



OPEN ACCESS

EDITED BY

Reinout A. Bem,
Amsterdam University Medical Center,
Netherlands

REVIEWED BY

Jerry John Zimmerman,
Seattle Children's Hospital, United States
Mary E. Hartman,
Washington University in St. Louis, United States

*CORRESPONDENCE

Jocelyn R. Grunwell
✉ jgrunwe@emory.edu

[†]These authors have contributed equally to this work and share first authorship

[‡]Deceased

SPECIALTY SECTION

This article was submitted to Pediatric Critical Care, a section of the journal Frontiers in Pediatrics

RECEIVED 06 February 2023

ACCEPTED 01 March 2023

PUBLISHED 17 March 2023

CITATION

Grunwell JR, Rad MG, Ripple MJ, Yehya N, Wong HR and Kamaleswaran R (2023) Identification of a pediatric acute hypoxemic respiratory failure signature in peripheral blood leukocytes at 24 hours post-ICU admission with machine learning. *Front. Pediatr.* 11:1159473. doi: 10.3389/fped.2023.1159473

COPYRIGHT

© 2023 Grunwell, Rad, Ripple, Yehya, Wong and Kamaleswaran. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of a pediatric acute hypoxemic respiratory failure signature in peripheral blood leukocytes at 24 hours post-ICU admission with machine learning

Jocelyn R. Grunwell^{1,2*†}, Milad G. Rad^{3†}, Michael J. Ripple^{1,2}, Nadir Yehya^{4,5}, Hector R. Wong^{6†} and Rishikesan Kamaleswaran^{2,7,8}

¹Division of Critical Care Medicine, Children's Healthcare of Atlanta, Atlanta, GA, United States, ²Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, United States, ³Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, United States, ⁴Department of Anesthesiology and Critical Care Medicine, University of Pennsylvania, Philadelphia, PA, United States, ⁵Division of Pediatric Intensive Care Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, United States, ⁶Division of Critical Care Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, ⁷Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, United States, ⁸Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States

Background: There is no generalizable transcriptomics signature of pediatric acute respiratory distress syndrome. Our goal was to identify a whole blood differential gene expression signature for pediatric acute hypoxemic respiratory failure (AHRF) using transcriptomic microarrays within twenty-four hours of diagnosis. We used publicly available human whole-blood gene expression arrays of a Berlin-defined pediatric acute respiratory distress syndrome (GSE147902) cohort and a sepsis-triggered AHRF (GSE66099) cohort within twenty-four hours of diagnosis and compared those children with a $P_aO_2/F_iO_2 < 200$ to those with a $P_aO_2/F_iO_2 \geq 200$. **Results:** We used stability selection, a bootstrapping method of 100 simulations using logistic regression as a classifier, to select differentially expressed genes associated with a $P_aO_2/F_iO_2 < 200$ vs. $P_aO_2/F_iO_2 \geq 200$. The top-ranked genes that contributed to the AHRF signature were selected in each dataset. Genes common to both of the top 1,500 ranked gene lists were selected for pathway analysis. Pathway and network analysis was performed using the Pathway Network Analysis Visualizer (PANEV) and Reactome was used to perform an over-representation gene network analysis of the top-ranked genes common to both cohorts. Changes in metabolic pathways involved in energy balance, fundamental cellular processes such as protein translation, mitochondrial function, oxidative stress, immune signaling, and inflammation are differentially regulated early in pediatric ARDS and sepsis-induced AHRF compared to both healthy controls and to milder acute hypoxemia. Specifically, fundamental pathways related to the severity of hypoxemia emerged and included (1) ribosomal and eukaryotic initiation of factor 2 (eIF2) regulation of protein translation and (2) the nutrient, oxygen, and energy sensing pathway, mTOR, activated *via* PI3K/AKT signaling.

Conclusions: Cellular energetics and metabolic pathways are important

Abbreviations

AHRF, acute hypoxemic respiratory failure; ARDS, acute respiratory distress syndrome; AUPRC, area under the precision recall curve; AUROC, area under the receive operating characteristic curve; eIF2, eukaryotic initiation of factor 2; GEO, Gene Expression Omnibus; HDAC, histone deacetylase; IPA, Ingenuity Pathway Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; PANEV, Pathway Network Visualizer; PEEP, positive end-expiratory pressure; PICU, pediatric intensive care unit; TCA, tricarboxylic acid.

mechanisms to consider to further our understanding of the heterogeneity and underlying pathobiology of moderate and severe pediatric acute respiratory distress syndrome. Our findings are hypothesis generating and support the study of metabolic pathways and cellular energetics to understand heterogeneity and underlying pathobiology of moderate and severe acute hypoxemic respiratory failure in children.

KEYWORDS

pediatric, acute respiratory distress syndrome, mechanical ventilation, gene expression profiling, machine learning, transcriptomics

1. Introduction

Clinical and biological heterogeneity are hallmark features of acute respiratory distress syndrome (ARDS), and correlating clinical phenotypes with biological endotypes using plasma biomarkers and clinical variables using machine learning methods has led to the discovery of a hyperinflammatory (reactive) and hypoinflammatory (unreactive) ARDS endotype with differential responses to high positive end-expiratory pressure (PEEP), low tidal-volume and restrictive fluid therapies (1–4). Systemic infections, such as sepsis, often caused by pneumonia, are a common trigger of ARDS; however, trauma, aspiration, near-drowning, burns, and blood transfusions are also instigators of ARDS.

Statistical and machine learning approaches have been used to perform a meta-analysis of publicly available whole-blood gene expression data to predict mortality from sepsis in adults and children (5). A multi-cohort analysis of adult and pediatric whole-blood gene expression data did not find a generalizable transcriptomics signature of ARDS (6), perhaps due to differences in the analytical approach, the cohort selection, or in the combination of children and adult data. Machine learning random forest models applied to clinically available laboratory and physiologic data within the first two days of intensive care outperformed multivariable logistic regression at predicting prolonged acute hypoxemic respiratory failure (AHRF) of a week or more following influenza infection in a large multicenter cohort of children enrolled in the PICFLU observational study (7). We have used machine learning to explore differentially expressed genes in the primary airway cells from tracheal aspirate samples of mechanically ventilated children with ARDS (8). Since publication of the ARDS transcriptomic meta-analysis incorporating adult ARDS and pediatric sepsis-triggered AHRF (6), pediatric ARDS-specific whole-blood microarray gene expression cohort is now publicly available, thus providing an opportunity to further explore the existence of pediatric ARDS-specific transcriptomic signatures (9).

The P_aO_2/F_iO_2 ratio is the ratio of arterial oxygen partial pressure (P_aO_2 in mmHg) to fractional inspired oxygen (F_iO_2 expressed as a fraction, not a percentage). The primary objective of this study was to determine a gene expression signature of mild ($P_aO_2/F_iO_2 \geq 200$) compared with moderate/severe ($P_aO_2/F_iO_2 < 200$) AHRF from two publicly available whole blood gene expression microarray datasets. Secondary objectives were to explore the common gene networks of moderate/severe vs. mild AHRF.

2. Methods

2.1. Ethics approval and consent to participate

Only deidentified clinical data was used in this study. The parent study protocols for collection and use of biological specimens and clinical data were approved by the Institutional Review Board (IRB) at the Children's Hospital of Philadelphia (CHOP IRB 13-010578, "Biomarkers of pediatric ARDS", Approved 7/2/2014; GSE147902) and by the IRBs of each of the 18 participating institutions for the sepsis AHRF cohort [Cincinnati Children's Hospital Medical Center parent protocol approved by Emory IRB 00079159, "Genomic Analysis of Pediatric Systemic Inflammatory Response Syndrome (SIRS)", Approved 02/21/2015; GSE66099]. Written informed consent was obtained from legal caregivers prior to enrollment in the parent studies (10–14). Procedures were followed in accordance with the ethical standards of the responsible committee on human experimentation (institutional or regional) and with the Helsinki Declaration of 1975.

2.2. Data collection

The pediatric sepsis dataset, GSE66099, and the pediatric ARDS dataset, GSE147902 were downloaded from the NCBI Gene Expression Omnibus (GEO) repository (9, 15) (<https://www.ncbi.nlm.nih.gov/geo/>). Gene expression data was deidentified and publicly available from the aforementioned two independent observational cohort studies. Both datasets contain gene expression profiles obtained from peripheral whole blood samples of patients who were admitted to the pediatric intensive care unit (PICU) within 24 h of a diagnosis of sepsis or ARDS using the Berlin definition, respectively. Children in GSE66099 were stratified by having a $P_aO_2/F_iO_2 < 200$ (vs. ≥ 200) and defined as having acute hypoxemic respiratory failure as chest radiograph findings were not available to ensure that all ARDS criteria were satisfied. In addition, some children in the GSE66099 cohort did not meet the hypoxemia threshold used in the Berlin definition of ARDS ($P_aO_2/F_iO_2 < 300$). By definition, children enrolled in the GSE147902 had to meet all Berlin criteria for ARDS (16). For this present analysis, AHRF was defined in both cohorts as a $P_aO_2/F_iO_2 < 200$. Pediatric-specific consensus criteria for septic shock and non-pulmonary organ failures were used to determine eligibility for study enrollment (17).

2.3. Normalization and background correction

The Affy package from R was used to remove technical variations in the gene expression data (18). Background noise correction and normalizations were performed using the R package gcrma (19). Surrogate Variable Analysis using the R package sva was used to correct for batch effects variation between datasets (20).

2.4. Probe to gene mapping

The GSE66099 and GSE147902 microarray dataset Affymetrix probes were matched to gene symbols using Affymetrix Human Genome U133 Plus 2.0 (hgu133plus2.db) and Human Gene 2.1 ST Array (hugene21sttranscriptcluster.db), respectively. Genes with multiple probes were matched to the same gene and the expressions were averaged (5).

2.5. Variable selection, classification models, and validation

Graphical modeling or cluster analysis of high-dimensional data [features (p) >> number of samples (n)]. Stability selection is a “wrapper” algorithm which combines a variety of distinct feature selection algorithms (LASSO, Decision Tree, Recurrent Feature Elimination, etc.) from the Python scikit-learn package to address the nature of this high-dimension data reduction problem (21). This approach is similar to popular “Baruta” R package (22). Stability selection can be seen as the python alternative, which uses subsampling (bootstrap sampling) of the data in combination with a feature selection algorithm such as LASSO to find the most important features in every sampled version of the data. The aggregate results from bootstrap sampling are used to calculate a stability score for each feature in the data. Features were then selected by choosing an appropriate threshold for the stability scores (importance coefficient). We selected candidate markers using stability selection by bootstrapping 100 simulations from all of the data to select stable gene markers using logistic regression as a classifier, with 0.6 as the threshold for selection to ensure enough genes were available for the primary analysis (21). The top-ranked genes that contributed to the AHRF signature were selected in each dataset. Genes common to both of the top 1,500 ranked gene lists were selected for pathway analysis. Leave-one-out cross-validation was used to validate the primary model. Area under the receiver operating characteristic (AUROC) and precision-recall curves were generated using the Python package scikit-learn 1.1.2 (23).

2.6. Pathway analysis

The most common pathways were explored using the 185 top-ranked genes common to both datasets using the Kyoto

Encyclopedia of Genes and Genomes (KEGG) (24, 25). The identified KEGG pathways are user defined first-level pathways that were fed into the R package Pathway Network Visualizer (PANEV) to identify and map the second- and third-level pathway networks associated with AHRF (26). The Python Reactome package was used to perform an over-representation gene network analysis using the 185 top-ranked genes common to both datasets (27, 28). We also explored canonical pathway enrichment, upstream regulator, causal network analysis, and molecular and cellular functions with the 185 top-ranked genes common to both datasets using the Ingenuity Pathway Analysis software (IPA, Qiagen, Redwood City, CA). The data analysis workflow using machine learning methods for common dataset gene selection and pathway analysis is shown in **Figure 1**.

2.7. Protein-protein interaction network analysis

We visualized the protein-protein interaction network using the software package Cytoscape 3.2.0 (29). Nodes represent proteins and edges represent interactions between two proteins. Hub genes in the protein-protein interaction network were defined at those with at least a degree of connectivity greater than or equal to ten. These analyses are available in the github repository at <https://github.com/ghiasirad/pahrf.git> as PPIN.pdf (**Supplementary File 5**), PPIN.sif (**Supplementary File 6**), and PPIN.csv (**Supplementary File 7**) files.

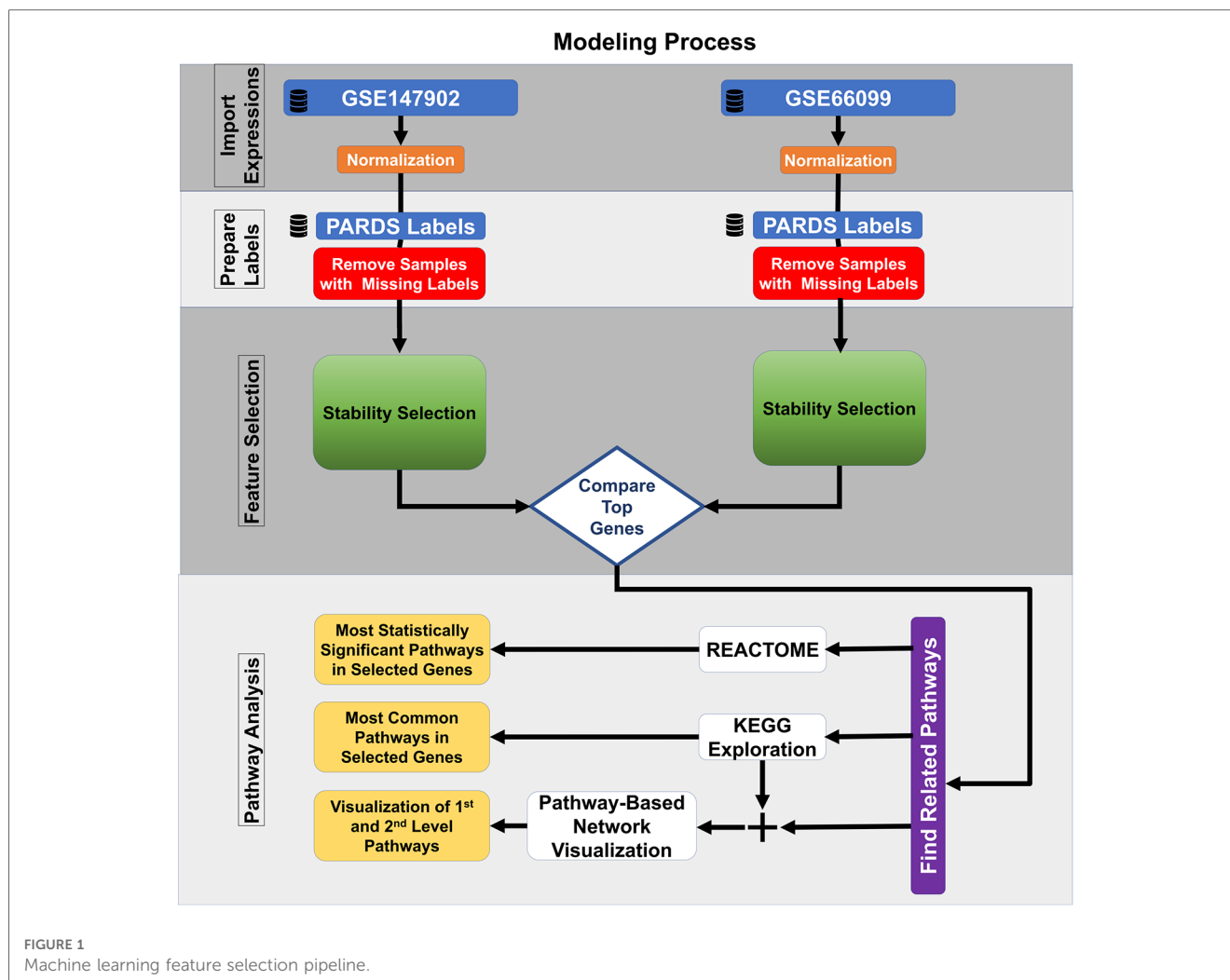
3. Results

3.1. Cohort description

There were forty-seven healthy control children, sixty-five children without (i.e., $P_aO_2/F_iO_2 \geq 200$) and ninety-six children with ($P_aO_2/F_iO_2 < 200$) sepsis-related AHRF from GSE66099 (sepsis dataset). There were twenty-six children without (P_aO_2/F_iO_2 201–300) and seventy children with ($P_aO_2/F_iO_2 < 200$) AHRF in GSE147902 (pediatric ARDS dataset). The available demographics and clinical characteristics of the participants stratified by P_aO_2/F_iO_2 ratio are summarized in **Table 1**.

3.2. Selection of common genes of an AHRF signature

Clinically, we were most interested in determining the underlying gene transcripts and molecular pathways that distinguish the most severely hypoxemic and critically ill children with $P_aO_2/F_iO_2 < 200$ mmHg from those with milder hypoxemia ($P_aO_2/F_iO_2 \geq 200$ mmHg). We first determined whether the differential gene expression pattern discovered by comparing P_aO_2/F_iO_2 200–300 with $P_aO_2/F_iO_2 < 200$ from GSE147902 could be applied to a cohort of children with AHRF secondary to sepsis. There was no overlap in differentially expressed genes for



children with a $P_aO_2/F_iO_2 < 200$ vs. a $P_aO_2/F_iO_2 \geq 200$ using the Benjamini-Hochberg false discovery rate for multiple-hypothesis corrections (30). We therefore used a machine learning stability selection approach to perform this comparison (21, 31).

We independently ranked the differentially expressed genes by $P_aO_2/F_iO_2 < 200$ vs. $P_aO_2/F_iO_2 \geq 200$ using the machine learning stability select algorithm (21, 31). From the top-ranked 1,500 genes in each data set, genes common to both ranked gene sets were selected in ten gene increments as depicted in **Figure 2A**. The optimal number of common genes was determined by plotting the area under the receiver operative curve (ROC) characteristic for each incremental addition of ten genes to the AHRF model as shown in **Figure 2B**. The red line in **Figure 2B** shows the change in the ROC with the addition of common genes for the pediatric ARDS dataset (GSE147902). The black line in **Figure 2B** shows the change in the ROC with the addition of common genes for the sepsis dataset (GSE66099). The ROC plateaus around 25–30 genes for each data set. The area under the receiver operating curves (AUROC) and area under the precision recall curves (AUPRC) for GSE147902 (**Figures 3A,B**) and for GSE66099 (**Supplementary Figures S1A,B**) are shown for the optimal

common genes for each incremental addition of ten genes to the model. For both datasets the AUROC was maximized at 21 genes with a value of 0.82 (95% CI: 0.72–0.89) for GSE147902 and 0.76 (95% CI: 0.69–0.82) for GSE66099 (genes listed in **Supplementary Data**).

Due to the imbalance in the proportion of cases with and without AHRF in the two cohorts, we report the area under the precision recall curve (AUPRC) for the model of $P_aO_2/F_iO_2 < 200$ vs. $P_aO_2/F_iO_2 \geq 200$. The AUPRC measures the ability of the models to correctly distinguish $P_aO_2/F_iO_2 \geq 200$ from $P_aO_2/F_iO_2 < 200$ in both GSE66099 and GSE147902. The pediatric ARDS GSE147902 dataset consists of 27% of children with a $P_aO_2/F_iO_2 \geq 200$ while the sepsis GSE66099 dataset consists of 40% of children with a $P_aO_2/F_iO_2 \geq 200$. The AUPRC was maximized at 92-genes with a value of 0.90 (95% CI: 0.75–0.91) for GSE147902 and 0.66 (95% CI: 0.5–0.76) for GSE66099.

3.3. Network analysis

The 185 genes selected using stability selection method were ranked by normalized importance coefficient for each dataset.

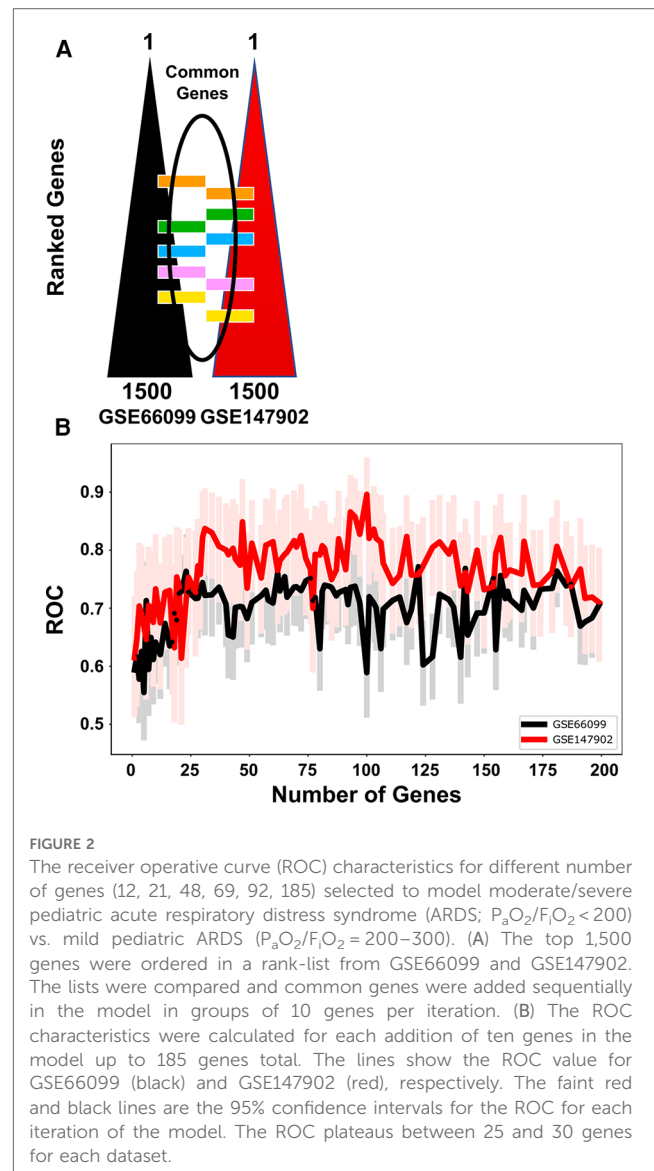
TABLE 1 Demographic and clinical characteristics of cohort participants.

Characteristics, <i>n</i> (%) or median (Q1, Q3)	GSE66099 P _a O ₂ /F _i O ₂		GSE147902 P _a O ₂ /F _i O ₂	
	≥200 <i>n</i> = 99	<200 <i>n</i> = 65	200–299 <i>n</i> = 26	<200 <i>n</i> = 70
Age (year)	2.4 (1.15, 5.75)	2.4 (0.8, 5.9)	6.7 (0.8, 14.1)	9.4 (2.9, 14.5)
Sex				
Female	39 (39.4)	29 (44.6)	7 (27)	31 (44)
Male	60 (60.6)	36 (55.4)	19 (73)	39 (56)
Race/Ethnicity				
Black	– ^a	– ^a	4 (15)	26 (37)*
White	– ^a	– ^a	15 (58)	24 (34)*
Hispanic	– ^a	– ^a	6 (23)	8 (11)*
Asian/Pacific Islander	– ^a	– ^a	0 (0)	3 (4.3)*
Other	– ^a	– ^a	1 (3.8)	9 (13)*
Severity of illness				
Pediatric Risk of Mortality III	13.0 (9, 19)	16.0 (11, 22)	10 (7, 17)	11 (6, 20)
Non-pulmonary organ failures, <i>n</i> (min, max)	2 (1, 6)*	1 (1, 6)	1 (1, 2.75)	1 (1, 2)
Comorbidities				
Any comorbidity	43 (44.3)	21 (32.3)	21 (81)	44 (63)
Malignancy	6 (6.2)	3 (4.6)	6 (29)	14 (32)
Immunocompromised	9 (9.3)	7 (10.8)	8 (31)	22 (31)
Bone Marrow or Stem Cell Transplant	4 (4.1)	2 (3.1)	5 (19)	10 (14)
Clinical Trigger				
Infectious pneumonia	14 (14.1)	17 (26.2)	13 (50)	43 (61)
Non-pulmonary sepsis	85 (85.9)	73 (73.8)	9 (35)	10 (14)
Aspiration pneumonia	NA	NA	1 (3.8)	9 (12.9)
Trauma	NA	NA	0 (0)	3 (4.3)
Other	NA	NA	3 (11.5)	5 (7.1)
Outcomes				
PICU mortality	9 (9.1)	13 (20.0)*	8 (31)	12 (17)
Complicated course	16 (16.7)	29 (44.6)*	9 (35)	19 (27)
28-day ventilator-free days	– ^a	– ^a	20 (0, 23)	18 (1, 22)

^aData is not available for reporting.

*Wilcoxon rank sum test; Fisher’s exact test; Pearson’s χ^2 test, *p* < 0.05.

The genes retained in the primary model all had importance coefficient scores greater than 0.9. The selected genes, ranked by the number of genes involved in the KEGG pathway, are shown in **Figure 4**. The top 3 KEGG pathways included metabolic, ribosomal, and coronavirus/COVID-19 disease pathways. Metabolic pathways involve the intricate networks of glycolysis, the tricarboxylic acid (TCA) cycle, the pentose phosphate pathway, gluconeogenesis, fatty acid oxidation and lipid metabolism, one-carbon metabolism, amino acid and cofactor metabolism, and nucleic acid metabolism for the production of energy and substrates for fundamental biological reactions. The coronavirus/COVID-19 disease pathway consists of anti-viral detection *via* the toll-like receptor (TLR) and retinoic acid-inducible gene I (RIG-I) pattern recognition receptor pathways, activation of the NLRP3 inflammasome, and signaling through the pro-inflammatory cytokines IL-6, JAK/STAT, PI3K, and MAPK signaling cascades, and activation of the transcription



factor NF-KB. We supported the KEGG pathway analysis using Ingenuity Pathway Analysis (IPA) software, and the top canonical IPA pathways included: the Coronavirus pathogenesis pathway (*p*-value 3.70e-07), Eukaryotic Initiation Factor 2 (eIF2) signaling (*p*-value 4.62e-05), mitochondrial dysfunction (*p*-value 2.95e-04), mammalian target of rapamycin (mTOR) signaling (*p*-value 5.16e-03), and cell cycle control of chromosomal replications (*p*-value 8.6e-03) (**Supplementary Table S1**). Upstream regulators included La Ribonucleoprotein 1 (*LARP1*, *p*-value 3.95e-07), Mitogen-Activated Protein Kinase Kinase Kinase 4 (*MAP4K4*, *p*-value 4.92e-06), Rapamycin-insensitive companion of mammalian target of rapamycin (*RICTOR*, *p*-value 9.91e-05), carbohydrate-responsive element-binding protein (*ChREBP*) also known as MLX-interacting protein-like (*MLXIPL*, *p*-value 2.60e-05), and Brain cytoplasmic 200 long-noncoding RNA (BC200 lncRNA or *BCYRN1*, *p*-value 1.76e-04). The main molecular and cellular functions included protein synthesis

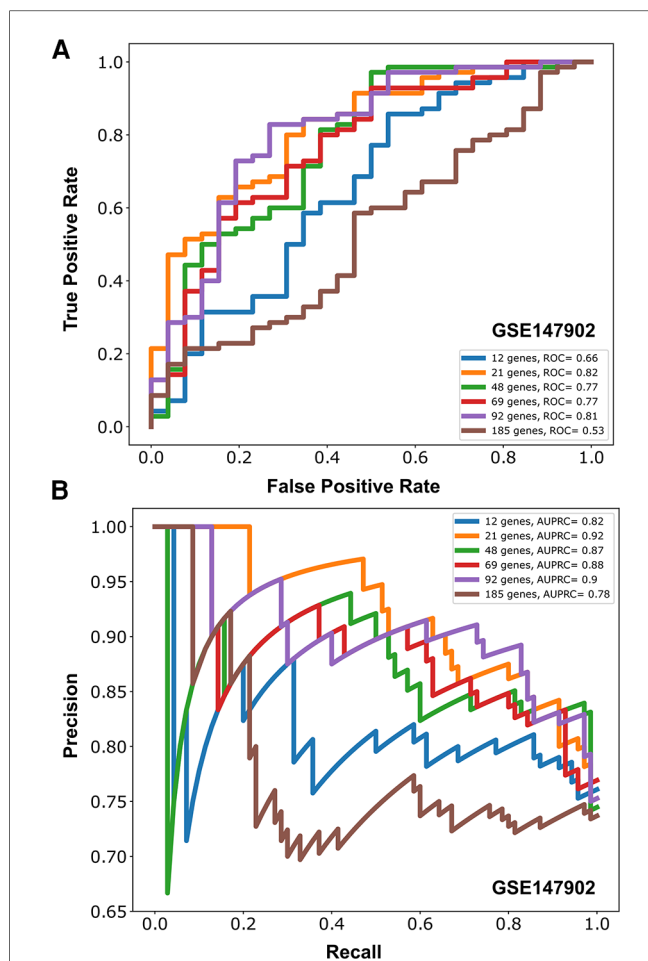


FIGURE 3
Area under the receiver operative curve (AUROC) (A) and area under the precision recall curve (AUPRC) (B) for different number of genes (12, 21, 48, 69, 92, 185) selected to model moderate/severe pediatric acute respiratory distress syndrome (ARDS; $P_{aO_2}/F_{iO_2} < 200$) vs. mild pediatric ARDS ($P_{aO_2}/F_{iO_2} = 200-300$) from GSE147902.

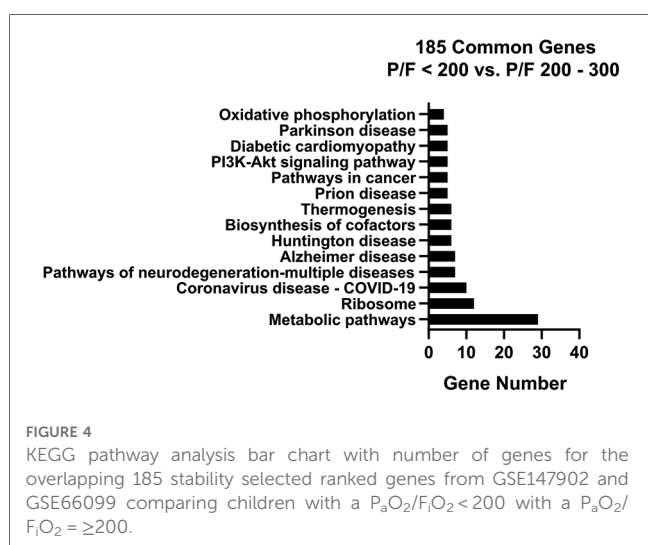


FIGURE 4
KEGG pathway analysis bar chart with number of genes for the overlapping 185 stability selected ranked genes from GSE147902 and GSE66099 comparing children with a $P_{aO_2}/F_{iO_2} < 200$ with a $P_{aO_2}/F_{iO_2} \geq 200$.

(p -value range: $1.49e-02-7.08e-13$), RNA Damage and Repair (p -value range: $1.48e-06-1.48e-06$), Cell Death and Survival (p -value range: $2.97e-02-6.06e-05$), Cell Morphology (p -value range: $2.97e-02-6.06e-05$), and Cellular Function and Maintenance (p -value range: $2.97e-02-6.06e-05$). The top networks were RNA Damage and Repair, Protein Synthesis, RNA Post-Transcriptional Modification (Score 47), Cell Death and Survival, Protein Synthesis, Cancer (Score 44), Protein Synthesis, Cancer, Hematological Disease (Score 20), Skeletal and Muscular System Development and Function, Tissue Morphology, Cell Death and Survival (Score 18), and Cancer, Endocrine System Disorders, Organismal Injury and Abnormalities (Score 16).

We next visualized networks of pathways involved in the moderate/severe AHRF gene expression signature by imputing the list of 185 common genes (Supplementary File 1) into the PANEV package in R. The first-, second-, and third-level pathways with the corresponding genes contributing to each pathway are shown in Supplementary Files 2–5 (Supplementary Files can be found at: <https://github.com/ghiasirad/pahrf.git>).

Finally, we used the curated Reactome database to perform a pathway over-representation analysis using the 185 common genes defining the moderate/severe AHRF gene expression signature. There were nineteen significantly enriched pathways shown in Table 2 with cytosolic and mitochondrial protein translation, selenoamino acid metabolism, selenocysteine synthesis, and nonsense codon mediated decay of messenger RNA. In addition of gene network analysis, we performed a protein-protein interaction network analysis using 185 top-ranked genes common to both datasets. The protein-protein interaction network analysis showed seventeen proteins with connections to ten or more other proteins (Supplementary Files 6–8).

We also examined the differential gene expression of pediatric ARDS severity from the pediatric ARDS-specific dataset GSE147902 compared to healthy controls from the GSE66099 dataset. Volcano plots of the mild PARDS vs. healthy controls and moderate/severe PARDS vs. healthy controls are shown in Supplementary Figure S2A,B, respectively. Bar graphs of the number of genes in each KEGG pathway for mild ($P_{aO_2}/F_{iO_2} 200-300$ from GSE147902) vs. healthy controls from GSE66099 (Supplementary Figure S3A) and for moderate/severe ($P_{aO_2}/F_{iO_2} < 200$ from GSE147902) vs. healthy controls from GSE66099 (Supplementary Figure S3B) are shown. Furthermore, we compared the differential gene expression of moderate/severe ($P_{aO_2}/F_{iO_2} < 200$) compared to mild ($P_{aO_2}/F_{iO_2} 200-300$) pediatric ARDS in GSE147902 summarized as a bar graph with the number of genes represented in each KEGG pathway (Supplementary Figure S4). The Reactome pathways sorted by p -value for the top 200 overlapping stability selected ranked genes from GSE147902 comparing children with a $P_{aO_2}/F_{iO_2} < 200$ with a $P_{aO_2}/F_{iO_2} \geq 200-300$ are shown in Supplementary Table S2. These analyses are presented and the findings discussed in the online Supplementary Materials (<https://github.com/ghiasirad/pahrf.git>).

TABLE 2 Reactome pathways sorted by p -value for the top 185 overlapping stability selected ranked genes from GSE147902 and GSE66099 comparing children with a $P_aO_2/F_iO_2 < 200$ with a $P_aO_2/F_iO_2 \geq 200-300$.

Pathway name	Entities			Reactions	
	Found	Ratio	FDR ^a	Found	Ratio
Peptide chain elongation	12/134	0.006	8.99e-04	5/5	3.70e-04
Selenoamino acid metabolism	12/198	0.009	0.003	5/33	0.002
Formation of a pool of free 40S subunits	9/111	0.005	0.003	2/2	1.48e-04
Eukaryotic Translation Termination	9/113	0.005	0.003	3/5	3.70e-04
Selenocysteine synthesis	9/115	0.005	0.003	2/7	5.18e-04
SRP-dependent cotranslational protein targeting to membrane	9/119	0.005	0.003	5/5	3.70e-04
Nonsense mediated Decay independent of the Exon Junction Complex	9/121	0.006	0.003	1/1	7.40E-05
Eukaryotic Translational Elongation	12/187	0.009	0.003	8/9	6.66e-04
L13a-mediated translational silencing of Ceruloplasmin expression	9/124	0.006	0.003	2/3	2.22e-04
Viral mRNA Translation	10/135	0.006	0.005	2/2	1.48e-04
Translation	24/755	0.035	0.011	52/99	0.007
GTP hydrolysis and joining of the 60S ribosomal subunit	9/166	0.008	0.018	3/3	2.22e-04
Ribosomal scanning and start codon recognition	6/70	0.003	0.019	2/2	1.48e-04
Mitochondrial translation elongation	8/135	0.006	0.02	5/8	5.92e-04
Mitochondrial translation termination	8/137	0.006	0.02	5/5	3.70e-04
Mitochondrial translation initiation	9/181	0.008	0.025	3/4	2.96e-04
Formation of the ternary complex, and subsequently, the 43S complex	6/86	0.004	0.039	1/3	2.22e-04
Nonsense-Mediated Decay	11/197	0.009	0.039	5/6	4.44e-04
Nonsense Mediated Decay enhanced by the Exon Junction Complex	11/197	0.009	0.039	4/5	3.70e-04

^aFalse Discovery Rate.

4. Discussion

We used publicly available microarray gene expression data sets and machine learning methods to determine metabolic and immune pathways involved in the systemic response of children within 24 h of an ARDS diagnosis. We evaluated the overlap amongst genes differentiating children with a $P_aO_2/F_iO_2 < 200$ vs. a $P_aO_2/F_iO_2 \geq 200$ using a pediatric ARDS-specific cohort and a pediatric sepsis cohort with AHRF. Although the sample collection for these pediatric cohorts predated the COVID-19 pandemic, the Coronavirus pathogenesis pathway was prominent and reflects the systemic inflammatory changes that occur with severe ARDS. Novel pathways including ribosomal and

eukaryotic initiation of factor 2 (eIF2) regulation of protein translation and the nutrient, oxygen, and energy sensing pathway, mTOR, which is activated *via* PI3K/AKT signaling, emerged in the KEGG and IPA analysis. We also identified gene networks involved in distinguishing children with mild or moderate/severe pediatric ARDS from healthy controls. Finally, we identified gene networks involved in differentiating children with mild from moderate/severe pediatric ARDS using data from the pediatric ARDS-specific cohort. Our results support further investigation of gene networks and signaling pathways involved in differentiating pediatric ARDS severity and understanding pediatric ARDS heterogeneity by underlying biological processes.

An attempt at finding a generalizable diagnostic gene expression signature for ARDS using publicly available human whole blood gene expression arrays in adults and children failed to find an ARDS-specific signature after adjusting for clinical severity (6). Since this publication, a pediatric-specific ARDS transcriptomics dataset from whole blood was used to identify three ARDS sub-phenotypes with divergent clinical characteristics and outcomes using k-means clustering (9). A limitation of the transcriptomic pediatric ARDS phenotyping study was the lack of a non-ARDS mechanically ventilated cohort as a control group (9). Using the definition of a $P_aO_2/F_iO_2 < 200$ for moderate/severe AHRF from GSE66099 and ARDS from GSE147902, we were able to find an overlapping pattern of gene pathways of importance to pediatric ARDS using stability selection modeling and determine the discriminatory ability of the model to predict moderate/severe acute hypoxemia using the intersection of the top 1,500 ranked genes in the pediatric ARDS-specific and pediatric sepsis AHRF cohorts.

Metabolic pathways were the top KEGG pathway in each analysis. Several recent studies have investigated metabolic changes in ARDS (32). The mechanism by which metabolic derangements, such as mitochondrial dysfunction, decreased oxidative phosphorylation, and oxidative stress, lead to bioenergetic failure in ARDS and metabolic reprogramming of the immune system are active areas of investigation. Metabolomics studies of the plasma and airway fluid of patients with ARDS have shown that changes in tricarboxylic acid (TCA) cycle intermediates, such as glucose, alanine and glutamine, are markers of energetic stress on lung epithelial cells (32–38). Microarray analysis of whole-blood gene expression in adults with sepsis-triggered ARDS revealed that the “reactive” or hyperinflammatory subgroup is enriched for genes associated with oxidative phosphorylation, and that this subgroup is also associated with high plasma lactate levels indicative of mitochondrial dysfunction (39).

By contrast, the mitogen-activated protein kinase (MAPK) pathways that control cell proliferation, differentiation, motility, and survival were enriched in the comparatively “uninflamed” subgroup (39). The MAPK pathway is a first-level or second-level pathway found in all of the analyses distinguishing pediatric ARDS from healthy controls and moderate/severe from mild pediatric ARDS. It was recently shown that children with ARDS exhibit similar “reactive/ hyperinflamed” and “uninflamed” phenotypes as shown in adults with ARDS (40); however, this

phenotypic classification was not available at the time of accrual of the publicly available gene expression dataset used in this analysis. Therefore, we cannot speculate on the pathobiology underlying class differences (3, 4, 39, 41). The three Children's Hospital of Philadelphia ARDS transcriptomic subtypes (CATS) described by clustering GSE147902 do not conform to the previously described adult phenotypes and are not yet externally validated (9).

Respiratory viral infections are a common trigger of ARDS in children. Pathogenic influenza and coronaviruses use host epigenetic reprogramming to evade the host immune response (42, 43). In the current analysis, histone deacetylases (HDACs) were found to be a top pathway associated moderate/severe ARDS. HDACs regulate chromatin structure and thereby influence gene expression. For example, HDACs repress the production of proinflammatory cytokines in alveolar macrophages in chronic inflammatory lung conditions such as chronic obstructive pulmonary disease (44, 45). HDAC class I and chromatin modifying enzymes were identified as top pathways associated with ARDS mortality in a study using a multi-omics approach with multiple publicly available datasets (46). DNA methylation is another epigenetic modification that emerged in the Reactome analysis that distinguished moderate/severe from mild pediatric ARDS. Alterations in DNA methylation sites in ARDS are related to an imbalance in inflammation, immunity, endothelial and epithelial function, and coagulation (47).

Although children in these cohorts predated the COVID-19 pandemic, we found that gene networks in the KEGG database labeled coronavirus/COVID-19 (hsa05171) were a prominent pathway. This finding is similar to our previous study profiling differentially expressed genes from primary airway cells obtained from tracheal aspirate samples from intubated children with PARDS (8). The COVID-19 pathway includes proinflammatory cytokines such as interleukin (IL)-6, tumor necrosis factor (TNF)- α , nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B), IL-1 β , IL-8, and IL-12 signaling through the Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway. Viral pathogen recognition pathways such as Toll-like receptor signaling through MyD88, retinoic acid-inducible gene I (RIG-1)/melanoma differentiation-associated protein 5 pathways, and the antiviral type I interferon (IFN α/β) response and the complement cascade are also highlighted in the COVID-19 gene signaling network.

Machine learning-based analysis of data with class imbalances can be plagued by a number of issues, stemming from the inherent mathematical assumption of equal case to control distributions among many learning algorithms. As a result, these models, when applied in a single train-test instance, may produce random effects that are poorly generalized in external datasets. The traditional AUROC metric poorly characterizes class-based performances, thus necessitating alternative metrics to evaluate the robustness of models. One of the methods to investigate the effects of class imbalance is by the use of metrics such as AUPRC, F1 (harmonic mean of the precision and recall), and bootstrap with replacement. In this work, we use AUPRC and stability selection using bootstrap with replacement to identify a

coherent set of "stable" genes that indicate robust predictive performance consistently across 100 bootstrap iterations. We illustrate the AUPRC plot to show strong performance even in the minority class, thus emphasizing the robustness of the learned model.

Our study is limited by the *post hoc* analysis focused on finding a systemic transcriptomic signature of children with moderate/severe ARDS. Only one single-center study of pediatric ARDS exists without a non-hypoxemic respiratory failure control group. We attempted to validate our findings using a multi-center pediatric sepsis-triggered AHRF cohort using the same definitions as in the pediatric ARDS gene array cohort without the bilateral infiltrate radiographic findings and the more stringent requirement of two arterial blood gases at least four hours apart with a $P_aO_2/F_iO_2 < 200$. These cohorts capture gene expression differences early in the course of AHRF. The use of a single time point without serial sampling or clinical trajectory information precludes our assessment of temporal changes in gene expression and is a study limitation. As remarked on prior attempts to discover an ARDS transcriptomic signature, whole-blood derived gene expression was used that may be different from the airway-specific transcriptomic response. Concomitant changes in metabolites are not readily available for analysis in these cohorts. Validation of these findings as part of a prospective, multi-center, observational trial enrolling children with PALICC-defined PARDS will be necessary to generalize our results.

5. Conclusions

In summary, our analysis demonstrated that changes in gene expression patterns corresponding to metabolic pathways involved in energy balance, fundamental cellular processes such as protein translation, mitochondrial function, oxidative stress, immune signaling, and inflammation are differentially regulated early in pediatric ARDS and sepsis-induced AHRF compared to both healthy controls and to milder, acute hypoxemia. Our findings support the hypothesis that differential regulation of metabolic pathway genes involved in cellular energetics and metabolic pathways are important mechanisms to consider to further our understanding of the heterogeneity and underlying pathobiology of moderate and severe pediatric acute respiratory distress syndrome. Future work should validate the findings from this secondary analysis of publicly available gene expression data using plasma biomarkers in a prospective observational study seeking to associate underlying pathobiology with outcomes in children with ARDS.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) at the Children's Hospital of Philadelphia (CHOP IRB 13-010578, "Biomarkers of pediatric ARDS", Approved 7/2/2014; GSE147902) and by the IRBs of each of the 18 participating institutions for the sepsis AHRF cohort [Cincinnati Children's Hospital Medical Center parent protocol approved by Emory IRB 00079159, "Genomic Analysis of Pediatric Systemic Inflammatory Response Syndrome (SIRS)", Approved 02/21/2015; GSE66099]. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

NY and HRW obtained the patient samples, performed the transcriptomics experiments. JG, NY, and RK conceived and developed the study, supervised the analysis and interpreted the data. MR and RK performed the informatics analysis. JG drafted and edited the manuscript. RK, MR, MJR, and NY assisted with drafting and editing the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Funding was provided by NIH grants K23 HL151897 JRG and K23-HL136688 to NY. RK was supported by the National Institutes of Health under Award Numbers R01GM139967 and UL1TR002378. HRW was supported by NIH R35GM126943.

References

- Bos LD, Schouten LR, van Vught LA, Wiewel MA, Ong DSY, Cremer O, et al. Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis. *Thorax*. (2017) 72(10):876–83. doi: 10.1136/thoraxjnl-2016-209719
- Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, et al. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med*. (2014) 2(8):611–20. doi: 10.1016/S2213-2600(14)70097-9
- Sinha P, Churpek MM, Calfee CS. Machine learning classifier models can identify acute respiratory distress syndrome phenotypes using readily available clinical data. *Am J Respir Crit Care Med*. (2020) 202(7):996–1004. doi: 10.1164/rccm.202002-0347OC
- Sinha P, Delucchi KL, Thompson BT, McAuley DF, Matthay MA, Calfee CS, et al. Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. *Intensive Care Med*. (2018) 44(11):1859–69. doi: 10.1007/s00134-018-5378-3
- Banerjee S, Mohammed A, Wong HR, Palaniyar N, Kamaleswaran R. Machine learning identifies complicated sepsis course and subsequent mortality based on 20 genes in peripheral blood immune cells at 24 H post-ICU admission. *Front Immunol*. (2021) 12:592303. doi: 10.3389/fimmu.2021.592303
- Sweeney TE, Thomas NJ, Howrylak JA, Wong HR, Rogers AJ, Khatri P. Multicohort analysis of whole-blood gene expression data does not form a robust diagnostic for acute respiratory distress syndrome. *Crit Care Med*. (2018) 46(2):244–51. doi: 10.1097/CCM.0000000000002839
- Sauthier MS, Jouvett PA, Newhams MM, Randolph AG. Machine learning predicts prolonged acute hypoxic respiratory failure in pediatric severe influenza. *Crit Care Explor*. (2020) 2(8):e0175. doi: 10.1097/CCE.0000000000000175
- Grunwell JR, Rad MG, Stephenson ST, Mohammad AF, Opolka C, Fitzpatrick AM, et al. Machine learning-based discovery of a gene expression signature in pediatric acute respiratory distress syndrome. *Crit Care Explor*. (2021) 3(6):e0431. doi: 10.1097/CCE.0000000000000431
- Yehya N, Varisco BM, Thomas NJ, Wong HR, Christie JD, Feng R. Peripheral blood transcriptomic sub-phenotypes of pediatric acute respiratory distress syndrome. *Crit Care*. (2020) 24(1):681. doi: 10.1186/s13054-020-03410-7
- Yehya N, Keim G, Thomas NJ. Subtypes of pediatric acute respiratory distress syndrome have different predictors of mortality. *Intensive Care Med*. (2018) 44(8):1230–9. doi: 10.1007/s00134-018-5286-6
- Yehya N, Thomas NJ, Margulies SS. Circulating nucleosomes are associated with mortality in pediatric acute respiratory distress syndrome. *Am J Physiol Lung Cell Mol Physiol*. (2016) 310(11):L1177–84. doi: 10.1152/ajplung.00067.2016
- Yehya N, Thomas NJ, Meyer NJ, Christie JD, Berg RA, Margulies SS. Circulating markers of endothelial and alveolar epithelial dysfunction are associated with mortality in pediatric acute respiratory distress syndrome. *Intensive Care Med*. (2016) 42(7):1137–45. doi: 10.1007/s00134-016-4352-1
- Wong HR, Salisbury S, Xiao Q, Cvijanovich NZ, Hall M, Allen GL, et al. The pediatric sepsis biomarker risk model. *Crit Care*. (2012) 16(5):R174. doi: 10.1186/cc11652
- Wong HR, Weiss SL, Giuliano JS, Wainwright MS, Cvijanovich NZ, Thomas NJ, et al. Testing the prognostic accuracy of the updated pediatric sepsis biomarker risk model. *PLoS One*. (2014) 9(1):e86242. doi: 10.1371/journal.pone.0086242
- Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust

Acknowledgments

JRG, NY, and RK are grateful for the mentorship, career guidance, and collaboration generously provided by the late HRW.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fped.2023.1159473/full#supplementary-material>.

- diagnostic gene set. *Sci Transl Med.* (2015) 7(287):287ra71. doi: 10.1126/scitranslmed.aaa5993
16. Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, et al. Acute respiratory distress syndrome: the Berlin definition. *JAMA.* (2012) 307(23):2526–33. doi: 10.1001/jama.2012.5669
 17. Goldstein B, Giroir B, Randolph A, International Consensus Conference on Pediatric S. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med.* (2005) 6(1):2–8. doi: 10.1097/01.PCC.0000149131.72248.E6
 18. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy-analysis of affymetrix GeneChip data at the probe level. *Bioinformatics.* (2004) 20(3):307–15. doi: 10.1093/bioinformatics/btg405
 19. Wu J, Irizarry R, MacDonald J, Gentry J. Gcrma: background adjustment using sequence information. Version. 2200:3–10. ed2012. p. R package.
 20. Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE, Zhang Y, et al. sva: Surrogate Variable Analysis. 3.40.0 ed2021. p. R package.
 21. Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Series B Stat Methodol.* (2010) 72(4):417–73. doi: 10.1111/j.1467-9868.2010.00740.x
 22. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw.* (2010) 36:1–13. doi: 10.18637/jss.v036.i11
 23. Pedregosa F, Varoquaux G, Michel V, Thirion B. Scikit-learn: machine learning in python. *J Mach Learn Res.* (2011) 12:2825–30. doi: 10.48550/arXiv.1201.0490
 24. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* (2021) 49(D1):D545–51. doi: 10.1093/nar/gkaa970
 25. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27
 26. Palombo V, Milanese M, Sferra G, Capomaccio S, Sgorlon S, D'Andrea M. PANEV: an R package for a pathway-based network visualization. *BMC Bioinformatics.* (2020) 21(1):46. doi: 10.1186/s12859-020-3371-7
 27. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* (2018) 46(D1):D649–55. doi: 10.1093/nar/gkx1132
 28. Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, et al. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics.* (2017) 18(1):142. doi: 10.1186/s12859-017-1559-2
 29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303
 30. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
 31. Shah RD, Samworth RJ. Variable selection with error control: another look at stability selection. *J R Stat Soc Series B Stat Methodol.* (2013) 75(1):55–80. doi: 10.1111/j.1467-9868.2011.01034.x
 32. Robinson MJ, Krasnodembkaya AD. Therapeutic targeting of metabolic alterations in acute respiratory distress syndrome. *Eur Respir Rev.* (2020) 29(156):200114. doi: 10.1183/16000617.0114-2020
 33. Izquierdo-Garcia JL, Nin N, Cardinal-Fernandez P, Ruiz-Cabello J, Lorente JA. Metabolomic profile of acute respiratory distress syndrome of different etiologies. *Intensive Care Med.* (2019) 45(9):1318–20. doi: 10.1007/s00134-019-05634-w
 34. Izquierdo-Garcia JL, Nin N, Jimenez-Clemente J, Horcajada JP, Arenas-Miras MDM, Gea J, et al. Metabolomic profile of ARDS by nuclear magnetic resonance spectroscopy in patients with H1N1 influenza virus pneumonia. *Shock.* (2018) 50(5):504–10. doi: 10.1097/SHK.0000000000001099
 35. Rogers AJ, Matthey MA. Applying metabolomics to uncover novel biology in ARDS. *Am J Physiol Lung Cell Mol Physiol.* (2014) 306(11):L957–61. doi: 10.1152/ajplung.00376.2013
 36. Viswan A, Ghosh P, Gupta D, Azim A, Sinha N. Distinct metabolic endotype mirroring acute respiratory distress syndrome (ARDS) subphenotype and its heterogeneous biology. *Sci Rep.* (2019) 9(1):2108. doi: 10.1038/s41598-019-39017-4
 37. Viswan A, Singh C, Kayastha AM, Azim A, Sinha N. An NMR based panorama of the heterogeneous biology of acute respiratory distress syndrome (ARDS) from the standpoint of metabolic biomarkers. *NMR Biomed.* (2020) 33(2):e4192. doi: 10.1002/nbm.4192
 38. Bos LDJ. Diagnosis of acute respiratory distress syndrome by exhaled breath analysis. *Ann Transl Med.* (2018) 6(2):33. doi: 10.21037/atm.2018.01.17
 39. Bos LD, Scicluna BP, Ong DY, Cremer O, van der Poll T, Schultz MJ, et al. Understanding heterogeneity in paediatric phenotypes of ARDS by leukocyte expression profiles. *Am J Respir Crit Care Med.* (2019) 200(1):42–50. doi: 10.1164/rccm.201809-1808OC
 40. Dahmer MK, Yang G, Zhang M, Quasney MW, Sapru A, Weeks HM, et al. Identification of phenotypes in paediatric patients with acute respiratory distress syndrome: a latent class analysis. *Lancet Respir Med.* (2022) 10(3):289–97. doi: 10.1016/S2213-2600(21)00382-9
 41. Sinha P, Delucchi KL, McAuley DF, O'Kane CM, Matthey MA, Calfee CS. Development and validation of parsimonious algorithms to classify acute respiratory distress syndrome phenotypes: a secondary analysis of randomised controlled trials. *Lancet Respir Med.* (2020) 8(3):247–57. doi: 10.1016/S2213-2600(19)30369-8
 42. Menachery VD, Eisfeld AJ, Schafer A, Josset L, Sims AC, Proll S, et al. Pathogenic influenza viruses and coronaviruses utilize similar and contrasting approaches to control interferon-stimulated gene responses. *mBio.* (2014) 5(3):e01174–14. doi: 10.1128/mBio.01174-14
 43. Menachery VD, Schafer A, Burnum-Johnson KE, Mitchell HD, Eisfeld AJ, Walters KB, et al. MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. *Proc Natl Acad Sci USA.* (2018) 115(5):E1012–21. doi: 10.1073/pnas.1706928115
 44. Barnes PJ, Adcock IM, Ito K. Histone acetylation and deacetylation: importance in inflammatory lung diseases. *Eur Respir J.* (2005) 25(3):552–63. doi: 10.1183/09031936.05.00117504
 45. Ito K, Ito M, Elliott WM, Cosio B, Caramori G, Kon OM, et al. Decreased histone deacetylase activity in chronic obstructive pulmonary disease. *N Engl J Med.* (2005) 352(19):1967–76. doi: 10.1056/NEJMoa041892
 46. Liao SY, Casanova NG, Bime C, Camp SM, Lynn H, Garcia JGN. Identification of early and intermediate biomarkers for ARDS mortality by multi-omic approaches. *Sci Rep.* (2021) 11(1):18874. doi: 10.1038/s41598-021-98053-1
 47. Zhang S, Wu Z, Xie J, Yang Y, Wang L, Qiu H. DNA Methylation exploration for ARDS: a multi-omics and multi-microarray interrelated analysis. *J Transl Med.* (2019) 17(1):345. doi: 10.1186/s12967-019-2090-1