering of time points compared to the average number assigned to the model profile in the permutation test.

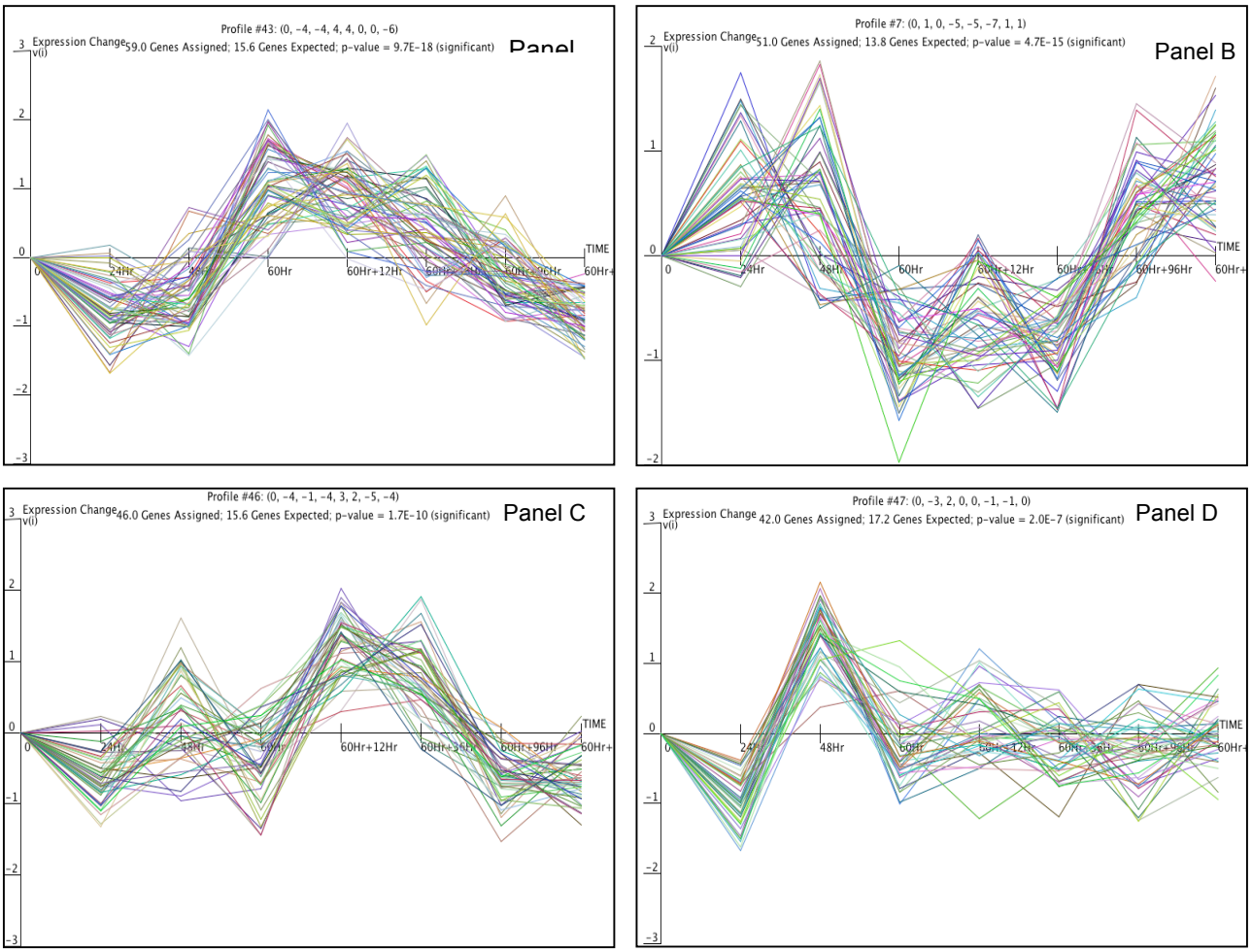


**Figure 5: Protein groups that demonstrated interesting temporal changes over hyperoxic injury and recovery using Short Time-series Expression Miner analysis.** Four groups of proteins (clusters) that had interesting temporal changes in protein abundance (relative to controls) during injury and recoveries were selected for further investigations are shown. Each of the profiles has a graph of expression pattern (y-axis, expression change in form of fold change) across the time-series (time, x-axis), a count of the number of proteins assigned, count of the number of proteins expected based on the permutation test and the profiles p-value. Gene symbols of the proteins in each of the profile are available in the supplemental table S5 and the GO enrichment analysis is shown in supplemental table S6.

| **Table 1: Proteins names and IPI number of the top ten proteins (significant) that changed during injury and repair** | | |
|---|---|---|
| **Gene Symbol** | **IPI Number** | **False Discovery Rate** |
| Ptprc Isoform 5 of Leukocyte common antigen | IPI00851132.1 | $4.16 \times 10^{-6}$ |
| LOC688509 similar to Alpha-enolase | IPI00767147.1 | $1.57 \times 10^{-4}$ |
| Actg1 Actin, cytoplasmic 2 | IPI00896224.1 | $1.70 \times 10^{-4}$ |
| Msn 68 kDa protein | IPI00778167.1 | $1.71 \times 10^{-4}$ |
| Dhx9 DEAH (Asp-Glu-Ala-His) box polypeptide 9 | IPI00366249.4 | $3.87 \times 10^{-4}$ |
| LOC689908 similar to heat shock protein 8 | IPI00566672.2 | $7.70 \times 10^{-4}$ |
| Tubb5 Isoform 1 of Tubulin beta-5 chain | IPI00197579.2 | $8.38 \times 10^{-4}$ |
| Sumo2 Small ubiquitin-related modifier 2 | IPI00192749.1 | $8.38 \times 10^{-4}$ |
| Aldh2 Protein | IIPI00778252.1 | $8.38 \times 10^{-4}$ |
| Krt10 58 kDa protein | IPI00551673.3 | $1.67 \times 10^{-3}$ |

Table 2: Gene Sets Differentially Expressed During Recovery From Hyperoxic Stress

| **Gene sets enriched during recovery from experiment lung injury** | **Systematic name** | **Biological Process** | **Size (Number of genes)** | ***FDR- q value** |
|---|---|---|---|---|
| Multicellular Organismal Development | M12777 | Progression from initial condition to later condition | 15 | 0.042 |
| Systems Development | M11617 | Progression of organ system over time | 14 | 0.044 |
| Regulation of Biological Quality | M10930 | Modulation of frequency, rate or extent of biological quality | 11 | 0.225 |
| Organ Development | M5412 | Development of tissue or tissues that work together | 10 | 0.242 |
| Chemical Homeostasis | M17520 | Maintenance of internal equilibrium | 6 | 0.21 |
| Multi organism Process | M11259 | Process by which organism has effect on other organism | 3 | 0.19 |

| Table 3: GO biological process enriched by proteins modules from STEM profiles | | | |
|---|---|---|---|
| (Figure 5, Panel A) | (Figure 5, Panel B) | (Figure 5, Panel C) | Figure 5, Panel D |
| Regulation of transcription | Oxido-reduction coenzyme metabolic process | ATP biosynthetic process | Actin filament based processes |
| Positive regulation of cell proliferation | Protein DNA complex assembly | Cytoskeletal organization | |
| Response to wounding | Regulation of apoptosis | Cellular homeostasis-response to hypoxia | |
| Protein complex biogenesis | | Cellular ion homeostasis | |
| Response to organic substance | | Response to extracellular stimulus | |
| | | Cell motility/ cell adhesions | |

**Supplementary Tables**

Table S1**:** IPI accession number and Official Gene Symbol of all the proteins identified in the individual iTRAQ experiments.

Table S2: List of 183 proteins identified to change during the course of injury/recovery with their associated FDR values

Table S3: Functional Annotation Clustering results obtained on the 183 proteins that demonstrated changes during injury and recovery.

Table S4: Genes Sets that are differentially expressed recovery in room air after hyperoxic stress

Table S5: Proteins in the four individual STEM profiles in Figure 5.

Table S6: Functional Annotation Clustering of the proteins in the four protein clusters identified by STEM algorithm.