

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

1 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 ClinicalTrials.gov (NCT#03539783).

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, Camarillo, CA, for patients >20 kg) or 3.0 mm Nasal Cytology Brush (Hobbs Medical, Stafford

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 -20°C for up to a week and then at -80°C . For RNA extraction, $10\mu\text{L}$ of β – *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 $50\mu\text{L}$ of nuclease-free water per manufacturer protocol. RNA quality and quantity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies).

Standard Library Preparation and Sequencing Specimen ribosomal RNA was depleted using 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 NEB-Next SingleCell/Low Input RNA Library Prep kit (New England Biolabs) per manufacturer in-

structions 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 ² with analysis of differential expression using DESeq2 ³. 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 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 ⁴ and ggplot2 ⁵ in R 3.6.3 ⁶. 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 ⁷ and Ingenuity Pathway Analysis (v 20.0) ⁸ with visualization of results using pheatmap.

Luminex Analysis Serum samples were analyzed using two different plates at 1:2 and 1:100 dilution respectively. The Luminex beads are the proprietary products of R&D Systems and were ordered by analyte as a custom assay.

Plate 1: Angiopoietin-2, Granzyme B, Intercellular Adhesion Molecule-1, Interferon gamma, Interferon alpha, Interleukin-10, Interleukin-17, Interleukin-18, Interleukin-6, Interleukin-8, Surfactant Protein D, Tumor Necrosis Factor-alpha, Tumor necrosis factor receptor soluble factor 1A, Receptor of Advanced Glycation End-products (RAGE)

Plate 2: Matrix Metalloproteinase-9, Plasminogen activator inhibitor-1, Myeloperoxidase

Serum samples were analyzed with standard curves and negative control on a Luminex 200 instrument using manufacturer protocol.

Statistical Analysis For bioinformatic analyses a Bonferroni-corrected p-value of less than 0.1 was considered significant. All analyses were performed and plots were generated using mainly R statistical package (version 3-5-3)⁹ and ggplot2 (v 3.3.0),¹⁰ ggpubr (v 0.3.3),¹¹ dunn.test (v1.3.5),¹² and finalfit (v 1.0.0)¹³ packages with the above DESeq2 analysis of differentially expressed genes. For all statistical 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 with Dunn's *post hoc* test for continuous data. Only non-parametric tests were used for consistency although some datasets were parametric.

2 Supplemental Data

Description of Failed mRNA-Seq Techniques We arrived at our current protocol after several failures. Library creation was successful in only 13 of 45 (29%) of RNA specimens using Ri-

boZero ribosomal RNA depletion, Illumina NexteraXT kit and HiSeq 2500 sequencing. To increase yield we performed ribosomal RNA depletion using NEBNext rRNA Depletion Kit (New England Biolabs) and amplified specimen RNA using the SeqPlex RNA kit (Sigma) with 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.

Serum Biomarker Data Violin plots of analyzed serum biomarkers by group assignment, nasal transcriptomic endotype, and PARDS Severity (Figure S1).

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**, 23–40 (2015). URL [10.1097/pcc.0000000000000432](https://doi.org/10.1097/pcc.0000000000000432).
2. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). URL [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635).

3. 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](https://doi.org/10.1186/s13059-014-0550-8).
4. Kolde, R. pheatmap: Pretty Heatmaps (2019).
5. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
6. Team, R. C. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2013).
7. 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](https://doi.org/10.1093/nar/gkp427).
8. 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](https://doi.org/10.1093/bioinformatics/btt703).
9. Team, R. C. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2013).
10. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
11. Kassambara, A. ggpubr: Publication Ready Plots - Articles - STHDA (2020). URL <https://CRAN.R-project.org/package=ggpubr>. R package version 0.3.0.
12. Dinno, A. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums.

13. Harrison, E., Drake, T. & Ots, R. Quickly Create Elegant Regression Results Tables and Plots when Modelling.

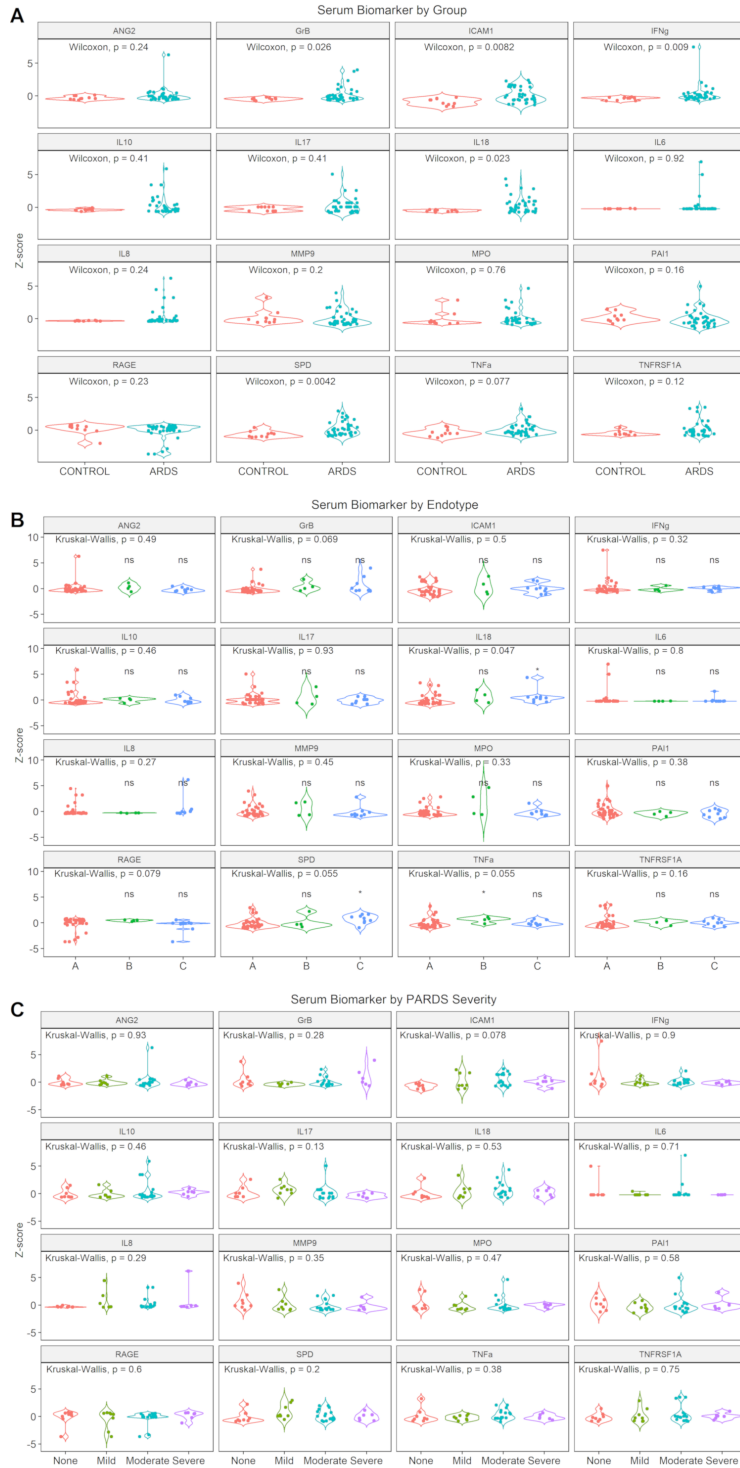


Figure 1: Serum Biomarker Levels Specimens From Subjects by (A) Group, (B) Endotype, and (C) PARDS Severity