medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Reviews

Identification of Three Endotypes in Pediatric Acute Respiratory Distress Syndrome by Nasal Transcriptomic Profiling

James "Garrett Williams[1*], Rashika Joshi[1†], Rhonda Jones[1‡], Toni Yunger[1§], Erin Stoneman[1¶] & Brian Varisco[1‖ **]

[1]*Critical Care Medicine, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC 7006, Cincinnati, 45229, OH, USA. Tel.: 513-803-2485, Fax: 513-803-3311*

Keywords: ARDS

1 ABSTRACT

Acute respiratory distress syndrome (ARDS) and pediatric ARDS (PARDS) can be triggered by multiple pulmonary and non-pulmonary insults and are the source of substantial morbidity and mortality. The nasal epithelium is similar to that of the lower conducting airways and nasal transcriptomic profiling identifies disease and endotypes in lung cancer, COPD, and asthma. We conducted a prospective trial of testing whether this technique could identify PARDS endotypes in 26

*E-mail: james.williams2@cchmc.org
†E-mail: rashika.joshi@cchmc.org
‡E-mail: rhonda.jones@cchmc.org
§E-mail: toni.yunger@cchmc.org
¶E-mail: erin.stoneman@cchmc.org
‖Corresponding author.
**E-mail: brian.varisco@cchmc.org

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

control and 25 PARDS subjects <18 admitted to the PICU extracting RNA from inferior turbinate brushings on days 1, 3, 7, and 14. Standard RNA processing and mRNA-seq yielded ~25% usable specimens and a modified, low-input RNA protocol yielded 95% usable specimens. Sixty-four low-input specimens from 10 control and 15 PARDS subjects were used for initial analysis. Control and some PARDS subjects clustered together in both by-day and combined analysis into Groups A while some day 1, 3, and 7 specimens from the same subjects clustered into Groups B and C with specimens from these subjects moving to Group A with PARDS resolution. In multivariate analysis, the only clinical variables predictive of Group B or C status was severity of lung injury or viral PARDS trigger. Compared to Group A, Group B had upregulation of innate immune processes and group C had upregulation of cilia-associated processes. Analysis of the 15 specimens processed and analyzed by standard techniques identified the same processes as differentiating Groups A, B, and C. Mortality appeared to be higher in group B (25%) and C subjects (28.6%) compared to A subject (5%, p=0.1). Our findings demonstrate three distinct PARDS endotypes which may be useful for prognostic and therapeutic enrichment. ClinicalTrials.gov Identifier NCT03539783

2 List of abbreviations

ARDS Acute Respiratory Distress Syndrome

COPD Chronic Obstructive Pulmonary Disease

ECMO Extracorporeal Membranous Oxygenation

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FiO2 Fraction of Inspired Oxygen

MAP Mean Airway Pressure

OI Oxygenation Index (FiO2 X MAP / PaO2)

OSI Oxygen Saturation Index (FiO2 X MAP / SaO2)

PaO2 Arterial Partial Pressure of Oxygen

PICU Pediatric intensive care unity

PARDS Pediatric Acute Respiratory Distress Syndrome

RNA Ribonucleic Acid

SaO2 Arterial Oxyhemoglobin Saturation

3 Background

In the Pediatric intensive care unity (PICU), pediatric acute respiratory distress syndrome (PARDS) is a leading source of morbidity and mortality [1]. Despite decades of research and may large, multi-center, randomized clinical trials in ARDS, the only consensus therapies are supportive: the use of low tidal volume ventilation and employing a restrictive fluid strategy [2 3]. A criticism of many studies in ARDS has been their failure to account for etiologic and physiologic differences [4]. ARDS

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

etiology can be broadly categorized as direct (i.e. pneumonia, aspiration, etc.) or indirect (i.e. sepsis, hemorrhagic shock, etc.) In adult ARDS, both direct and indirect ARDS confer similar mortality despite lower illness severity scores in direct ARDS, suggesting different pathophysiological processes [5]; differences that are dramatically exemplified in a recent publication showing near-complete replacement of epithelial cells with macrophages upon lung biopsy in two young adults placed on extracorporeal membranous oxygenation (ECMO) for presumed community acquired pneumonia [6]. However, categorizing the inciting injury as direct or indirect has proven inadequate to guide therapy [7 8]. Nonetheless, several serum biomarkers indicate differences in the pathophysiology of direct vs. indirect ARDS [9], suggesting the possibility of using biomarkers PARDS endotypes (defined as a non-apparent but clinically meaningful subclassification). Several serum biomarkers are under investigation. Elevated serum levels of advanced glycosylation end-product specific receptor (AGER, also RAGE) predict ARDS progression [10]. The endothelial cell protein angiopoietin-2 (Ang2) was positively associated with mortality in PARDS [11], and adult surgery or trauma-related ARDS [12]. In latent group analysis of the FACTT trial [3], subjects with higher levels of inflammation (as evidenced by higher interleukin-8 and tumor necrosis factor receptor-1 and lower serum bicarbonate levels) were less likely to survive with a fluid liberal strategy, while subjects with lower levels of inflammation were more likely to survive with a fluid-liberal strategy [13]. Thus, serum biomarkers demonstrate some ability to differentiate two ARDS subtypes, may help to guide some therapies, and may help predict course.

However, there is perhaps a limit to the extent to which peripheral blood-based assays accurately reflect lung pathology. While peripheral blood gene expression profiling in pediatric sepsis

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

had identified important endotypes that correlate with outcome [14], there is only ~40% concordance of leukocyte gene expression between lung and peripheral leukocytes with ~20% of genes being expressed discordantly [15]. While RAGE is a lung-specific protein, Ang2 and inflammatory cytokines are present in multiple tissues and may lack specificity. We propose that directly assaying the gene expression of respiratory epithelial cells can better characterize lung pathobiology and longitudinally assessing this expression to identify key pathways and processes in PARDS recovery. Such profiling has been well-established in other respiratory diseases. Bronchial and nasal gene expression profiling was highly diagnostic for the presence of lung cancer in smokers [16], and for corticosteroid sensitivity in asthma [17]. Nasal and bronchial gene expression profiling can differentiate between COPD and non-COPD in smokers [18]. To accomplish this in the PICU environment, collection, storage, transportation, and processing techniques must minimize the degradation of sample RNA and be compatible with available resources (specimen storage temperatures, absence of noxious odors, easy to handle and manage, etc). Here, we report the development of a technique to collect quality RNA specimens from nasal respiratory epithelial cells and present data that nasal epithelial cell transcriptomics can be used to identify three distinct PARDS endotypes.

4 METHODS AND DESIGN

Human Subjects Research Studies were approved by the Cincinnati Children's Hospital Institutional Review Board (IRB 2015-8514 & 2017-1345) and registered with ClinicalTrial.gov (NCT#03539783).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Study Procedure Patients admitted to the PICU <18 years of age who were intubated and meeting consensus PARDS criteria [1] or who were admitted to the PICU without apparent lung disease and with expected duration of hospital admission 7 days or more were eligible for enrollment. Exclusion criteria were orders for limitation of resuscitation, known nasal pathology, a high risk of nasal bleeding as determined by the clinical team, or baseline oxygen requirement of 2 liters per minute or more. Enrollment was permitted at any time during PARDS course so long as the first day meeting PARDS criteria was known. Nasal brushings and serum samples were collected on study days 1, 3, 7, and 14. Clinical and biometric parameters at the times of enrollment and specimen collection were recorded in a RedCap database.

Nasal Brushing Using a soft cytology brush (CytoSoft, CYB-1, Medical Packaging Corporation, Camarrilo, CA, for patients >20 kg) or 3.0 mm Nasal Cytology Brush (Hobbs Medical, Staford Springs, CT, for patients <= 20 kg), nasal brushing of an inferior turbinate was performed on all subjects. The end of the brush was cut using wire cutters and preserved in 1 mL of RNAProtect buffer (Qiagen).

Specimen Preservation and RNA Extraction Brushes in RNAProtect were stored at -20C for up to a week and then at -80C. For RNA extraction, 10μL of $\beta - mercaptoethanol$ was added during thawing on ice to reduce mucus disulfide bonds and the specimen was passed through a QIAShredder (Qiagen) and RNA collected using the AllPrep DNA/RNA \ Mini kit* (Qiagen) eluting RNA with 50 μL of nuclease-free water per manufacturer protocol. RNA quality and quantity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Standard Library Preparation and Sequencing Specimen ribosomal RNA was depleted us- ing RiboZero (Illumina) and libraries created using NexteraXT (Illumina), and sequenced on a HiSeq2500 using paired end sequencing of 150 base pairs at a sequencing depth of 10 million reads per sample.

Unsuccessful RNA Amplification Specimen ribosomal RNA was depleted using NEBNext rRNA Depletion kit (New England Biolabs) and amplified using SeqPlex (Sigma) with 20 cycles of amplification. Barcoding was performed first with PlexWell96 (SeqWell) and then with NexteraXT. Sequencing was not undertaken due to low cDNA concentration.

Low-Input Library Preparation Specimen RNA was amplified and barcoded using the NEBNext SingleCell/Low Input RNA Library Prep kit (New England Biolabs) per manufacturer instructions with 20 cycles of amplification using NEBNext multiplex oligos for barcoding. Specimen DNA concentration was normalized, and sequencing was performed using a Novaseq sequencer and a single S4 flow well with paired end sequencing of 150 base pairs each yielding ~10 million reads per sample.

Group Assignment and Bioinformatic Analysis Sequence files were aligned to GRCh38 and specimen gene counts assessed using STAR [19] with analysis of differential expression using DESeq2 [20]. Briefly, a count file was created from STAR output and uploaded into DESeq2. The datasets were first analyzed by day using group (Control or PARDS) as the only variable and determining clustering by Euclidean distance and principal component analysis. As clusters consistently contained specimens from the same subject, re-analysis of specimens from all days was

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

undertaken with the same analysis. A similar analysis was undertaken using the standard-technique RNA-Seq dataset. From these results, groupings were determined using principal component analysis with specimen assignment added to metadata. Data was visualized using pheatmap [21] and ggplot2 in R 3.6.3 [22]. The differential gene expression of each PARDS cluster against the control cluster was determined in a new DESeq2 analysis with pathway and upstream regulator analysis of two-fold increased or decreased genes performed using ToppGene [23] and Ingenuity Pathway Analysis [24] with visualization of results using pheatmap.

Statistical Analysis For bioinformatic analyses a Bonferroni-corrected q-value of less than 0.1 was considered significant. The R statistical package was used for all analyses using ggplot2 and finalfit [25] packages. For all non-bioinformatic analyses p-values of <0.05 were considered significant with use of Fisher Exact test for nominal and ordinal data and Wilcoxon-Rank Sum test for continuous data.

5 RESULTS

Recruitment and Specimen Collection From January 1, 2018 to November 30, 2019 we enrolled 21 control and 19 PARDS subjects. From these subjects, we collected 11 nasal brushings. PARDS subjects differed from control in illness severity (PELOD2), severity of lung injury, length of PICU and hospital stay, and exposure to corticosteroids (Table 1).

Reasons for Failure of Non-low Input RNA Techniques We arrived at our current protocol after several failures. Library creation was successful in only 13 of 45 (29%) of RNA specimens

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

using RiboZero ribosomal RNA depletion, Illumina NexteraXT kit and HiSeq 2500 sequencing. To increase yield we attempted we performed ribosomal RNA depletion using NEBNext rRNA Depletion Kit (New England Biolabs) and amplified specimen RNA using the SeqPlex RNA kit (Sigma) with of all but one of 48 specimens having amplification. Thirty-five (73%) specimens had amplicons in the desirable 150-400 base pair range, and twelve (25%) had a predominance of ~150 base pair amplicons. Barcoding of these specimens with PlexWell96 kit (SeqWell) was unsuccessful, and barcoding using Nextera DNA Flex kit was successful in the 35 desirable-range specimens but the pooled DNA concentration too low for sequencing. We then used specimens with at least 20 ng of RNA remaining and RNA from specimens acquired in the interim with the New England Biosystems NEBNext Single Cell/Low Input RNA Library Prep Kit with NEBNext Multiplex Oligos and were able to create barcoded libraries in 95% (64/67) of specimens. The sequences from these specimens was used for downstream analysis (Figure 1)

Assessing Nasal Specimen Similarity 64 nasal brushing specimens from 15 PARDS and 10 control subjects collected on study days 1, 3, 7, and 14, were analyzed by DESeq2. On days 1 and 3, four clusters were readily apparent by principal component analysis with two of these clusters (A1 and A2) containing control subjects (Figure 2 A&B). On study days 7 and 14, fewer samples were available for analysis, but clusters A1, A2, and C were still apparent (Figure 2 C&D). Analysis of all specimens together yielded the same specimens in clusters B and C, but specimens from A1 and A2 were no longer distinct and were combined for downstream analysis (cluster A, Figure 2 E). By Euclidean distance of the combined specimens, cluster B was clearly distinct from A and C, and the specimens in cluster C were grouped together, but cluster C was located within the larger

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

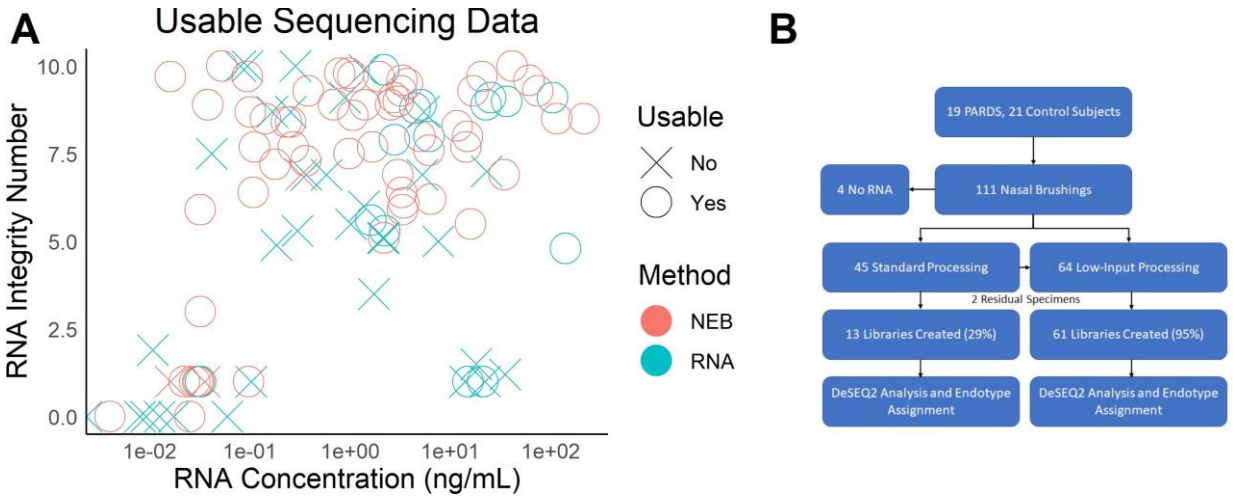

Figure 1: Study Specimens. (A) Comparison of standard (RNA) and low-input (NEB) techniques of nasal brushing cDNA library creation. NEB permitted creation of libraries from specimens of lower RNA concentration and lower quality (RNA Integrity Number). (B) Study specimen flow diagram.

cluster A (Figure 2 F). PARDS severity is classified as mild, moderate, or severe based on impairment in oxygenation [1]. While cluster A contained several specimens obtained during moderate or severe PARDS, clusters B and C were exclusively composed of such specimens (Figure 2 G). By linking the temporally obtained specimens for each subject, a clear trajectory from cluster B to cluster A and from cluster C to cluster A (with one exception) was noted (Figure 2 H). Lastly, we analyzed the 13 specimens collected and processed without amplification and with standard cDNA library prep using the same analysis method. Three groups were again identified, although not all control specimens were clustered together (Figure 3). Notably however, among the two control subjects in cluster C one developed PARDS, and one nearly met mild PARDS criteria (oxygen saturation index 4.8). This longitudinal data with validation in a second dataset indicate that nasal

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

transcriptomic gene expression in PARDS patients can be classified into one of three gene expression profiles. One similar to that of control subjects (Group A), and two others that are largely distinct from those of control subjects (Groups B&C).

Characterization of Three PARDS Endotypes For this study, we define "endotype" as a biologically distinct disease subgroup that is clinically indistinguishable from other subgroups of that disease. To first determine whether or not there were clinical characteristics of the subjects or clinical measures present at the time of specimen collection that were predictive of cluster assignment, we performed a descriptive analysis of subjects (assigning subjects as "B" or "C" if any one specimen was so classified and "A" if all specimens were cluster A) and of specimens. We then performed univariate analysis of group A vs combined Group B&C, and multivariate logistic regression for any variables with $p<0.1$. For categorical variables Fisher's exact test was used for statistical comparisons and for continuous and ordinal variables Wilcoxon-Rank Sum test was used. In comparing Group A, B, and C subjects, only disease severity (PELOD2) was statistically significant (Supplemental Table 1), and for individual specimens, PARDS classification (None, Mild, Moderate, or Severe), and the presence of direct lung injury, a viral or combined viral/bacterial cause of ARDS were significantly different between groups A, B, and C (Supplemental Table 2). In subject univariate analysis, group assignment (Control or PARDS) neared statistical significance ($p=0.06$) and both disease severity and PARDS severity were significantly higher in group B&C compared to group A (Supplemntal Table 3). In specimen univariate analysis, disease severity and direct lung injury neared statistical significance (both $p=0.05$), and PARDS severity, viral infection, and combined viral/bacterial infection were significant (Supplemental Table 4). In multivariate analy-

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

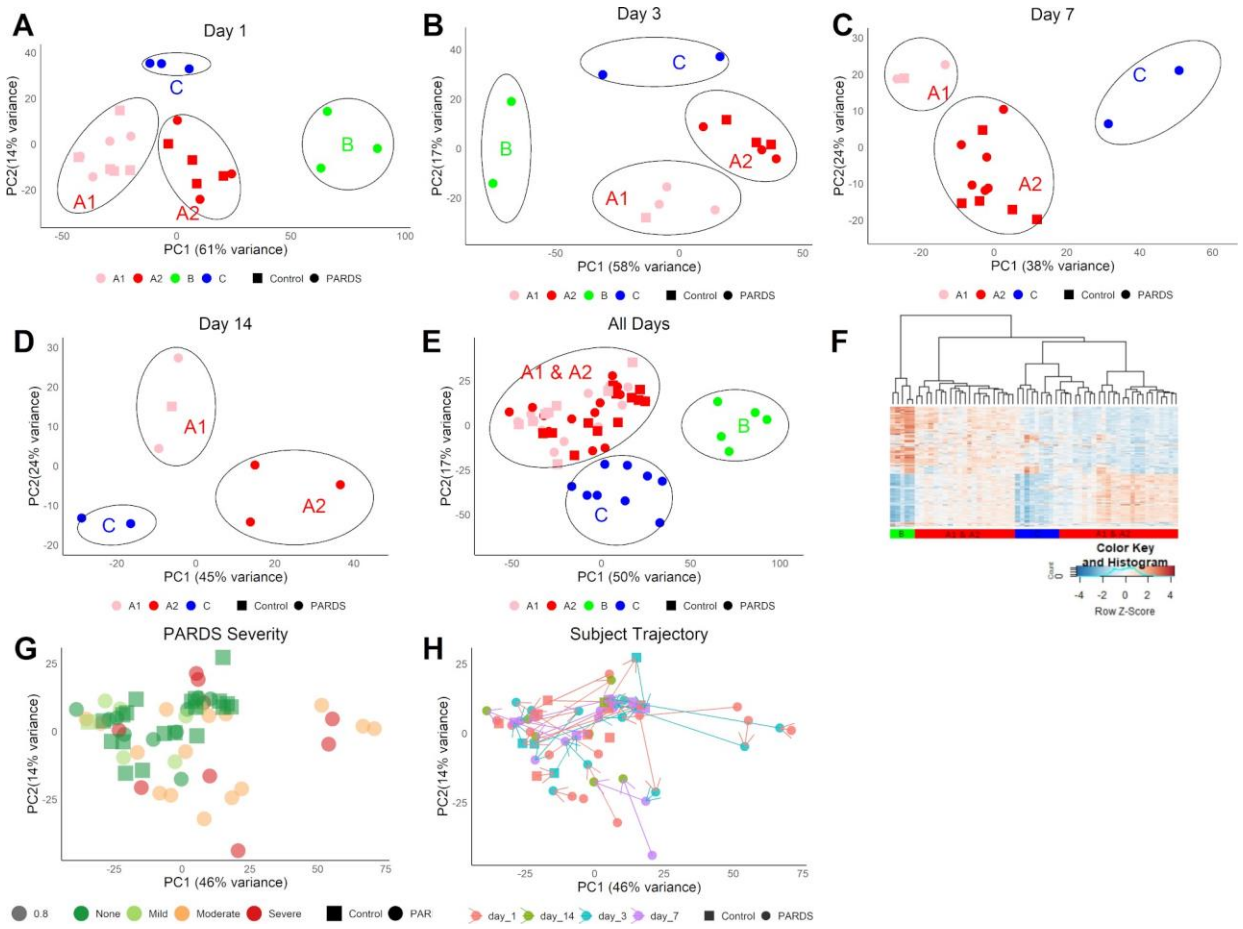

Figure 2: Nasal Specimen Transcriptomic Cluster Analysis. (A) Principal component analysis (PCA) of mRNA-Seq data from day 1 and (B) day 3 nasal specimens showed four clusters, A1, A2, B, and C. Specimens from control subjects were restricted to clusters A1 and A2. (C) Day 7 and (D) day 14 specimens had only A1, A2, and C clusters. (E) In the combined dataset, clusters A1 and A2 were no longer distinct but clusters B and C remained so. (F) A heatmap of Euclidean distance clearly differentiated the five PARDS specimens comprising cluster B. Specimens from cluster C were most similar to each other although not entirely separated from cluster A. (G) PARDS severity was greatest in specimens from clusters B and C although moderate and severe PARDS were present in cluster A. (H) Following specimens from each subject over time, subjects
with specimens that were initially in clusters B and C moved into cluster A over time.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

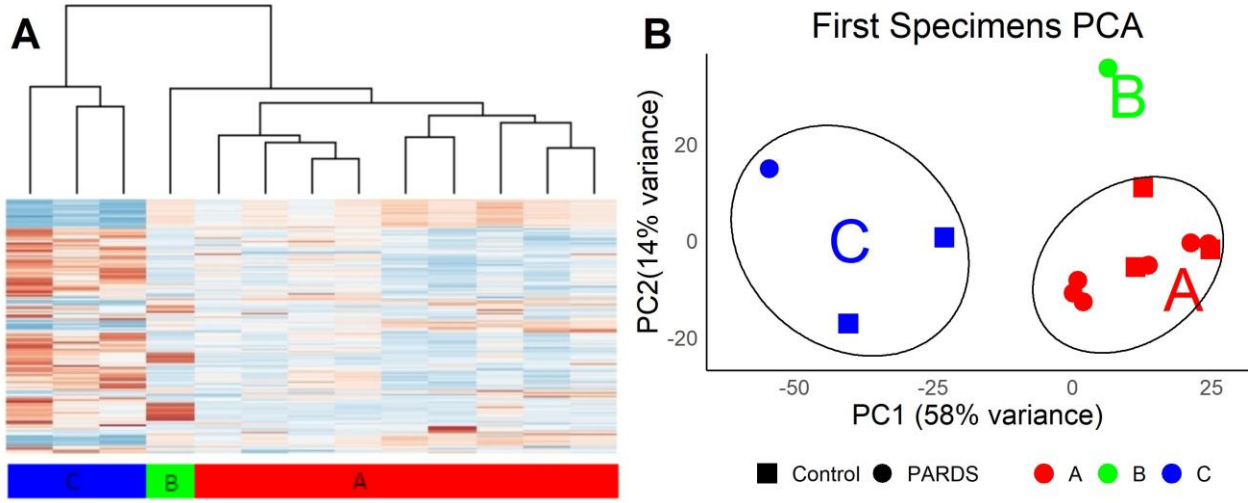

Figure 3: Similar Clustering in Distinct Group of Nasal Specimens. (A) Heatmap of the RNA processed by standard techniques for mRNA-seq (only 25% successful) showed clustering into three groups. (B) Similar to the low-input technique, specimens processed by this technique clustered into three groups.

sis of subjects, no variable was statistically significant (Supplemental Table 5), but in multivariate analysis of specimens, PARDS severity (p=0.005) and viral PARDS trigger (p=0.04) were significant (Supplemental Table 6). Although these findings (Figure 4) should be interpreted with caution due to low numbers, these data demonstrate that (1) none of the collected demographic or disease-specific data apart from a viral PARDS trigger significantly influences group assignment and (2) changes in the nasal transcriptome mirror changes in disease severity in PARDS subjects. Notably, an equal percentage (67%) of specimens classified as group B or C had a viral ARDS trigger.

Since specimens from groups B and C were collected from subjects at times of greater lung injury severity and group A contained the specimens from these same subjects at times of lesser

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

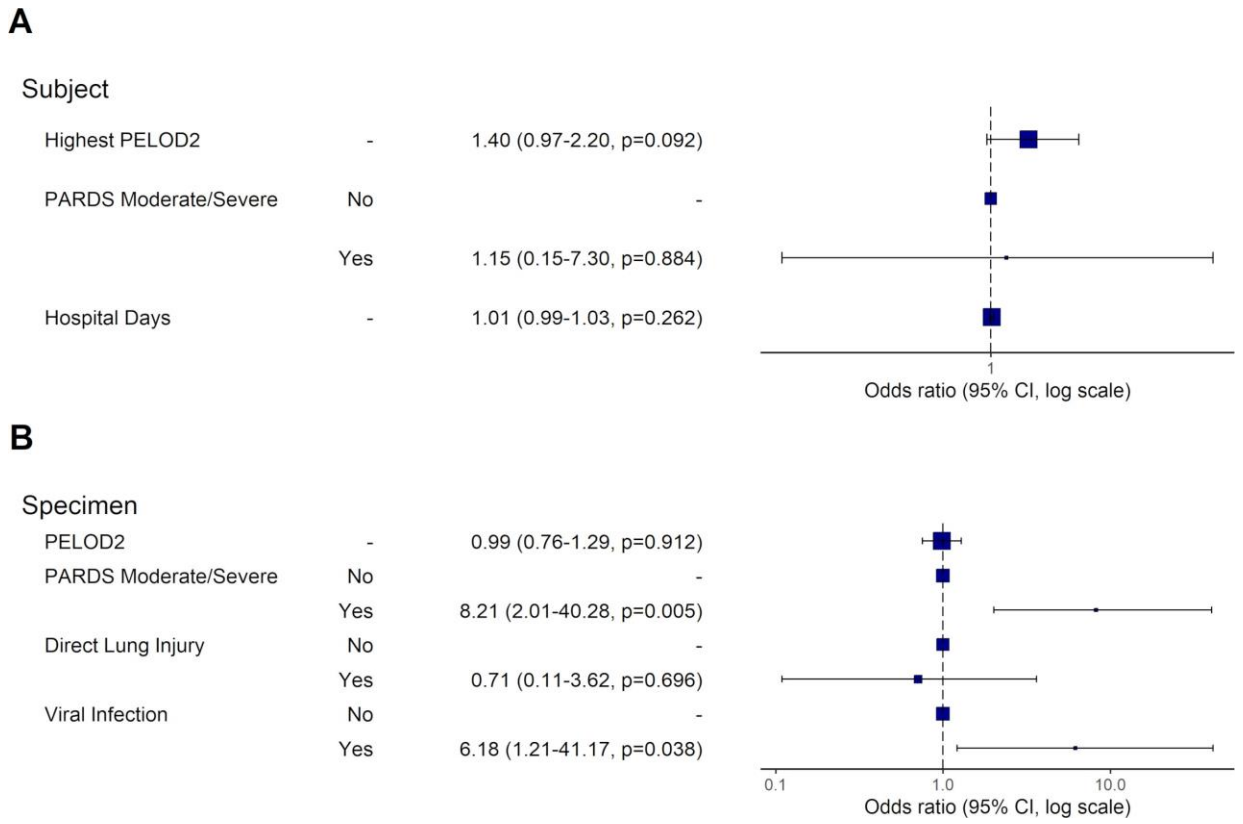

Figure 4: Variables Associated with Group B or C in Multivariate Logistic Regression. (A) Vari- ables with p<0.1 in univariate analysis were subjected to multivariate analysis. Subjects with any specimens classified as B or C tended to have greater organ dysfunction, but this finding was not statistically significant. (B) Specimens from subjects with moderate or severe PARDs or with viral infection as a PARDS trigger were more likely to be classified as group B or C.

lung injury severity in addition to control subjects, we analyzed sets of deferentially expressed genes (DEGs) of group B compared to group A and group C compared to group A without respect to collection day. After filtering for expression level and fold-change of 2 or greater, the Group B analysis contained 1192 up and 873 downregulated DEGs. Group C contained 533 up and 86 downregulated DEGs. Toppgene analysis of these datasets revealed that compared to Group A,

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Group B had decreased and Group C had increased representation of genes related to ciliary function microtubule motor activity. Group B had increases in many genes related to innate immunity (Figure 5 A-B). The pathways and gene family overepresented in the two datasets contained terms related to neutrophil degranulation, cytokines, and tumor necrosis factor for Group B and cilium assembly and dyneins for Group C (Figure 5 C-D). There was a notable absence of adaptive immune terms in this ToppGene analysis. Ingenuity Pathway Analysis was undertaken to identify potential regulators. Interferon-γ and tumor necrosis factor-related signaling were notable in Group B (Figure 5 E, G), and in analyzing network depth, suppression of FOXP1 and the proteosomal protein UBE3C were identified as potential master regulators of the Group B inflammatory cascade (Figure 5 F). Considering the differences in nasal epithelial cell gene expression in groups A, B, and C in conjunction with our findings that the only clinical variables that differentiated specimens from groups B and C from A were severity of lung injury and a viral cause of ARDS, and that viruses were the most common trigger of ARDS in both groups B and C, these data demonstrate that nasal epithelial transcriptomics can identify three distinct endotypes in PARDS: Endotypes A, B, and C.

Possible Sub-endotypes in B but not C We were interested in the lack of an adaptive immune signature in Endotype B. In analyzing Endotype B and C subjects separately, we found that subjects in Endotype B clustered together tightly with differences driven by adaptive immune genes (Figure 6 A), while in Endotype C there was no such clustering apparent (Figure 6 B). While preliminary, these data suggest that within the inflammatory Endotype B, there may exists sub-endotypes that may represent distinct inflammatory processes in these subjects. It seems unlikely that the sub-endotypes represent resolution of inflammation because the subsequent specimen for

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

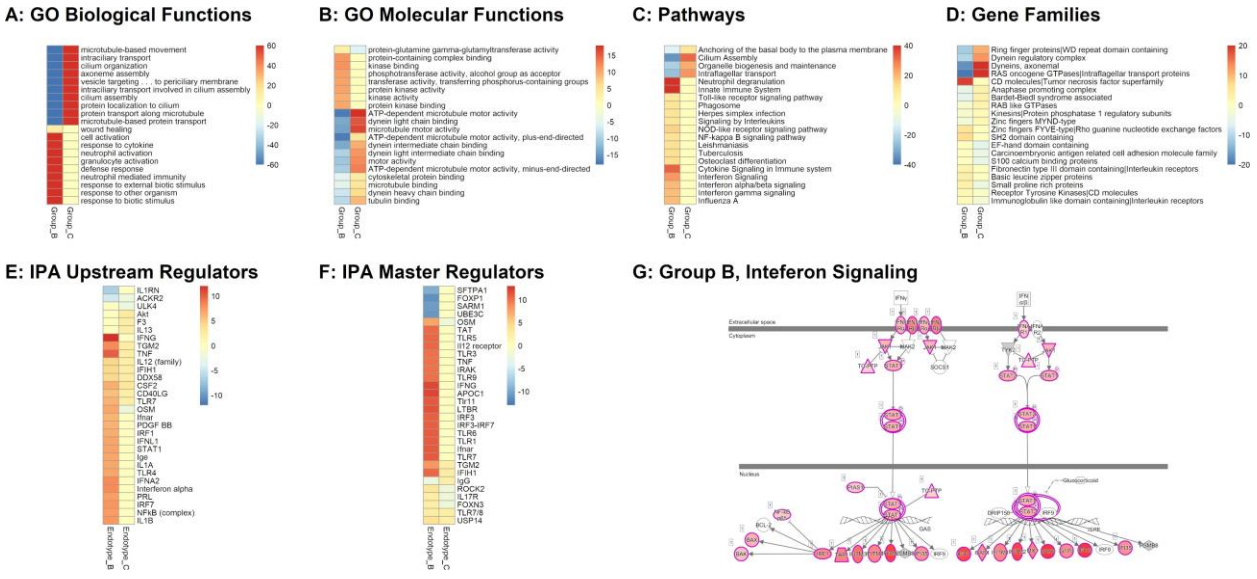

Figure 5: Characterization of ARDS Endotypes. (A) ToppGene and IPA were used to analyze DEGs in groups B and C compared to group A, which contained all of the control subjects in the dataset analyzed (the original dataset could not be normalized to the larger, second dataset because of differences in RNA processing and amplification). Innate immune processes were increased in group B and cilia-related processes increased in group C and decreased in group C. (B) GO molecular function analysis demonstrated a similar up- and downregulation of dynein and microtubule motor processes in groups C and B respectively. (C) Pathway and (D) Gene Family analyses were consistent with an increase in innate immune processes in group B and cilia-related ones in group C. All color scales for panels A-D represent -log10(Bonferroni-corrected p-value) for upregulated and log10(corrected p-value) for downregulated terms. (E) IPA analysis of upstream regulators showed increased Interferon-γ and tumor necrosis factor in group B and increased Akt and Interleukin-13 signaling in group C. (F) Network depth analysis of master regulators identified a potential role for FOXP1 suppression in Group B in addition to upregulation of many members of inflammatory signaling cascades. Oncostatin-M (OSM) was the only term with reciprocal scoring between the groups and is known to regulate cell cytokine production. Color scales for panels E-F represent activation or inhibition z-score. (G) Upregulation of many components of the Interferon-γ *signaling* cascade is shown. Red represents degree of upregulation. Green is absent but represents downregulation.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

each subject in Endotype B was Endotype A.

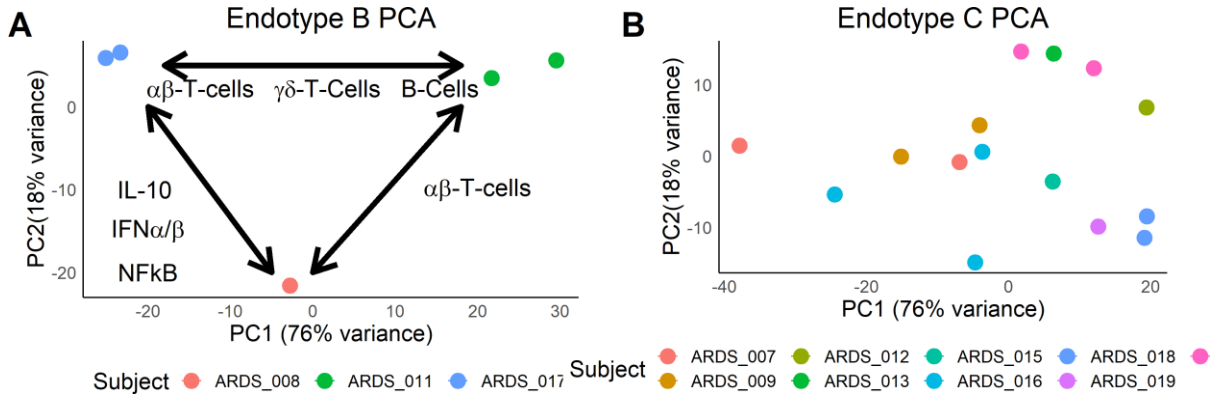

Figure 6: Sub-Endotypes in PARDS. (A) When analyzed separately, Endotype B specimens clus- tered by subject with pathways and cell-specific processes related to adaptive immunity differenti- ating them. (B) Subject separation was not noted for specimens in Endotype C.

6 Discussion

Our findings in conjunction with reports on the ability of serum biomarkers in pediatric [11] and adult [9, 13] ARDS to identify clinically meaningful endotypes provide hope that future trials in ARDS can incorporate diagnostic and prognostic enrichment in subject recruitment and enrollment thus increasing the likelihood of discovering meaningful therapies. The lack of such enrichment could explain the frequency with which small pilot clinical trials show positive results only to fail in larger trials. Given sufficient infrastructure and support, nasal transcriptomics based endotyping could be performed in 48-72 hours to provide this diagnostic and prognostic enrichment.

Our findings of an inflammatory endotype (Endotype B) is in good agreement with the find-

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ings of Calfee, *et al.* [13] showing that hyper-inflammatory adult ARDS responds less well to high levels of PEEP and liberal fluid administration. However, our findings also suggest that the nasal epithelium harbors cell-specific information of adaptive immune processes that could be exploited for targeted immunotherapies. Additional data is required. Endotype C was less distinct from Endotype A which contained control and resolved PARDS specimens, but the upregulation of ciliary genes in this endotype paired with their downregulation in Endotype B is interesting because β-agonists improve ciliary function. It is perhaps this aspect of ARDS and not enhanced alveo- lar fluid resorption that drive the positive findings in the BALTI-1 trial but were not replicated in BALTI-2 [26]. Further characterization of these ciliary findings are required.

Our findings in conjunction with other endotyping studies discredit the idea that ARDS progresses through predictable exudative, fibroproliferative, and resolution stages [27] but do lend credence to the concept that regardless of injury, there are stereotyped injury-response and repair pathways. In our data, subjects with similar ARDS triggers could be grouped into one of three endotypes. Similar findings were found in hyper-inflammatory and hypo-inflammatory groups in adult ARDS. While it's clearly important to resolve underlying ARDS triggers, emerging data indicates that identifying what type of lung response a subject is having to that ARDS trigger is likely preferable to a one-size-fits-all approach.

Several study strengths and weaknesses are noted. First, additional subjects and specimens would increase the confidence that only three endotypes exist. Endotype B consisted of only four subjects. However, the identification of these same endotypes at different time points and in a

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

largely distinct cohort of subjects and specimens increases our confidence that Endotypes A, B, and C represent true biologically distinct entities. Future studies should use a first-specimen approach to confirm these endotypes. Second, we did not restrict the time frame in which subjects could be enrolled. Our rationale for such an approach was that the underlying pathobiology of intubated ARDS subjects is well-established by the time of endotracheal intubation. However, it also likely introduced a greater number of subjects with resolving PARDS and caused an over representation of Endotype A. Third, we did not exclude subjects with baseline ventilator support or immunodeficiency so long as baseline oxygen requirement was less than 2 liters per minute. These patients are likely predisposed to worse PARDS with a given insult and this too may have caused over-representation of Endotype A. Lastly, although nasal transcriptomics reflect lung processes in other diseases, future studies should correlate nasal, bronchial, and bronchoalveolar lavage gene expression. A notable strength of our study that differentiates it from other ARDS endotyping studies to date is the inclusion of a critically-ill control group. This demonstrated that not all PARDS subjects have nasal gene exxpression profiles different from that of controls and that with PARDS resolution, the nasal transcriptome returns to normal.

7. References

1. Khemani, R. G., Smith, L. S., Zimmerman, J. J. & Erickson, S. Pediatric acute respiratory distress syndrome: definition, incidence, and epidemiology: proceedings from the Pediatric Acute Lung Injury Consensus Conference. *Pediatr Crit Care Med* **16**, S23–40 (2015). URL 10.1097/pcc.0000000000000432.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. *N Engl J Med* **342**, 1301–8 (2000). URL 10.1056/NEJM200005043421801.

3. Wiedemann, H. P. *et al.* Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med* **354**, 2564–75 (2006). URL NEJMoa062200[pii]10.1056/NEJMoa062200.

4. Gattinoni, L., Marini, J. J. & Quintel, M. Time to Rethink the Approach to Treating Acute Respiratory Distress Syndrome. *Jama* **319**, 664–666 (2018). URL 10.1001/jama.2017.21900.

5. Luo, L. *et al.* Clinical Predictors of Hospital Mortality Differ Between Direct and Indirect ARDS. *Chest* **151**, 755–763 (2017). URL 10.1016/j.chest.2016.09.004.

6. Taylor, M. S. *et al.* Delayed Alveolar Epithelialization: A Distinct Pathology in Diffuse Acute Lung Injury. *Am J Respir Crit Care Med* **197**, 522–524 (2018). URL 10.1164/rccm.201706-1094LE.

7. Willson, D. F., Truwit, J. D., Conaway, M. R., Traul, C. S. & Egan, E. E. The Adult Calfactant in Acute Respiratory Distress Syndrome Trial. *Chest* **148**, 356–364 (2015). URL 10.1378/chest.14-1139.

8. Willson, D. F. *et al.* Effect of exogenous surfactant (calfactant) in pediatric acute lung injury: a randomized controlled trial. *Jama* **293**, 470–6 (2005).

9. Calfee, C. S. *et al.* Distinct molecular phenotypes of direct vs indirect ARDS in single-center and multicenter studies. *Chest* **147**, 1539–1548 (2015). URL 10.1378/chest.14-2454.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

10. Jabaudon, M. *et al.* Receptor for advanced glycation end-products and ARDS prediction: a multicentre observational study. *Sci Rep* **8**, 2603 (2018). URL 10.1038/s41598-018-20994-x.

11. Yehya, N. *et al.* Circulating markers of endothelial and alveolar epithelial dysfunction are associated with mortality in pediatric acute respiratory distress syndrome. *Intensive Care Med* **42**, 1137–45 (2016). URL 10.1007/s00134-016-4352-1.

12. Gallagher, D. C. *et al.* Circulating angiopoietin 2 correlates with mortality in a surgical population with acute lung injury/adult respiratory distress syndrome. *Shock* **29**, 656–61 (2008). URL 10.1097/shk.0b013e31815dd92f.

13. Famous, K. R. *et al.* Acute Respiratory Distress Syndrome Subphenotypes Respond Differently to Randomized Fluid Management Strategy. *Am J Respir Crit Care Med* **195**, 331–338 (2017). URL 10.1164/rccm.201603-0645OC.

14. Wong, H. R. *et al.* Validation of a gene expression-based subclassification strategy for pediatric septic shock. *Crit Care Med* **39**, 2511–7 (2011). URL 10.1097/CCM.0b013e3182257675.

15. Morrell, E. D. *et al.* Peripheral and Alveolar Cell Transcriptional Programs Are Distinct in Acute Respiratory Distress Syndrome. *Am J Respir Crit Care Med* **197**, 528–532 (2018). URL 10.1164/rccm.201703-0614LE.

16. Silvestri, G. A. *et al.* A Bronchial Genomic Classifier for the Diagnostic Evaluation of Lung Cancer. *N Engl J Med* **373**, 243–51 (2015). URL 10.1056/NEJMoa1504601.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

17. Kuo, C. S. *et al.* A Transcriptome-driven Analysis of Epithelial Brushings and Bronchial Biopsies to Define Asthma Phenotypes in U-BIOPRED. *Am J Respir Crit Care Med* **195**, 443–455 (2017). URL 10.1164/rccm.201512-2452OC.

18. Boudewijn, I. M. *et al.* Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression. *Respiratory Research* **18**, 213 (2017). URL 10.1186/s12931-017-0696-5.

19. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). URL 10.1093/bioinformatics/bts635.

20. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014). URL 10.1186/s13059-014-0550-8.

21. Kolde, R. pheatmap: Pretty Heatmaps (2019).

22. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).

23. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, 305–11 (2009). URL 10. 1093/nar/gkp427.

24. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014). URL 10.1093/bioinformatics/btt703.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

25. Harrison, E., Drake, T. & Ots, R. Quickly Create Elegant Regression Results Tables and Plots when Modelling. URL https://finalfit.org/index.html.

26. Smith, F. G. *et al.* Effect of intravenous beta-2 agonist treatment on clinical outcomes in acute respiratory distress syndrome (BALTI-2): a multicentre, randomised controlled trial. *Lancet* **379**, 229–35 (2012). URL 10.1016/S0140-6736(11)61623-1.

27. Ware, L. B. & Matthay, M. A. The Acute Respiratory Distress Syndrome. *New England Journal of Medicine* **342**, 1334–1349 (2000). URL 10.1056/NEJM200005043421806.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083451; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1 - Group Descriptive Data

label	levels	ARDS	Control	p
Age	Median (IQR)	5.3 (6.8)	3.3 (5.4)	0.280
Sex	Female	8 (44.4)	5 (38.5)	1.000
	Male	10 (55.6)	8 (61.5)	
Race	Non-White	4 (22.2)		0.120
	White	14 (77.8)	13 (100.0)	
Genetic Synd. or D.D.	No	9 (50.0)	9 (69.2)	0.462
	Yes	9 (50.0)	4 (30.8)	
Baseline Lung Disease	No	12 (66.7)	7 (53.8)	0.710
	Yes	6 (33.3)	6 (46.2)	
Baseline Kidney or Liver Disease	No	17 (94.4)	11 (84.6)	0.558
	Yes	1 (5.6)	2 (15.4)	
Immunocompromised	No	11 (61.1)	12 (92.3)	0.095
	Yes	7 (38.9)	1 (7.7)	
Highest PELOD2	Median (IQR)	7.0 (4.0)	5.0 (0.0)	0.016
Highest PARDS Category	Mild	3 (16.7)	1 (7.7)	<0.001
	Moderate	10 (55.6)		
	Severe	5 (27.8)		
	None		12 (92.3)	
Any Corticosteroids	No	6 (33.3)	10 (76.9)	0.029
	Yes	12 (66.7)	3 (23.1)	
Ventilator Free Days	Median (IQR)	13.9 (17.3)	14.9 (23.9)	0.935
PICU Days	Median (IQR)	18.0 (14.5)	6.0 (13.0)	0.004
Hospital Days	Median (IQR)	33.0 (87.8)	18.0 (21.0)	0.012
Mortality	Died	4 (22.2)		0.120
	Survived	14 (77.8)	13 (100.0)	