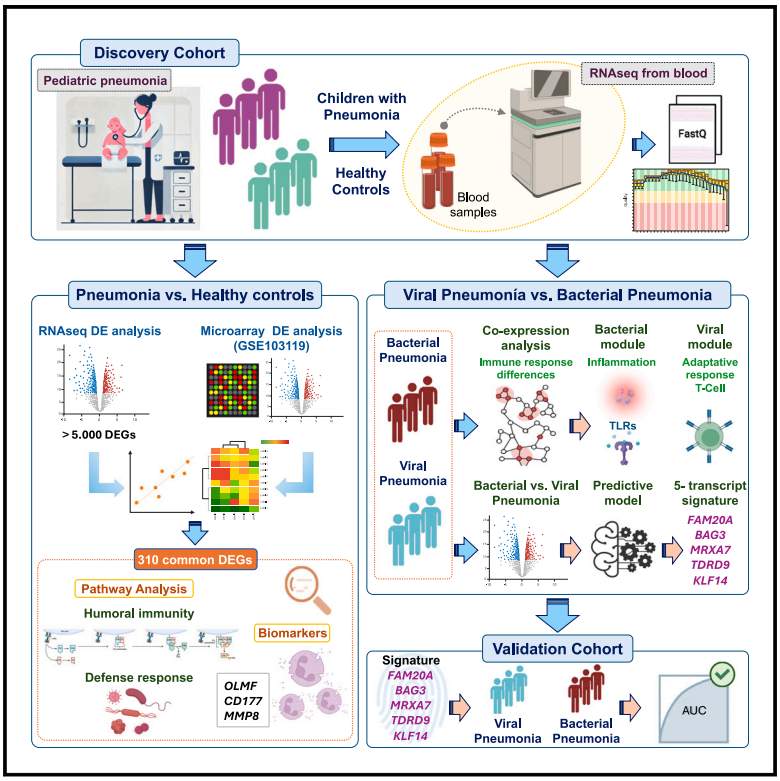
iScience

Article

A 5-transcript signature for discriminating viral and bacterial etiology in pediatric pneumonia

Graphical abstract



Highlights

- Over 5,000 genes are differentially expressed in pneumonia vs. healthy controls
- Pneumonia DEPs: neutrophil degranulation, humoral response, and inflammation
- A 5-transcript host signature differentiates bacterial from viral pneumonia
- Enhances diagnosis, reducing unnecessary tests and treatments

Authors

Sandra Viz-Lasheras, Alberto Gómez-Carballa, Jacobo Pardo-Seco, ..., DIAMONDS, GENDRES and, PERFORM consortia

Correspondence

antonio.salas@usc.es

In brief

Pediatrics; Diagnostics; Body substance sample; Clinical microbiology; Transcriptomics

Viz-Lasheras et al., 2025, iScience *28*, 111747
February 21, 2025
https://doi.org/10.1016/j.isci.2025.111747

CellPress

iScience

CellPress
OPEN ACCESS

Article

A 5-transcript signature for discriminating viral and bacterial etiology in pediatric pneumonia

Sandra Viz-Lasheras,[1,2,3,37] Alberto Gómez-Carballa,[1,2,3,37] Jacobo Pardo-Seco,[1,2,3] Xabier Bello,[1,2,3] Irene Rivero-Calle,[2,3,4] Ana Isabel Dacosta,[2,3,4] Myrsini Kaforou,[5] Dominic Habgood-Coote,[5] Aubrey J. Cunnington,[5] Marieke Emonts,[6,7,8] Jethro A. Herberg,[5] Victoria J. Wright,[5] Enitan D. Carrol,[9,10] Stephane C. Paulus,[11] Werner Zenz,[12] Daniela S. Kohlfürst,[12] Michiel Van der Flier,[13,14] Ronald de Groot,[14] Luregn J. Schlapbach,[15] Philipp Agyeman,[16] Andrew J. Pollard,[11] Colin Fink,[17] Taco T. Kuijpers,[18] Suzanne Anderson,[19] Cristina Calvo,[20,21,22,23,24]

(Author list continued on next page)

SUMMARY

Pneumonia stands as the primary cause of death among children under five, yet current diagnosis methods often result in inadequate or unnecessary treatments. Our research seeks to address this gap by identifying host transcriptomic biomarkers in the blood of children with definitive viral and bacterial pneumonia. We performed RNA sequencing on 192 prospectively collected whole blood samples, including 38 controls and 154 pneumonia cases, uncovering a 5-transcript signature (genes *FAM20A*, *BAG3*, *TDRD9*, *MXRA7*, and *KLF14*) that effectively distinguishes bacterial from viral pneumonia (area under the curve (AUC): 0.95 [0.88–1.00]). Initial validation using combined definitive and probable cases yielded an AUC of 0.87 [0.77–0.97], while full validation in a new prospective cohort of 32 patients achieved an AUC of 0.92 [0.83–1.00]. This robust signature holds significant potential to enhance diagnostics accuracy for pediatric pneumonia, reducing diagnostic delays and unnecessary treatments and potentially transforming clinical practice.

INTRODUCTION

Pneumonia is an acute respiratory infection and one of the leading causes of morbidity and mortality worldwide. In children younger than 5 (excluding the neonatal period), it remains the primary cause of death.[1,2] Common symptoms in children include fever, tachypnea, cough, chest pain, and difficulty in breathing.[3] Pneumonia can be caused by a variety of microorganisms, including bacteria, viruses, and fungi, depending on age groups and the specific epidemiological situation. However, *Streptococcus pneumoniae* is the most common bacterial cause of pneumonia across all age group.[2,4] The causative microorganisms also vary significantly between the two main types of pneumonia: community-acquired pneumonia (CAP) and hospital-acquired pneumonia (HAP). *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Enterobacter* are the most common causes of HAP[5] whereas *Streptococcus pneumoniae*; *Staphylococcus aureus*; *Streptococcus pyogenes*; *Haemophilus influenzae*; *Mycoplasma pneumoniae*; *Legionella pneumophila*; and respiratory viruses, such as respiratory syncytial virus, rhinovirus, or influenza virus A/B, are the most common microorganisms responsible for CAP.[3,6] The introduction of improved conjugate vaccines against *Haemophilus influenzae* type b and *Streptococcus pneumoniae* has led to a reduction in the incidence and severity of childhood pneumonia, resulting in significant changes in the proportions of CAP cases etiologies.[7,8]

The clinical presentation of pneumonia is diverse and may overlap with other respiratory conditions such as asthma, bronchiolitis, pertussis, lung abscess, bronchiectasis, or malaria, among others.[6,9–14] Accurate assessment of disease severity[15] is crucial for effective management, influencing decisions on antibiotic prescriptions and hospitalization.[16,17] Despite efforts

[1]Unidade de Xenética, Instituto de Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, and Genética de Poblaciones en Biomedicina (GenPoB) Research Group, Instituto de Investigación Sanitaria (IDIS), 15706 Hospital Clínico Universitario de Santiago (SERGAS), Galicia, Spain
[2]Genetics, Vaccines and Infections Research Group (GenViP), Instituto de Investigación Sanitaria de Santiago, 15706 Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain
[3]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBER-ES), Madrid, Spain
[4]Translational Pediatrics and Infectious Diseases, Department of Pediatrics, 15706 Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Galicia, Spain
[5]Department of Infectious Disease, Imperial College London, London W2 1PG, UK
[6]Great North Children's Hospital, Paediatric Immunology, Infectious Diseases & Allergy, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne NE1 4LP, UK

(Affiliations continued on next page)

María del Carmen Martínez-Padilla,[25] Ana Pérez-Aragón,[26] Esteban Gómez-Sánchez,[27] Juan Valencia-Ramos,[27] Francisco Giménez-Sánchez,[28] Paula Alonso-Quintela,[29] Laura Moreno-Galarraga,[30,31] Ulrich von Both,[32] Marko Pokorn,[33] Dace Zavadska,[34] María Tsolia,[35] Clementien L. Vermont,[36] Henriëtte A. Moll,[36] Michael Levin,[5] Federico Martinón-Torres,[2,3,4] and Antonio Salas[1,2,3,38,*] on behalf of EUCLIDS, DIAMONDS, GENDRES and, PERFORM consortia

[7]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne NE2 4HH, UK
[8]NIHR Newcastle Biomedical Research Centre based at Newcastle upon Tyne Hospitals NHS Trust and Newcastle University, Newcastle upon Tyne NE4 5PL, UK
[9]Department of Infectious Diseases, Alder Hey Children's NHS Foundation Trust, Liverpool L12 2AP, UK
[10]Department of Clinical Infection, Microbiology and Immunology, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool L69 7BE, UK
[11]Department of Paediatrics, University of Oxford and the NIHR Oxford Biomedical Research Centre, Oxford OX3 9DU, UK
[12]Department of General Paediatrics, Medical University of Graz, Graz, Auenbruggerplatz 34/2 8036, Graz, Austria
[13]Pediatric Infectious Diseases and Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht 3508 AB, the Netherlands
[14]Pediatric Infectious Diseases and Immunology, Amalia Children's Hospital, and Section Pediatric Infectious Diseases, Laboratory of Medical Immunology, Department of Laboratory Medicine, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen 6500 HB, the Netherlands
[15]Department of Intensive Care and Neonatology, and Children's Research Center, University Children's Hospital Zürich, University of Zürich, Zürich, Switzerland
[16]Department of Pediatrics, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
[17]Micropathology Ltd, University of Warwick, Warwick CV4 7EZ, UK
[18]Division of Pediatric Immunology, Rheumatology and Infectious diseases, Emma Children's Hospital, Amsterdam Univiersyt Medical Center (Amsterdam UMC), Amsterdam 1105 AZ, the Netherlands
[19]Medical Research Council Unit at the London School of Hygiene & Tropical Medicine, Banjul, The Gambia
[20]General Pediatrics, Infectious and Tropical Diseases Department, Hospital La Paz, 28046 Madrid, Spain
[21]La Paz Research Institute (IdiPAZ), 28029 Madrid, Spain
[22]Faculty of Medicine, Universidad Autónoma de Madrid (UAM), 28049 Madrid, Spain
[23]Centro de Investigación Biomédica en Red en Enfermedades Infecciosas (CIBERINFEC), Madrid, Spain
[24]Red de Investigación Traslacional en Infectología Pediátrica (RITIP), Madrid, Spain
[25]Unidad de Cuidados Intensivos Pediátricos, Complejo Hospitalario de Jaen, Jaen, Spain
[26]Hospital Universitario Virgen de las Nieves, Servicio de Pediatría, Granada, Spain
[27]Department of Pediatric Intensive Care Unit, Hospital Universitario de Burgos, Burgos, Spain
[28]Instituto Hispalense de Pediatría, Instituto Balmis de Vacunas, Almería, Spain
[29]Neonatal Intensive Care Unit, Complejo Asistencial Universitario de León, León, Spain
[30]Department of Pediatrics, Complejo Hospitalario de Navarra, Servicio Navarro de Salud, Pamplona, Spain
[31]IdiSNA (Instituto de Investigación Sanitaria de Navarra), Navarra Institute for Health Research, Pamplona, Spain
[32]Infectious Diseases, Department of Pediatrics, Dr von Hauner Children's Hospital, University Hospital, LMU Munich, Munich, Germany
[33]Division of Paediatrics, University Medical Centre Ljubljana and Medical Faculty, University of Ljubljana, Ljubljana, Slovenia
[34]Children's Clinical University Hospital, Rīga Stradins University, Rīga, Latvia
[35]Second Department of Paediatrics, National and Kapodistrian University of Athens (NKUA), School of Medicine, Panagiotis & Aglaia, Kyriakou Children's Hospital, Athens, Greece
[36]Department of Pediatrics, Erasmus MC, Rotterdam, the Netherlands
[37]These authors contributed equally
[38]Lead contact
*Correspondence: antonio.salas@usc.es
https://doi.org/10.1016/j.isci.2025.111747

to enhance international guidelines,[9,10,18] the variability in causative microorganisms and symptoms poses a significant challenge for clinicians in healthcare centers.[14,19–22] Current diagnostic criteria rely on nonspecific symptoms, chest imaging, and inconclusive laboratory analysis,[14,19,23,24] with limitations in differentiating the etiology of the disease (i.e., viral and bacterial). Obtaining lower respiratory tract samples in children is often difficult.[19,20] Consequently, misdiagnosis of bacterial pneumonia is common, leading to increased costs, unnecessary medical tests, hospital admissions, and incorrect antibiotic prescriptions.[6,25]

Given these challenges, numerous studies have suggested the potential use of biomarkers as supplementary tools in the managing and diagnosis of pneumonia.[9,16,25–27] Serum or plasma C-reactive protein (CRP) and procalcitonin (PCT) have been identified as potential biomarkers to assist in the diagnosis and prognosis of pneumonia patients.[9,16,28,29] The study of host transcriptomic in the field of infectious disease has become increasingly important in recent years, leading to advancements in understanding host-pathogen interactions and the development of useful tools for diagnosing and prognosing diseases, including the development of point-of-care devices.[30–35] There are a growing number of studies investigating host gene expression biomarkers related to different infectious diseases, such as tuberculosis,[36–38] H1N1,[39] Respiratory Syncitial Virus (RSV),[40] and rotavirus,[41] and also aiming at differentiating bacterial from viral infection.[30,35,42,43] However, the study of host transcriptomics in pneumonia patients has not received as much attention, and only a few studies

describing blood transcriptomic biomarkers related to CAP in adults are available in the literature.[27,44] Some studies have recently focused on pneumonia outcome prediction[45–47] or pneumonia etiology discrimination.[48,49]

Regarding pediatric pneumonia, most studies have focused on the potential use of already described biomarkers, such as white blood cell (WBC) count, neutrophil percentage (NP), serum CRP, and PCT. However, results indicate that these biomarkers are not good predictors for pediatric pneumonia due to their inability to differentiate pneumonia from bacterial and viral origins.[50,51] To our knowledge, there is only one study investigating blood transcriptomics in the context of pediatric pneumonia etiology, but in a very specific population from a malaria-endemic area and using a signature with a high number of genes.[52] Therefore, further studies are necessary to approach this complex disease in children.

In light of these challenges in pneumonia diagnosis and management in children, we analyze the whole blood transcriptome of the largest pediatric pneumonia cohort recruited to date with the aim of (1) shedding light on the molecular mechanisms underlying pediatric pneumonia, providing insights on new therapeutic targets; (2) investigating host's differential response to pneumonia of bacterial and viral origin; and (3) proposing a minimal transcriptomic signature that allows for the differentiation between viral and bacterial pneumonia in children.

RESULTS

Discovery cohort description

In this study, blood samples were collected from 192 children (see Figure 1 and STAR Methods for details on the experimental design). Among them, 154 were hospitalized with pneumonia (median age: 3.4 years; 51.3% male) and 38 were healthy controls (median age: 3.3 years; 68.4% male). Clinical and demographic details of the entire cohort and sub-categories used in different analyses are provided in Table 1. Among the pneumonia patients, 25% were admitted to Pediatric Intensive Care Unit (PICU), 3% died, and 39% required oxygen support (16.2% with invasive ventilation). No statistically significant differences were found between the bacterial and viral pneumonia sub-groups for these parameters. The majority of bacterial pneumonias were caused by *Streptococcus pneumoniae* (47.5%) followed by *Staphylococcus aureus* (10.0%) and *Streptococcus pyogenes* (7.5%) among others (35.0%). Viral pneumonias were caused by influenza virus (22.2%), bocavirus (22.2%), RSV (22.2%), rhinovirus (11.1%), adenovirus (11.1%), and parainfluenza (11.1%).

As expected, CRP values were significantly higher in definitive bacterial (DB) + probable bacterial (PB) compared to definitive viral (DV) + probable viral (PV) (p value = 0.001). CRP values were higher in the DB group compared to DV; this difference approached the significance threshold (p value = 0.062), likely due to limited statistical power, especially given the low number of samples in the DV group, coupled with missing data. For the remaining blood tests, there were no significant statistical differences between bacterial and viral sub-groups.

Differentially expressed genes in pneumonia

Through a comparative analysis of transcriptomes in children with pneumonia vs. healthy controls, we identified 5,486 differentially expressed genes (DEGs), using a significance threshold of Benjamini-Hochberg false discovery rate (FDR) 5% (see Table S1). Within this set of DEGs, 2,716 were found to be upregulated, while 2,770 were downregulated.

To gain insight into the overall variation in gene expression, we performed a principal component analysis (PCA) on a subset of 100 highly variable genes out of a total of 192 samples. The first principal component (PC1), explaining approximately 30% of the variation, effectively segregated the samples into two main clusters, corresponding to pneumonia samples and healthy control samples (Figure 2A).

To prioritize significant changes in gene expression, we applied a selective criterion, including an adjusted p value < 0.05 and a $\log_2$FC > |2|. This selection process yielded 272 DEGs (Figure 2B), with 208 exhibiting upregulation and 64 genes showing downregulation, indicating a prevailing pattern of overexpression among these genes.

Next, we compared the DEGs identified in our RNA sequencing (RNA-seq) cohort, with those inferred from the analysis of the Wallihan et al. study[53] (accession number GSE103119). The dataset by Wallihan et al. included transcriptomic data from blood samples of pediatric pneumonia cases (n = 152) and healthy controls (n = 20). Our comparison revealed 1,729 common DEGs (Table S2). Focusing on the DEGs with significant expression changes compared to healthy controls (adjusted p value < 0.05 and $\log_2$FC > |1|; n = 310 genes), we found a strong positive correlation (ρ = 0.81; p value < $2.2e^{-16}$) between $\log_2$FC values obtained from both datasets (Table S2; Figure 2C). Furthermore, the expression pattern of these 310 shared genes clearly segregated pneumonia cases from healthy controls, forming two distinct clusters in both our RNA-seq discovery cohort and Wallihan's microarray dataset (Figure 2D).

Functional enrichment analysis of DEGs

To further characterize DEGs from a functional point of view, we conducted an over-representation analysis (ORA) and a gene set enrichment analysis (GSEA) based on Gene Ontology (GO) and Reactome using only validated DEGs with a positive correlation between the discovery and the validation datasets. We set a threshold of an adjusted p value < 0.05 for significance and a $|\log_2FC|$ > 1.5. The analysis revealed that the most significant pathways (adjusted p value < 0.01) are associated with essential immune processes, primarily related to the innate response to severe CAP (Table S3; Figure 2E).

The Reactome-ORA analysis indicated a robust association between the selected DEGs, including *matrix metalloproteinase-8 (MMP8)*, *OLFM4*, or *CD177*, and an innate response characterized by "neutrophil degranulation." This pathway displayed the highest significance (adjusted p value = $6.5e^{-12}$; Table S3). Consistently, GO-ORA results identified differences in neutrophil-related immune pathways, including "leukocyte mediated immunity" (adjusted p value = $7.4e^{-5}$), "myeloid leukocyte mediated immunity" (adjusted p value = 0.001), "neutrophil mediated immunity" (adjusted p value = $5.1e^{-5}$), "myeloid leukocyte activation" (adjusted p value = $1.4e^{-5}$), "granulocyte activation" (adjusted p value = 0.002), and "neutrophil activation" (adjusted p value = 0.001) (Table S3; Figure 2E).

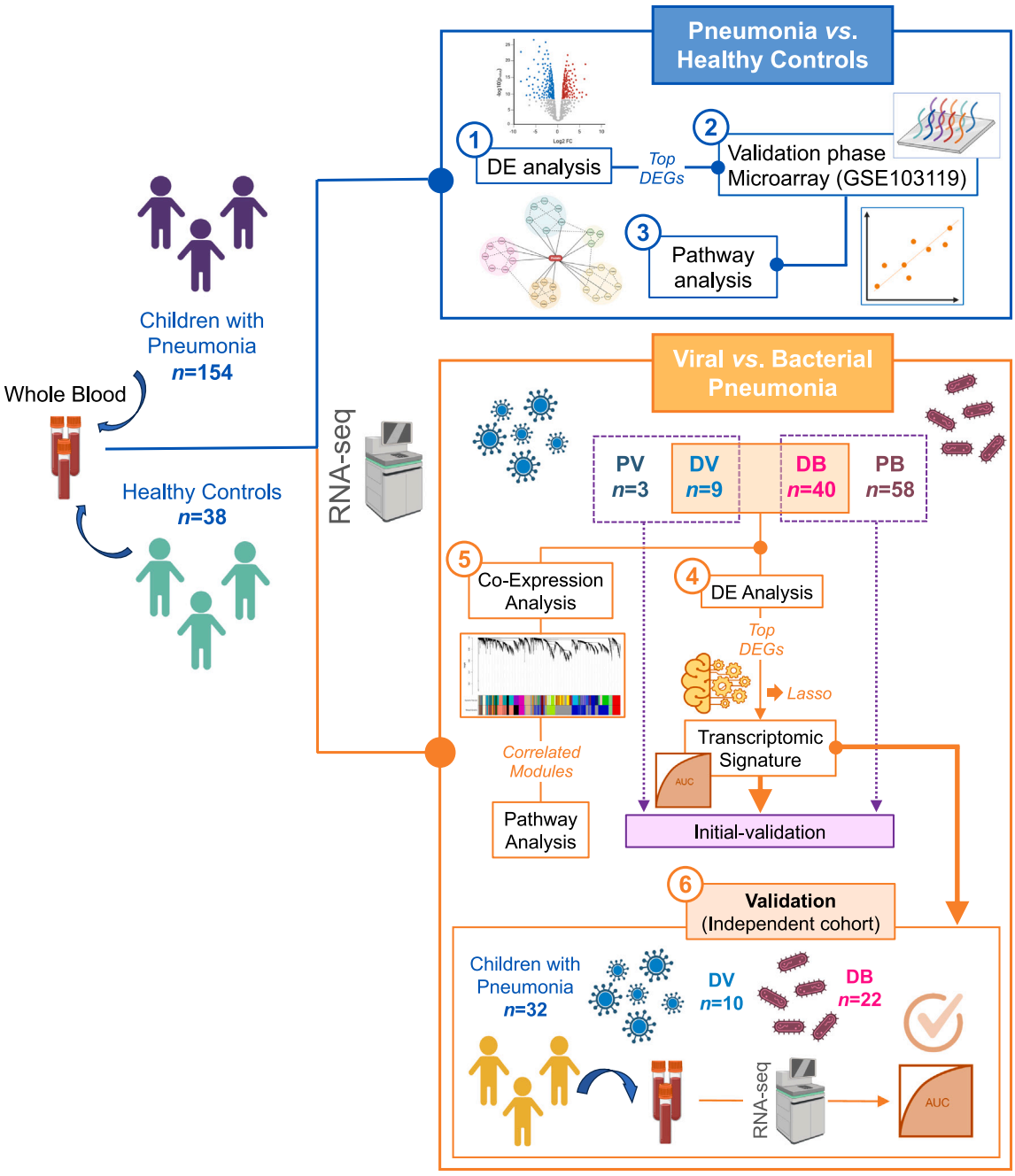


Figure 1. Scheme of the study design

The figure was built using BioRender resources; created in BioRender. BioRender.com/b32x666.

Furthermore, there is a significant overlap in enrichment results from GO and Reactome, particularly regarding humoral immunity processes. Notably, GO-ORA identified "humoral immune response" (adjusted p value = $4.4e^{-5}$), "antimicrobial humoral response" (adjusted p value = $1.4e^{-5}$), and "antimicrobial humoral immune response mediated by antimicrobial peptide" (adjusted p value = 0.0003) (Table S3; Figure 2E) as highly significant pathways. Likewise, Reactome-ORA highlighted the involvement of various humoral responses in severe CAP, including the complement system ("regulation of complement cascade" [adjusted p value = 0.001] and "complement cascade" [adjusted p value = 0.003]), as well as humoral immunity mediated by "antimicrobial peptides" (adjusted p value = $1.2e^{-9}$) (Table S3; Figure 2E).

Most significant pathways from GO-ORA were those implicated in "defense response to bacterium" (adjusted p value = $2.3e^{-8}$),

Table 1. Clinical features of the pneumonia patient's cohort

	Pneumonia vs. healthy controls			DV vs. DB					DV + PV vs. DB + PB				
	Controls (*n* = 38)	Pneumonia (*n* = 154)	*n*	DB (*n* = 40)	*n*	DV (*n* = 9)	*n*	*p* value	DB + PB (*n* = 98)	*n*	DV + PV (*n* = 12)	*n*	*p* value
Gender (male)	68.4% (36/38)	51.3% (79/154)	154	62.5% (25/40)	40	55.6% (5/9)	9	0.72	53.1% (52/98)	98	58.3% (7/12)	12	0.769
Age (years)	3.3 (2.3–8.2)	3.36 (1.24–6.17)	153	2.77 (1.25–5.78)	40	2.75 (1.52–4.76)	8	0.923	3.78 (1.54–6.285)	98	1.93 (0.94–3.88)	11	0.096
Days hospitalized	–	5 (3–10)	149	6.5 (3.5–14)	38	6 (4.75–8.5)	8	0.621	5 (3–9)	95	6 (3.5–9)	11	0.536
PICU admission	–	24.7% (38/154)	154	30% (12/40)	40	22.2% (2/9)	9	1	22.4% (22/98)	98	25% (3/12)	12	1
PICU days	–	7 (3–11)	37	11 (6.5–16)	11	6.5 (4.25–8.75)	2	0.372	6 (3–11)	21	2 (2–6.5)	3	0.356
Death	–	3.2% (5/154)	154	7.5% (3/40)	40	11.1% (1/9)	9	0.569	4.1% (4/98)	98	8.3% (1/12)	12	0.445
Treatments													
Inotropes	–	9.7% (15/154)	154	5% (2/40)	40	11.1% (1/9)	9	0.464	6.1% (6/98)	98	8.3% (1/12)	12	0.565
IV	–	16.2% (25/154)	154	15% (6/40)	40	22.2% (2/9)	9	0.628	11.2% (11/98)	98	25% (3/12)	12	0.18
NIV	–	11% (17/154)	154	15% (6/40)	40	11.1% (1/9)	9	1	8.2% (8/98)	98	8.3% (1/12)	12	1
Oxygen	–	39% (60/154)	154	40% (16/40)	40	33.3% (3/9)	9	1	34.7% (34/98)	98	41.7% (5/12)	12	0.751
Analytical parameters													
CRP (mg/L)	–	81 (30.1–215.5)	59	146 (34.5–244)	11	28.5 (11.85–30.15)	6	0.062	190 (81–273)	33	27 (5.9–30.1)	7	0.001*
Platelets (10^9/L)	–	322 (259–433)	135	287.5 (175.5–405.5)	28	491 (283–507)	5	0.173	324 (262.75–421.5)	84	398 (273.5–495)	8	0.458
White cells (10^9/L)	–	17.2 (10.2–25)	133	18.75 (10.075–26.075)	28	7.2 (6.35–16.25)	3	0.423	18.65 (12.18–26.72)	84	6.85 (5.75–20.775)	6	0.089
Lactate (mmol/L)	–	1.9 (1.18–2.65)	30	2.5 (2.3–2.7)	5	2.5 (2.5–2.5)	1	1	2.2 (1.48–2.55)	16	2.5 (2.5–2.5)	1	0.539
Neutrophils (10^9/L)	–	11.87 (6.47–19.95)	128	13.4 (8.05–20.88)	26	12.06 (7.94–16.18)	2	0.762	14.61 (7.82–23.62)	82	4.16 (3.82–18.6)	5	0.111
Monocytes (10^9/L)	–	0.8 (0.17–1.7)	84	0.61 (0.125–1.2)	14	0.8 (0.4–1.15)	3	0.801	0.955 (0.168–1.92)	56	0.02 (0–0.8)	5	0.068
Lymphocytes (10^9/L)	–	3.17 (1.67–5.4)	130	1.95 (1.4–5.65)	27	4.1 (3.42–8.55)	3	0.213	2.6 (1.5–5.4)	83	3.42 (2.1475–6.35)	6	0.381

Wilcoxon and Fisher exact test were used to assess statistical significance between groups in numerical and categorical variables, respectively. IV, invasive ventilation; NIV, non-invasive ventilation. * indicates statistical significance.

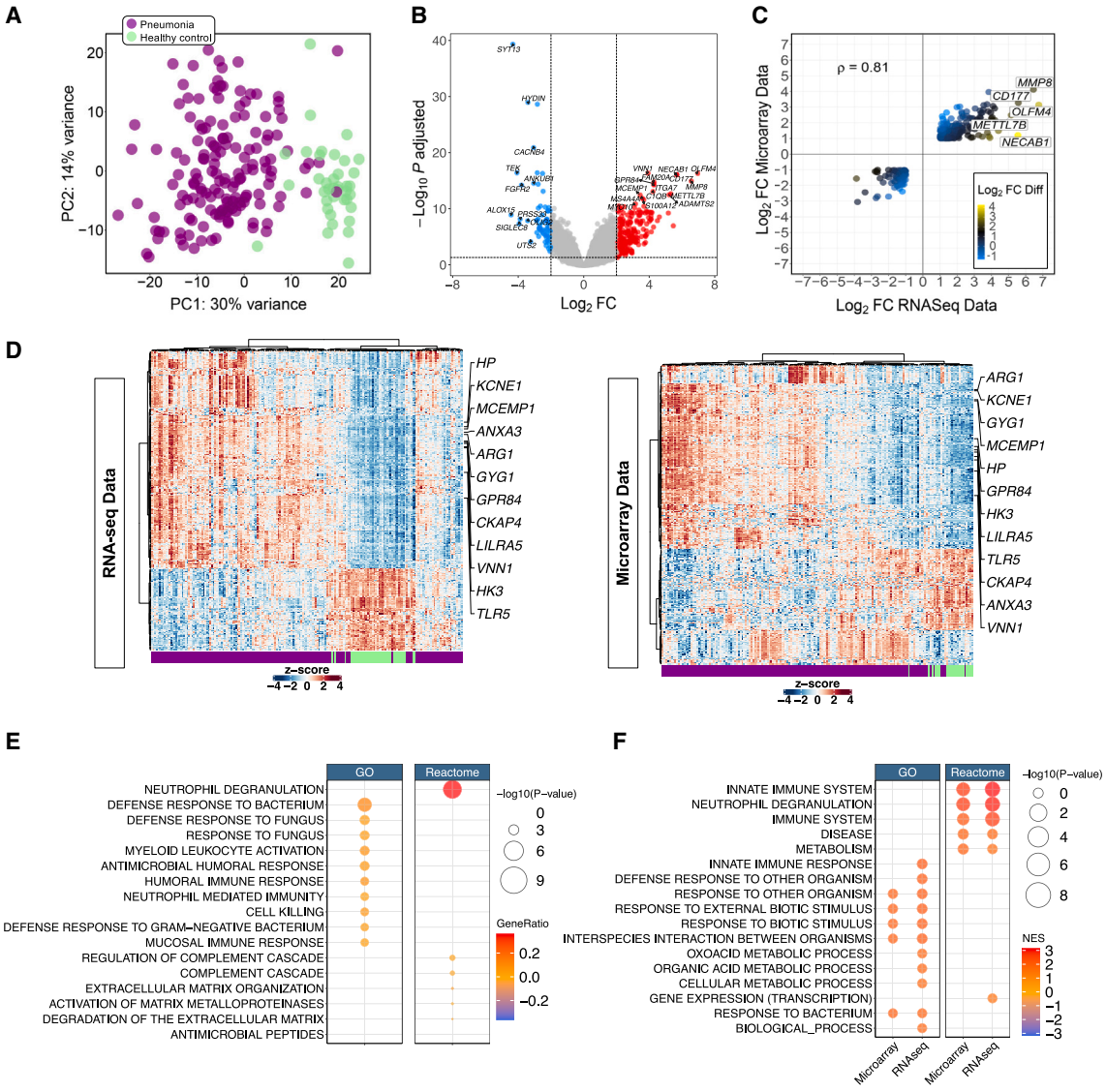


Figure 2. Pneumonia patients vs. healthy controls

(A) PCA of transcriptome profiles from pneumonia and healthy control samples. Two first principal components (PC1 and PC2) are shown.

(B) Volcano plot showing the DEGs between conditions: pneumonia vs. healthy control. Downregulated genes are colored blue, and upregulated genes are colored red (thresholds: adjusted p value = 0.05, $\log_2$FC = |2|).

(C) Correlation between $\log_2$FC of DEGs obtained from the comparison pneumonia vs. healthy controls in RNA-seq/microarray data. Only genes with adjusted p value < 0.05 and $\log_2$FC > |1| are displayed. The color scale represents the differences between the $\log_2$FC values of both analyses. The p value of the correlation is $2.2e^{-16}$, and only names of genes with a $\text{Log}_2\text{FC} > |5|$ are shown.

(D) Two-way hierarchical clustering analysis heatmap of DEGs between pneumonia and healthy control samples in RNA-seq and microarray validation cohorts. Each row represents one transcript; each column represents one patient. The bar at the bottom indicates the sample phenotype. Only genes with a $\log_2$FC > |1| and adjusted p value < 0.05 were represented in the heatmap, and only the genes that were common in the top 40 with the lower p value in both analysis were printed. Expression intensity is indicated by color (red, high expression; blue, low expression).

(E) Dot plot from over-representation analysis (ORA) pathway analysis of common DEGs between RNA-seq and microarray cohort with adjusted p value (FDR) < 0.05 and a $\log_2$FC > |1.5| for the comparison pneumonia patients vs. healthy controls using Gene Ontology (GO) and Reactome as reference. Size along the x axis indicates the number of genes in the input list that are annotated to the corresponding term/number of genes in the input list (gene ratio). Dot colors correspond to the different FDR p values associated with the pathways.

(F) Dot plot from GSEA pathway analysis of common DEGs between RNA-seq and microarray cohort with FDR p value < 0.05 and a $\log_2$FC > |1.5| for the comparison pneumonia patients vs. healthy controls using GO and Reactome as reference. Size of the dots corresponds to FDR p values associated with the pathways. Dot colors correspond to the pathway normalized enrichment scores (NESs) values.

"defense response to fungus" (adjusted p value = $1.0e^{-5}$), and related processes ("defense response to gram-negative bacterium" and "antibacterial humoral response") (Table S3; Figure 2E). It is not surprising that genes related to defense against bacteria were overrepresented in the overall analysis as most of the samples in the dataset correspond to bacterial-origin pneumonia.

Additionally, GO-ORA analysis highlighted a significant association of CAP with biological processes that play essential roles not only in innate protection against viral and bacterial invasion but also in disease pathogenesis. This encompasses regulatory pathways related to transcription factors, such as the pro-inflammatory nuclear factor κB (NF-κB) ("regulation of DNA-binding transcription factor activity," "positive regulation of DNA-binding transcription factor activity," and "positive regulation of NF-κB transcription factor activity"), as well as pathways indicating inflammatory response ("regulation of inflammatory response," "positive regulation of inflammatory response," and "acute inflammatory response") and activation of the cytokine signaling system ("positive regulation of cytokine production" and "cytokine-mediated signaling pathway") (Table S3; Figure 2E).

The results from the GSEA closely mirrored those obtained from both microarray[53] and our RNA-seq cohorts. Remarkably, all differentially regulated pathways exhibited significant overactivation in pneumonia cases (normalized enrichment score [NES] > 0) (Table S3; Figure 2F). The most significant pathways identified in the GO and Reactome analysis overlapped with those detected in the ORA, highlighting key processes such as the "response to other organism" and "response to bacterium" in GO. Additionally, pathways related to the "innate immune system" and processes like "neutrophil degranulation" in Reactome were also prominently featured.

Bacterial vs. viral pneumonia: Diagnostic signature

We conducted an additional PCA analysis to compare the transcriptomic profiles of pneumonia in DB and DV infections. Our goal was to investigate how these two groups cluster in relation to each other and in comparison to the control group. In the first PCA (Figure 3A), pneumonias separated distinctly from the control group along PC1, which accounted for 45% of the variation. Specifically, the DB profiles clustered at one pole of the plot, the controls in the opposite side, and the DV profiles in between. A separate PCA, including only DB (n = 40) and DV (n = 9) transcriptomic samples, revealed a relatively weak separation between the two groups. DV profiles primarily clustered to one side along its PC1 (28% of the variation), albeit with some slight mixing with other DB samples (Figure 3A).

Our differential expression analysis between DB and DV pneumonias identified 282 DEGs (FDR of 5%; Table S4). The majority showed under-expression in DB with respect to DV pneumonia (n = 217; 77%). However, when considering genes with more substantial expression changes (setting a threshold of adjusted p value < 0.05 and $|\log_2 FC| > 2$), we found that the proportion of over-expressed genes in DB pneumonia (n = 30; 55%) was slightly higher than the proportion of under-expressed genes (n = 25; 45%; Table S4; Figure 3B). Interestingly, *ALB* was the gene yielding the lowest adjusted p value ($1.88e^{-7}$) in the comparison between DB and DV patients (Table S4; Figures 3B and S1).

To further investigate, we selected a subset of DEGs candidates from the DV vs. DB differential expression in pneumonia samples (n = 36; see STAR Methods). We then employed the least absolute shrinkage and selection operator (LASSO) regression algorithm to identify the optimal transcriptomic signature for distinguishing between these two types of infection.

The best model for differentiation between viral and bacterial pneumonia comprised five genes (*FAM20A*, *BAG3*, *TDRD9*, *MXRA7*, and *KLF14*), forming the optimal signature (Table S5). Among these genes, four were over-expressed (*FAM20A*, *TDRD9*, *MXRA7*, and *KLF14*) while one (*BAG3*) showed under-expression in bacterial compared to viral pneumonia. This transcriptomic signature statistically discriminated between viral and bacterial pneumonia with high accuracy (p value = $3.0e^{-5}$; optimal cutoff value = 2.204; Figure 3C) and demonstrated excellent performance in the training set, generating an area under the curve (AUC) value of 0.95 [0.88–1.00] (sensitivity = 0.80 and specificity = 1.00) (Figure 3C; Table S6). Moreover, the signature reliably differentiated bacterial pneumonia from healthy controls (AUC: 0.91 [0.84–0.98]) (Figure S2).

Due to the limited number of samples classified as viral pneumonia (DV = 9; PV = 3), we were unable to split the dataset into training and test sets conventionally. Instead, we adopted an initial validation approach, evaluating the performance of the transcriptomic signature in a test dataset that included not only the samples confidentially classified as DB and DV, which had been already used in the training dataset for the signature discovery (DB = 40; DV = 9), but also all PB (n = 58) and PV (n = 3) pneumonias (DB + PB = 98; DV + PV = 12; Figure 1). This 5-transcripts signature identified in the training cohort exhibited robust discriminatory capacity when validated in this test set (p value = $2.8e^{-5}$, Figure 3D). In this scenario, the receiver operating characteristic (ROC) curve also indicates high accuracy, resulting in an AUC value of 0.87 [0.77–0.97] (sensitivity = 0.76 and specificity = 0.83; Figure 3D; Table S6). Considering only PV and PB samples, we obtained an expected decrease in the AUC value (0.69 [0.40–0.97]), as these samples are coded with a phenotype of lower clinical confidence and are therefore more likely to contain viral samples within the PB set, bacterial samples within the PV set, or mixed infections (see score values for classification in Figures 3C and 3D; see also Figure S3A).

The performance of our 5-transcriptomic signature was also compared to the previously described 2-transcript signature that generically differentiates viral from bacterial infections in febrile children using *IFI44L* and *FAM89A* transcripts.[30] Although the 2-transcript signature showed a good performance discriminating viral and bacterial pneumonia (AUC: 0.87 [0.76–0.96]) in our discovery dataset, the AUC value was lower than that obtained from the specific 5-transcript signature we propose (Figure S3B). When we tested the 2-transcripts viral/bacterial signature in the PB vs. PV subset, the AUC value was also lower than the AUC from the pneumonia signature (AUC: 0.44 [0.00–0.89]; Figure S3C). Finally, to check the best performance that we could achieve with these two transcripts in our discovery dataset, we re-trained the model using *IFI44L* and *FAM89A* transcripts. In this case we

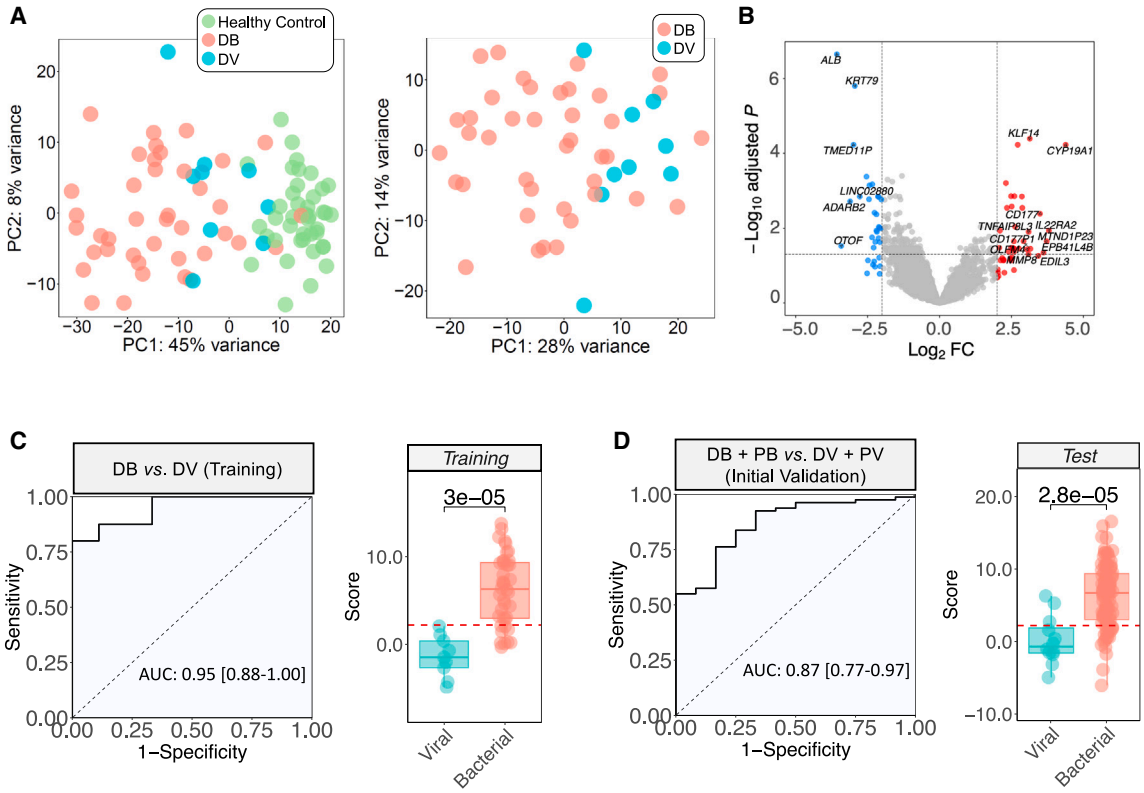


Figure 3. Viral and bacterial pneumonias

(A) PCA of transcriptome profiles of DV and DB pneumonias and healthy control samples. Two first principal components (PC1 and PC2) are shown.

(B) Volcano plot showing the DEGs between conditions: DB pneumonia vs. DV pneumonia (DV). Downregulated genes are colored blue, and upregulated genes are colored red (thresholds: adjusted *p* value = 0.05, $\log_2$FC = |2|).

(C) Receiver operating characteristic (ROC) curves based on the specific 5-transcript signature from the training cohort including the area under the curve (AUC) and 95% confident intervals (CIs) values (left). Boxplots of the predicted values using the optimal model in the training cohort. Wilcoxon *p* value is also displayed (right).

(D) Receiver operating characteristic (ROC) curves based on the specific 5-transcript signature from the test cohort including the area under the curve (AUC) and 95% confident intervals (CIs) values (left). Boxplots of the predicted values using the optimal model in the training cohort. Wilcoxon *p* value is also displayed (right). Red dashed line represents the optimal cutpoint. The boxes are defined by the upper and lower quartile (Q1 and Q3); whiskers extend to the most extreme data point, which is no more than 1.5 times the IQR from the box; the median is shown as a bold-colored horizontal line.

obtained an AUC value of 0.88 [0.77–0.98] (Figure S3D), which is still below the AUC from the specific pneumonia signature (AUC: 0.95 [0.88–1.00]).

Additionally, we tested the 5-transcripts signature and the 2-transcript signature from Herberg et al.[30] in the bacterial and viral pneumonia samples (excluding co-infections) from the dataset of Wallihan et al.[53] Both signatures yielded AUC values close to 0.5 for this dataset, with the 5-transcript signature achieving an AUC of 0.52 [0.40–0.64] and the 2-transcripts signature from Herberg et al. reaching an AUC of 0.56 [0.45–0.68]. These results indicate limited discrimination ability for these signatures, within this specific dataset.[30]

Validation of the 5-transcripts pneumonia's signature

We validated the diagnostic accuracy of the proposed signature using new gene expression data obtained from an independent cohort of children with pneumonia (Tables 2 and S7). The validation results confirm that the 5-transcript signature can accurately differentiate between DV and DB pneumonias in this new cohort, achieving an impressive AUC of 0.92 [0.83–1.00] (Figures 4A and 4B).

To ensure that the effectiveness of the signature was not influenced by disease severity or specific pathogens, we conducted a stratified analysis based on the causal pathogens of DB pneumonias—*S. pneumoniae*, *S. aureus*, and *S. pyogenes*—and by severity (mild/moderate vs. severe). The results affirmed the robustness of the signature across different bacterial pathogens: AUC values of 0.95 [0.87–1.00] for *S. pneumoniae*, 0.82 [0.54–1.00] for *S. aureus*, and a perfect AUC of 1.00 [1.00–1.00] for *S. pyogenes* (Figure 5A). Moreover, the signature consistently demonstrated high performance in distinguishing between viral and bacterial pneumonias across both mild/moderate and severe cases, with AUC values of 1.00 [1.00–1.00] for mild/moderate cases and 0.92 [0.77–1.00] for severe cases (Figure 5B).

Co-expression analysis of viral vs. bacterial pneumonia

To understand the distinct host mechanisms and key genes involved in the specific response to viral and bacterial

Table 2. Clinical features of the validation cohort

	Healthy controls	DB vs. DV				
	Controls (n = 35)	DB (n = 22)	n	DV (n = 10)	n	p value
Gender (male)	54.3% (19/35)	63.6% (14/22)	22	60% (6/10)	10	1
Age (years)	4.2 (1.24–6.17)	4.8 (0.1–17.5)	22	2.4 (0.1–15.6)	10	0.167
Days hospitalized	–	14.0 (3.5–14)	22	6 (4.75–8.5)	8	0.038
PICU admission	–	54.5% (12/22)	22	40%% (4/10)	10	0.703
PICU days	–	2.0 (0.0–14.0)	22	7,5 (2.00, 8.00)	4	0.843
Death	–	4.5% (1/22)	22	0% (0/9)	10	1
Treatments						
Inotropes	–	32% (7/22)	22	20% (2/10)	2	0.379
IV	–	31.8% (7/22)	22	30% (3/10)	9	0.750
NIV	–	13,6% (3/22)	22	30% (3/10)	9	1
Oxygen	–	54.5% (12/22)	22	60% (6/10)	9	0.637
Analytical parameters						
CRP (mg/L)	–	208.4 (46.2–375.4)	16	18.2 (1.3–52.5)	10	0.084
Platelets (10^9/L)	–	249.0 (17.0–1299.0)	19	271.5 (141.0–522.0)	10	0.449
White cells (10^9/L)	–	13.5 (4.8–28.3)	19	5.8 (2.8–11.5)	10	0.001*
Lactate (mmol/L)	–	1.4 (0.6–5.9)	8	0.7 (0.5–1.7)	4	0.147
Neutrophils (10^9/L)	–	9.6 (3.7–24.0)	15	3.0 (0.9–5.1)	9	0.001*
Monocytes (10^9/L)	–	0.8 (0.1–3.4)	15	0.4 (0.1–2.0)	8	0.258
Lymphocytes (10^9/L)	–	2.5 (0.2–6.0)	15	2.6 (0.9–4.7)	9	0.765

Wilcoxon and Fisher exact test were used to assess statistical significance between groups in numerical and categorical variables, respectively. IV, invasive ventilation; NIV, non-invasive ventilation. * indicates statistical significance.

pneumonia, we conducted a co-expression network analysis using the subset of pneumonia samples with DB or DV origin. The analysis detected 16 modules of co-expressed genes (Figure S4), with three of them significantly correlated (p value < 0.05) with viral/bacterial phenotypes (Figures 6A and 6B). The blue module (Enah//Vasp-like [EVL] module, $R = -0.4$; p value = 0.0043; 661 genes) and darked module (Transforming Growth Factor Beta 3 [TGFB3] module, $R = -0.32$; p-value = 0.025; 73 genes) exhibited negative correlation with bacterial pneumonia. In contrast, the salmon module (mitogen-activated protein kinase [MAPK]14]) showed a positive correlation (R = 0.32; p value = 0.026; 1,132 genes) with bacterial pneumonia (Figures 6A and 6B). The EVL and TGFBR3 modules clustered together on the dendrogram, indicating their involvement in similar biological processes globally activated during viral pneumonia (Figure 6A). Genes within these significant modules displayed a strong correlation between trait significance (GS) and module membership (MM), suggesting that highly interconnected genes (higher MM) within the module are closely related to the causal pathogen of pneumonia (Figure 6B). These driver genes, located in the upper right side of the plots, are pivotal for predicting viral or bacterial pneumonia.

Functional assessment of the significant modules unveiled different immune responses for viral and bacterial pneumonia. The modules over-regulated in viral pneumonia showed enrichment in terms associated with both innate and adaptative immune responses. For example, the EVL module was linked to adaptative T cell response (T cell selection; T cell receptor signaling; and T cell activation, differentiation, and proliferation) and interferon production (interferon gamma production) (Figure 6C; Table S8), while the TGBR3 module participated in "natural killer mediated immunity" (innate response) and, as in the case of the EVL module, in an adaptative response involving "leukocyte mediated cytotoxicity" and "T cell mediated immunity" (Table S8).

Conversely, the significant pathways related to genes from the bacterial MAPK14 module primarily represent a different innate response characterized by "positive regulation of cytokine production," "cytokine-mediated signaling pathway," "acute inflammatory response," "pattern recognition receptor signaling," and "myeloid leukocytes activation" (granulocytes, monocytes, macrophages, and dendritic cells). These findings were further supported by an ORA using the Reactome database, emphasizing the role of MAPK14 module genes in innate immune responses, including neutrophil degranulation, various Toll-like receptor (TLR) cascades, and signaling by interleukins (Figure 6D; Table S8).

DISCUSSION

Pneumonia is a complex disease characterized by diversity in its etiology, symptomatology, and clinical management. This complexity extends to decisions regarding the prescription of antibiotics and hospitalization for children, which remains a significant issue in terms of medical cost.[13] Some efforts have been made to identifying molecular biomarkers to assist in diagnosing and predicting pneumonia outcomes. However, there is still not an optimal test to accurately diagnose pneumonia in children or

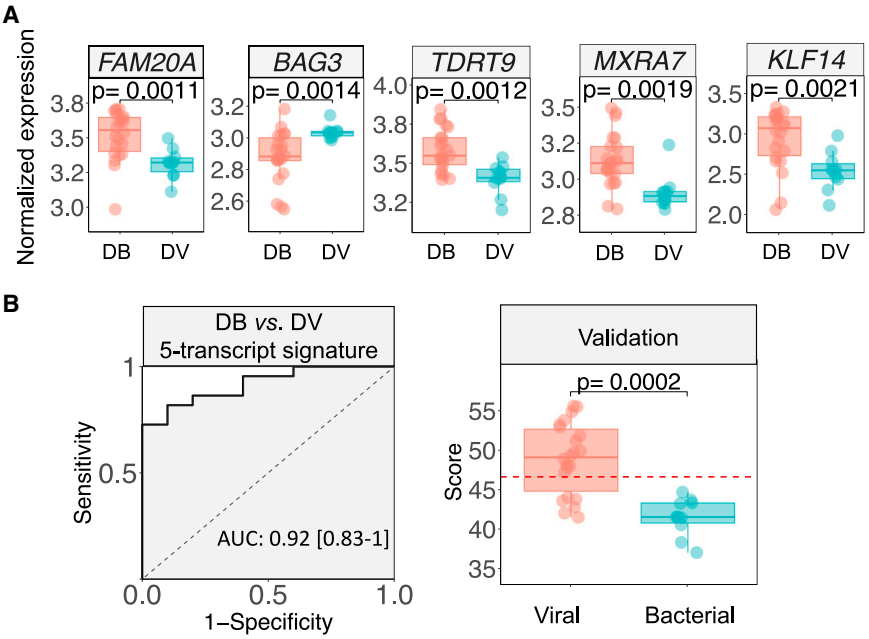


Figure 4. Validation cohort

(A) Boxplots showing the expression values of the genes included in the 5-transcript signature in the validation cohort. The boxes are defined by the upper and lower quartile (Q1 and Q3); the median is shown as a bold-colored horizontal line; whiskers extend to the most extreme data point, which is no more than 1.5 times the IQR from the box.

(B) ROC curve, AUC value with confidence interval, and boxplot of the predicted value obtained from applying the 5-transcript viral/bacterial signature coefficients in the validation cohort. Wilcoxon test *p* values are displayed in the boxplots.

to distinguish between bacterial and viral pneumonia. Consequently, antibiotics are often prescribed when a bacterial infection cannot be ruled out leading to unnecessary cost and contributes to the development of antimicrobial resistance.[9,54]

Our current analysis reveals that pneumonia in children induces extensive modifications in the host's transcriptome. Our findings indicate that pneumonia alters the expression of more than 5,000 genes in blood of affected children. Subsequent validation of these differential expression results in an independent cohort of pediatric pneumonia cases confirmed these findings. Additionally, two independent enrichment approaches (ORA and GSEA) using GO and Reactome as reference databases revealed a convergence of altered pathways within the immune system. Specifically, these pathways involve the activation and migration of leukocytes, with significant focus on neutrophil transcripts. During pneumonia, there is recruitment of neutrophils into the lungs. In cases of acute lung injury, a distinct pattern of activated neutrophils, referred to as the neutrophil activation signature,[55] emerges. This pattern is unique to the process of neutrophils migrating toward infected lung tissue, leading to various responses, including the release of bactericidal enzymes through degranulation, the discharge of cytokines and reactive oxygen species, and the generation of neutrophil extracellular traps. While these activities facilitate the elimination of causative pathogens, they may also result in tissue damage and exacerbate acute inflammation.[56–58]

Most prominent overexpression changes detected in our RNA-seq cohort, which were subsequently validated in an independent dataset, were represented by genes that produce proteins secreted by secondary granules released from neutrophils. These genes include *OLFM4*, *MMP8*, and *CD177*, and some of them have been previously reported as candidate biomarkers for the diagnosis/prognosis of CAP and predictors of sepsis.[59–61] MMP8 is a collagenase derived from neutrophils that has implications in a wide variety of pulmonary pathologies. It can degrade all components of the extracellular matrix and plays a role in neutrophil migration.[62] Its expression has already been shown to be significantly higher in pneumonia patients compared to healthy controls, representing one of the most over-expressed gene in adult pneumonia samples from various studies,[27,63] and also in our pediatric cohort. Serum and plasma levels of MMP8 protein have been proposed for diagnosing CAP, and as prognostic biomarker of fatal outcome in septic patients and predictive of serious bacterial infection, suggesting that could be an inflammatory modulator in sepsis.[64–66] However, its poor specificity limits its use as a single protein diagnostic biomarker. Our findings support the evidence that *MMP8* gene is also over-expressed in children with pneumonia. Similarly, we observed a significant overexpression of the *OLFM4* gene in pneumonia patients. *OLFM4* is a matrix glycoprotein of neutrophil-specific granules and has been described as a marker of severity in infectious diseases. In children, it has been identified as an independent risk factor for poor outcomes in sepsis,[67,68] and its inhibition has been proposed as a possible therapeutic approach in infected patients.[68] *CD177* is a glycosylphosphatidylinositol-anchored protein that plays a crucial role in regulating neutrophils by modulating their migration and activation. *CD177* has been identified as the most dysregulated molecule in purified neutrophils from patients with septic shock and severe influenza infection.[69] Recently, it has also been associated with worse outcome and higher mortality in patients infected with SARS-CoV-2[70] and included in a diagnostic four-gene signature for pediatric sepsis.[59]

ABL was the gene yielding the lowest significant value when comparing DB and DV patients. *ALB* encodes for the albumin protein, the most abundant protein in human blood, but it is primarily expressed in the liver and pancreas (https://www.gtexportal.org/). Although the overall expression of ALB in our blood samples was low (median $\log_2$ values of 2.07, 5, and 3.09 for DB, DV, and healthy controls, respectively), it showed a significant overexpression in DV compared to DB patients. Some studies have showed that decreased levels of albumin in serum and plasma may indicate a higher risk of ICU admission and death in sepsis,[71,72] serve as a diagnostic biomarker for infectious diseases,[73] and act as a risk

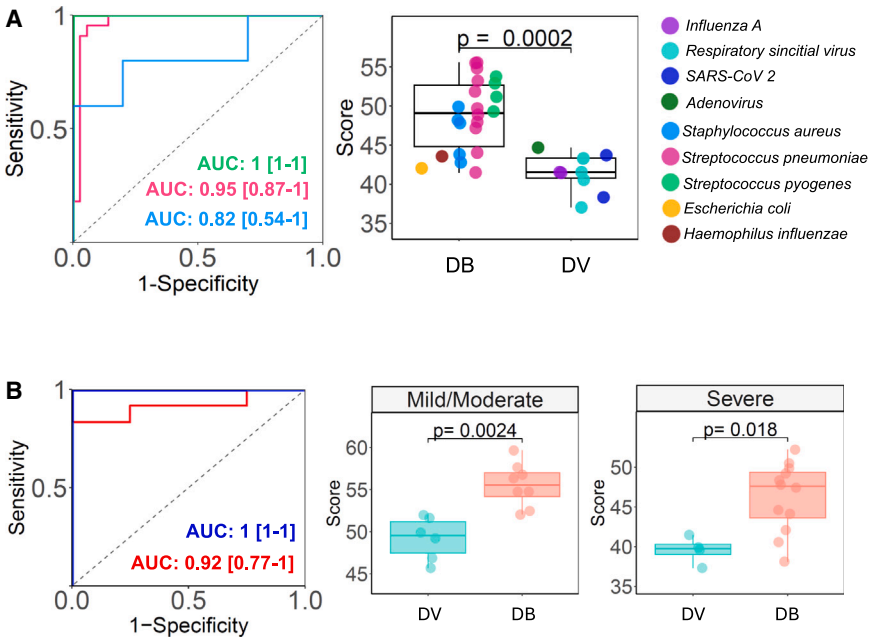


Figure 5. Performance of the 5-transcript signature

(A) ROC curves, AUC values, and boxplots of the performance of the 5-transcript signature in pneumonias with different causal pathogens. Only bacterial groups with more than 3 samples were considered for the AUC and ROC analysis. The boxes are defined by the upper and lower quartile (Q1 and Q3); the median is shown as a bold-colored horizontal line; whiskers extend to the most extreme data point, which is no more than 1.5 times the IQR from the box.

(B) ROC curves, AUC values, and boxplots of the performance of the 5-transcript signature in pneumonias with different severities (severe or mild/moderate).

factor for more severe clinical manifestations in COVID-19 patients.[74,75]

In addition to the neutrophil-related innate response, our analysis identified a common enrichment of terms related to humoral immunity processes, consistent with previous results from pediatric pneumonia cases,[76] as well as an enrichment of innate processes commonly activated in response to viral and bacterial infections, such as the cytokines release and inflammation. At the same time, however, the over-activation of these processes can also be responsible for the clinical consequences of severe pneumonia, increasing the risk of severe sepsis and death.[77,78]

Since pneumonia can be caused by a broad range of viral or bacterial pathogens, it is essential to investigate new approaches to differentiate the pathogenic origin of the disease, which may support decisions regarding antibiotic prescriptions. Molecular methods for detecting causative bacteria remain less effective in pneumonia patients, mainly due to challenges in obtaining samples from the infection site and the frequent absence of positive results from accessible sites such as blood.[79] Moreover, detecting viruses in nasopharyngeal secretions from pneumonia patients is common due to colonization, but this does not exclude the possibility of a concurrent bacterial infection.[80,81] A comparison of gene expression between DV vs. DB pneumonia revealed 282 DEGs between both categories. Functional studies of the co-expression modules driving the host-specific responses to viral or bacterial pneumonia highlighted differentially activated immunological pathways in both pneumonia etiologies. We identified two significant modules associated with viral pneumonia (EVL and TGBR3) and one associated with bacterial pneumonia (MAPK14). Genes from the MAPK14 module are primarily involved in innate responses related to cytokine production, acute inflammation, neutrophil degranulation, deficiencies in TLR cascade, and interleukin signaling. TLR receptors are critical in constraining the proliferation and dissemination of invading bacteria. Activation of TLRs triggers innate immune signaling pathways and transcriptional upregulation of pro-inflammatory mediators that recruit neutrophils to pulmonary tissue during bacterial infection.[82] An effective and timely inflammatory response is key to pathogen clearance and control the infection in pneumonia. However, an exacerbated and prolonged release of pro- and anti-inflammatory cytokines, followed by neutrophil hyper-responsiveness, is associated with an increased risk of severe sepsis and poor outcomes in CAP.[77,83] TLRs can detect various conserved patterns from both viral and bacterial pathogens. Interestingly, significant pathways involving TLRs detected in the bacterial module are among the TLR cascades commonly activated by bacterial pathogen-associated molecular patterns (PAMPs; e.g., lipoproteins TLR1/TLR2/TLR6 and lipopolysaccharide TLR4).[84] Most TLRs, except TLR3, signal via myeloid differentiation primary response 88 (MyD88) and interleukin-1 receptor-associated kinase 4 (IRAK-4). IRAK-4 is critical for MyD88-dependent TLR signaling, and patients with IRAK-4 mutations are susceptible to recurrent bacterial infections.[85,86] Patients with both IRAK-4 and MyD88 deficiencies are predisposed to severe bacterial infection,[87] and defects in MyD88-mediated TLR signaling make children much more susceptible to pneumonia.[88] The canonical TLR pathway can activate NF-κB and MAPKs through different mediators; in this regard, the hub gene of our bacterial module, *MAPK14*, is a key gene in the development of ventilator-associated bacterial pneumonia in adults.[89]

Collectively, these findings suggest that immune responses to severe bacterial pneumonia are primarily characterized by dysfunction in the innate TLR signaling. This dysfunction leads to an exacerbated inflammatory response, driven by cytokine release and neutrophil degranulation, which may play a pivotal role in the development of the severe clinical symptoms observed in these patients.

In the case of viral pneumonia-correlated modules (EVL and TGBR3), we identified a significant representation of adaptative immunity mediated by T cells. Genes within the EVL module were associated with processes such as T cell selection; T cell receptor signaling; and T cell activation, differentiation, and

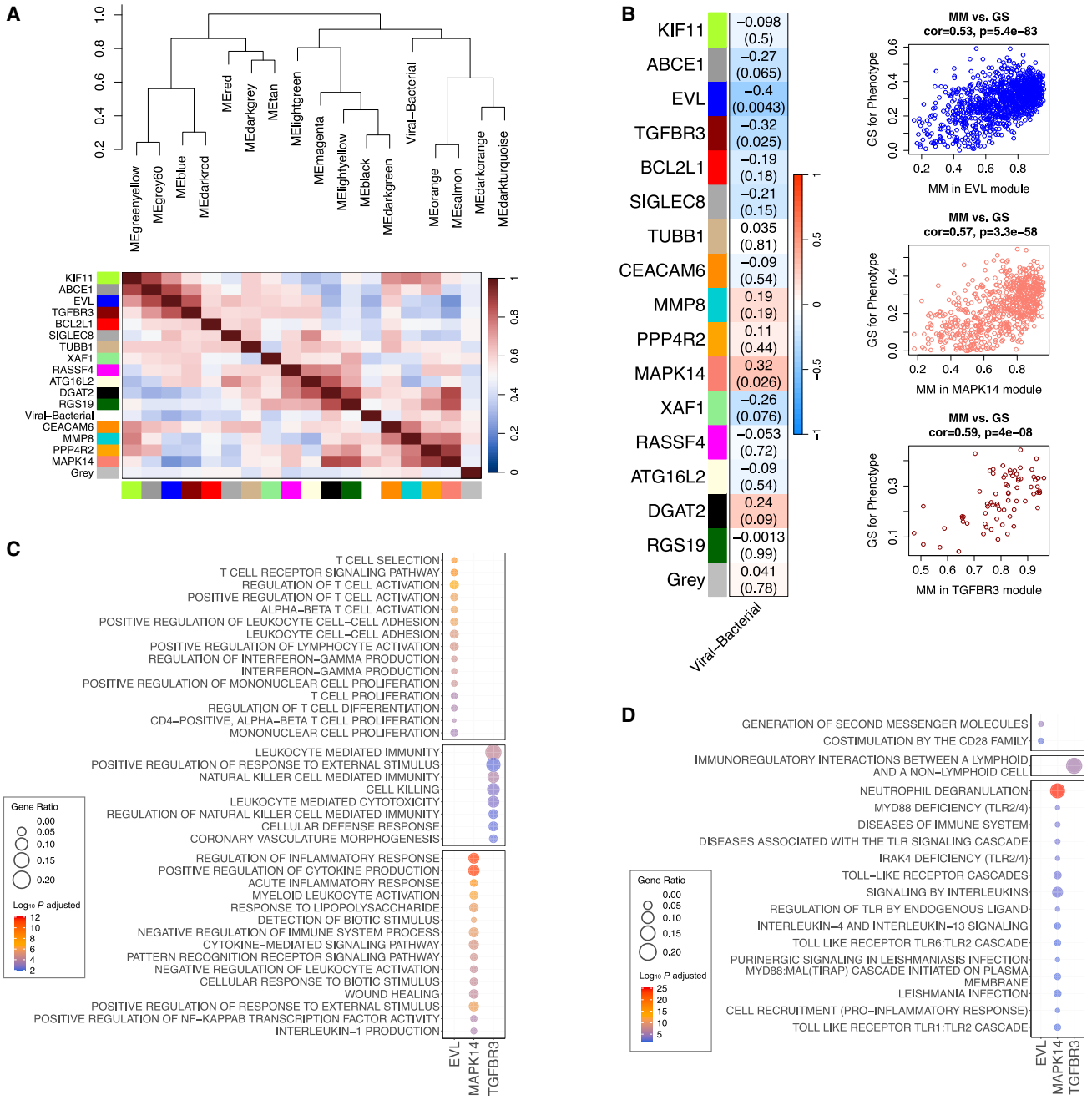


Figure 6. Co-expression analysis of viral vs. bacterial pneumonia

(A) Hierarchical clustering eigengene dendrogram and heatmap showing relationships among the modules and pneumonia etiology (DB and DV pneumonias). Gene names on the left of the heatmap are the hub genes of each module.

(B) Correlation values heatmap between co-expression modules and DB and DV conditions (DV was considered the reference group). Values show individual correlations of the module with DB phenotype (left). Correlation values are also indicated with different colors (legend gradient color bar). *p* values of these correlations are represented in brackets (lower values). Plots showing comparison between MM (module membership) and GS (correlation with bacterial/viral condition) of genes from the most significant modules detected.

(C) Top 15 statistically associated GO biological processes detected for each of the significantly correlated modules.

(D) Top 15 statistically associated Reactome pathways detected for each of the significantly correlated modules.

proliferation. The TGFBR3 module also reveals pathways related to adaptative immunity, such as "leukocyte mediated cytotoxicity" and "T cell mediated immunity." T cells are essential in combating viral infections, by inducing cell death through cytokines release.[90] Notably, both $CD4^+$ and $CD8^+$ T cells increase in population during the first days to weeks of an acute viral infection.[91] $CD8^+$ cells have been described to be critical in clearing various viral infections in mouse models,[92] while $CD4^+$ T cells serve as helpers for B cells and $CD8^+$ T cells, also possessing independent antiviral function through cytokine secretion.[93] Of interest, polymorphisms in the hub gene *TGFBR3* were previously associated with a history of pneumonia in sickle cell disease patients.[94] *EVL* expression was associated with poor prognosis in sepsis.[95]

We have identified a 5-transcript blood signature that can accurately differentiate bacterial from viral pneumonias with an AUC of 0.95 (sensitivity = 0.80; specificity = 1.00), demonstrating superior discriminant power between viral and bacterial pneumonia compared with previously published non-syndrome-specific signatures for differentiating viral from bacterial infections in children (AUC: 0.87 [0.75–0.96] and 0.88 [0.78–0.98] in the re-trained model).[30] However, as the general signature also performs effectively in this cohort, both methods could be implemented complementarily in a clinical context for the definitive confirmation of the diagnosis. The 5-transcript signature's performance remains high when including pneumonia samples of both PB and PV origin (AUC = 0.87; sensitivity = 0.76; specificity = 0.83), affirming the robustness and discriminatory power of the transcriptomic signature. Among the genes selected in this signature, some are associated with antiviral or bacterial activity. For instance, *BAG3* is a cochaperone recently linked to intracellular bacterial proliferation[96]. *FAM20A*, a pseudokinase, exhibits increased expression in the lung and liver,[97] in sepsis,[98] and in blood neutrophils of acute respiratory distress syndrome patients.[55] The *TDRD9* gene was previously described in a transcriptomic signature for sepsis derived from CAP.[99] *MXRA7* is highly expressed in murine and human ocular tissues and may play a role in pathological processes involving injury, neovascularization, and wound healing, although its function remains mainly unknown.[100] Finally, the Krüppel-like transcription factor *KLF14* exhibits a modulatory role in macrophages-driven inflammation[101] and has been found to be upregulated in septic patients.[102] Its potential regulatory function in sepsis positions this gene as a therapeutic target candidate for sepsis.

We have also observed that the 5-transcript signature identified in the present study and the signature identified by Herberg et al.[30] for differentiating viral from bacterial infection did not prove effective in the pneumonia cohort from Wallihan et al..[53] This outcome can be explained by the fact that the vast majority of patients with bacterial pneumonia in the Wallihan et al. cohort have infections caused by *Mycoplasma pneumoniae* (30/35; 86%). These results suggest that the host gene expression response to this pathogen is specific and distinct from that triggered by other pneumonia-causing bacteria. The fact that *M. pneumoniae* is generating worrying outbreaks in the post-COVID-19 pandemic era[103,104] deserves further dedicated investigation.

The validation of the transcriptomic signature in an independent cohort of children with pneumonia further reinforces its high specificity and sensitivity in accurately distinguishing bacterial from viral infections. Importantly, our analysis confirmed that the effectiveness of the signature is independent of both the causal bacterial pathogen responsible and the severity of the disease. This robust performance across diverse clinical contexts underscores the potential of this signature as a reliable diagnostic tool for pediatric pneumonia, offering a significant advancement in precision and reliability over current diagnostic methods.

Our study involved hospitals from different countries, reducing the potential bias from specific populations or areas. It has characterized and independently validated transcriptomic changes caused by CAP in hospitalized children.

We have uncovered various immune mechanisms and genes that are distinctly activated in viral and bacterial pneumonia, making a significant advance in the field. Most notably, our 5-transcript signature stands out for its ability to accurately differentiate between bacterial and viral pneumonia, even in cases of uncertain origin. Crucially, this signature might make such distinctions before the full clinical symptoms emerge. This early diagnostic capability holds the potential to drastically reduce the time to a definitive diagnosis compared to traditional culture-based methods, ultimately shortening hospital stays, and mitigate the consequences of delayed or missed diagnoses, such as unnecessary hospital investigations and antibiotic prescriptions when a bacterial infection is suspected. These benefits position our findings as a major step forward in improving pneumonia management.

Limitations of the study

The primary limitation of our study is the limited number of samples from children with viral pneumonia. Therefore, further validation is required, utilizing additional datasets that encompass well-defined pneumonia infections of both viral and bacterial origins. Despite this limitation, our results reaffirm the classification of pediatric patients into viral and bacterial pneumonia categories based on the clinical algorithm developed by Herberg et al..[30]

CONSORTIA

GENDRES network

Miriam Cebey López, Antonio Salas Ellacuriaga, Ana Vega Gliemmo, José Peña Guitián, Alexa Regueiro, Antonio Justicia Grande, Leticia Pías Peleteiro, María López Sousa, María Jose de Castro, Carmen Curros Novo, Elena Rodrigo, Miriam Puente Puig, Rosaura Leis Trabazo, Nazareth Martinón Torres, Alberto Gómez Carballa, Jacobo Pardo Seco, Sara Pischedda, José María Martinón Sánchez, Belén Mosquera Pérez, Isabel Villanueva González, Lorenzo Redondo Collazo, Carmen Rodríguez-Tenreiro, María del Sol Porto Silva and Federico Martinón Torres, Máximo Francisco Fraga Rodríguez, Orlando Fernández Lago, José Ramón Antúnez, Enrique Bernaola Iturbe, Laura Moreno Galarraga, Jorge Álvarez, Mercedes Herranz, Francisco Gil, Eva Gembero, Jorge Rodríguez, Teresa González López, Delfina Suarez Vázquez, Ángela Vázquez Vázquez, Susana Rey García, Nathalie Carreira Sande, Ana López Fernández, Nuria Romero Pérez, José Antonio Couceiro Gianzo, Nazareth Fuentes Perez, Francisco Giménez Sánchez, Miguel Sánchez Forte, Cristina Calvo Rey, María Luz García García, Iciar Olabarrieta Arnal,

Adelaida Fernández Rincón, Ignacio Oulego Erroz, David Naranjo Vivas, Santiago Lapeña, Paula Alonso Quintela, Jorge Martínez Sáenz de Jubera, Estibaliz Garrido García, Cristina Calvo Monge, Eider Oñate Vergara, Jesús de la Cruz Moreno, M[a] Carmen Martínez Padilla, Beatriz Jiménez Jurado, Carmen Santiago Gutiérrez, María Esther Vidaurreta del Castillo, Manuel Baca Cots, David Moreno Pérez, Ana Cordón Martínez, Antonio Urda Cardona, José Miguel Ramos Fernández, Esmeralda Núñez Cuadros, Susana Beatriz Reyes, María Cruz León León, Santiago Alfayate, Cristina Calvo, Carlos Grasa, Cristian Quintana Ortega, Leticia La Banda Montalvo, María Lopez Cerdán, Ana Dominguez Castells, Francisco Giménez Sánchez, Andrés J. Alcaraz Romero, Diego Bautista Lozano, Sara Uillen Martin, Roi Piñeiro, Juan Ignacio Sánchez Díaz, Alba Palacios Cuesta, Elvira González Salas, Sira Fernández De Miguel, Belen Joyanes Abancens, Esther Aleo Lujan, Alfredo Tagarro García, María Luisa Herreros, Rut del Valle, Libertad Latorre Navarro, María Concepción Zazo Sanchidrián, Mariano Esteban, Marta González Lorenzo, M[a] Carmen Vicent Castello, Lorena Moreno Requena, Juan Luis Santos Pérez, César Gavilán Martín, Lucía González-Moro Azorín, Monterrat López Franco, Manuel Silveira Cancela, María José Cilleruelo Ortega, Luz Golmayo, Francisca Portero Azorín, Andrés Concha Torre, Lucía Rodríguez García, Carlos Rodrigo Gonzalo de Liria, Andrés Antón Pagarolas, Maria Méndez, Cristina Prat, Jesús López-Herce, Miriam García Samprudencio, Gema Manrique Martín, Paula García Casas, Débora Sanz Álvarez, María Jesús Cabero, Miguel Lillo Lillo, Marta Pareja, Pablo Rojo, Cristina Epalza (Hospital 12 Octubre); Adriana Navas Carretero, Estefanía Barral, Miriam Herrera, Elvira Cobo Vázquez, Elena del Castillo Navío, Patricia Flores Perez, Paula García Casas, Esteban Gómez Sánchez, Juan Valencia Ramos, Francisco Javier Pilar Orive, Elva Rodríguez Merino, Ana Pérez Aragón, M[a] Yolanda Ruiz del Prado, David Moreno, Beatriz Carazo, Jordi Antón, and María teresa Rives Ferreiro. Further details may be consulted at http://www.gendres.org, and www.genvip.eu.

EUCLIDS consortium

Michael Levin, Lachlan Coin, Stuart Gormley, Shea Hamilton, Jethro Herberg, Bernardo Hourmat, Clive Hoggart, Myrsini Kaforou, Vanessa Sancho-Shimizu, Victoria Wright, Amina Abdulla, Paul Agapow, Maeve Bartlett, Evangelos Bellos, Hariklia Eleftherohorinou, Rachel Galassini, David Inwald, Meg Mashbat, Stefanie Menikou, Sobia Mustafa, Simon Nadel, Rahmeen Rahman, Clare Thakker, Sumit Bokhandi, Sue Power, Heather Barham, Nazima Pathan, Jenna Ridout, Deborah White, Sarah Thurston, Saul Faust, Sanjay Patel, Jenni McCorkell, Patrick Davies, Lindsey Crate, Helen Navarra, Stephanie Carter, Ramesh Ramaiah, Rekha Patel, Catherine Tuffrey, Andrew Gribbin, Sharon McCready, Mark Peters, Katie Hardy, Fran Standing, Lauran O'Neill, Eugenia Abelake, Akash Deep, Eniola Nsirim, Andrew Pollard, Louise Willis, Zoe Young, C. Royad, Sonia White, PM. Fortune, Phil Hudnott, Federico Martinón Torres, Antonio Salas Ellacuriaga, Fernando Álvez González, Ruth Barral-Arca, Miriam Cebey-López, María José Curras-Tuala, Natalia García, Luisa García Vicente, Alberto Gómez-Carballa, Jose Gómez Rial, Andrea Grela Beiroa, Antonio Justicia Grande, Pilar Leboráns Iglesias, Alba Elena Martínez Santos, Nazareth Martinón-Torres, José María Martinón Sánchez, Beatriz Morillo Gutiérrez, Belén Mosquera Pérez, Pablo Obando Pacheco, Jacobo Pardo-Seco, Sara Pischedda, Irene Rivero Calle, Carmen Rodríguez-Tenreiro, Lorenzo Redondo-Collazo, Sonia Serén Fernández, María del Sol Porto Silva, Ana Vega, Jose Manuel Fernández García, María Elena Gamborino Caramés, María Sol Rodríguez Calvo, Marta Aldonza Torres, Vanesa Álvarez Iglesias, Carmen Curros Novo, Isabel Ferreirós Vidal, Narmeen Mallah, Laura Navarro Ramón, Isabel Rego Lijo, Alba Camino Mera, Lúa Castelo Martínez, Ana Isabel Dacosta Urbieta, Wiktor Dominik Nowak, Julia García Currás, Nour El Zaharaa Mallah, Julián Montoto Louzao, Sara Rey Vázquez, Sandra Viz Lasheras, Patricia Regueiro Casuso, Miriam Taboada Puga, Lucía Vilanova Trillo, Susana Beatriz Reyes, María Cruz León León, Álvaro Navarro Mingorance, Xavier Gabaldó Barrios, Eider Oñate Vergara, Andrés Concha Torre, Ana Vivanco, Reyes Fernández, Francisco Giménez Sánchez, Miguel Sánchez Forte, Pablo Rojo, J. Ruiz Contreras, Alba Palacios, Cristina Epalza Ibarrondo, Elizabeth Fernandez Cooke, Marisa Navarro, Cristina Álvarez Álvarez, María José Lozano, Eduardo Carreras, Sonia Brió Sanagustín, Olaf Neth, M[a] del Carmen Martínez Padilla, Luis Manuel Prieto Tato, Sara Guillén, Laura Fernández Silveira, David Moreno, Ronald de Groot, A. Marceline Tutu van Furth, Michiel van der Flier, Navin P. Boeddha, Gertjan J. A. Driessen, Jan A. Hazelzet, Taco W. Kuijpers, Dasja Pajkrt, Elisabeth A. M. Sanders, Diederik van de Beek, A. van der Ende, H.L.A. Philipsen, A.O.A. Adeel, M.A. Breukels, D.M.C. Brinkman, C.C.M.M. de Korte, E. de Vries, W.J. de Waal, R. Dekkers, A. Dings-Lammertink, R.A. Doedens, A.E. Donker, M. Dousma, T.E. Faber, G.P.J.M. Gerrits, J.A.M. Gerver, J. Heidema, J. Homan-van der Veen, M.A.M. Jacobs, N.J.G. Jansen, P. Kawczynski, K. Klucovska, M.C.J. Kneyber, Y. Koopman-Keemink, V.J. Langenhorst, J. Leusink, B.F. Loza, I.T. Merth, C.J. Miedema, C. Neeleman, J.G. Noordzij, C.C. Obihara, A.L.T. van Overbeek – van Gils, G.H. Poortman, S.T. Potgieter, J. Potjewijd, P.P.R. Rosias, T. Sprong, G.W. ten Tussher, B.J. Thio, G.A. Tramper-Stranders, M. van Deuren, H. van der Meer, A.J.M. van Kuppevelt, A.M. van Wermeskerken, W.A. Verwijs, T.F.W. Wolfs, Luregn J Schlapbach, Philipp Agyeman, Christoph Aebi, Christoph Berger, Eric Giannoni, Martin Stocker, Klara M Posfay-Barbe, Ulrich Heininger, Sara Bernhard-Stirnemann, Anita Niederer-Loher, Christian Kahlert, Paul Hasters, Christa Relly, Walter Baer, Enitan Carrol, Stéphane Paulus, Hannah Frederick, Rebecca Jennings, Joanne Johnston, Rhian Kenwright, Colin G Fink, Elli Pinnock, Marieke Emonts, Rachel Agbeko, Suzanne Anderson, Fatou Secka, Kalifa Bojang, Isatou Sarr, Ngane Kebbeh, Gibbi Sey, Momodou Saidykhan, Fatoumatta Cole, Gilleh Thomas, Martin Antonio, Werner Zenz, Daniela S. Klobassa, Alexander Binder, Nina A. Schweintzger, Manfred Sagmeister, Hinrich Baumgart, Markus Baumgartner, Uta Behrends, Ariane Biebl, Robert Birnbacher, Jan-Gerd Blanke, Carsten Boelke, Kai Breuling, Jürgen Brunner, Maria Buller, Peter Dahlem, Beate Dietrich, Ernst Eber, Johannes Elias, Josef Emhofer, Rosa Etschmaier, Sebastian Farr, Ylenia Girtler, Irina Grigorow, Konrad Heimann, Ulrike Ihm, Zdenek Jaros, Hermann Kalhoff, Wilhelm Kaulfersch, Christoph Kemen, Nina Klocker, Bernhard Köster, Benno Kohlmaier, Eleni Komini, Lydia Kramer, Antje Neubert, Daniel Ortner, Lydia Pescollderungg, Klaus Pfurtscheller, Karl Reiter, Goran Ristic, Siegfried Rödl,

Andrea Sellner, Astrid Sonnleitner, Matthias Sperl, Wolfgang Stelzl, Holger Till, Andreas Trobisch, Anne Vierzig, Ulrich Vogel, Christina Weingarten, Stefanie Welke, Andreas Wimmer, Uwe Wintergerst, Daniel Wüller, Andrew Zaunschirm, Ieva Ziuraite, and Veslava Žukovskaja.

DIAMONDS consortium

Michael Levin, Lachlan Coin, Stuart Gormley, Shea Hamilton, Jethro Herberg, Bernardo Hourmat, Clive Hoggart, Myrsini Kaforou, Vanessa Sancho-Shimizu, Victoria Wright, Amina Abdulla, Paul Agapow, Maeve Bartlett, Evangelos Bellos, Hariklia Eleftherohorinou, Rachel Galassini, David Inwald, Meg Mashbat, Stefanie Menikou, Sobia Mustafa, Simon Nadel, Rahmeen Rahman, Clare Thakker, Sumit Bokhandi, Sue Power, Heather Barham, Nazima Pathan, Jenna Ridout, Deborah White, Sarah Thurston, Saul Faust, Sanjay Patel, Jenni McCorkell, Patrick Davies, Lindsey Crate, Helen Navarra, Stephanie Carter, Ramesh Ramaiah, Rekha Patel, Catherine Tuffrey, Andrew Gribbin, Sharon McCready, Mark Peters, Katie Hardy, Fran Standing, Lauran O'Neill, Eugenia Abelake, Akash Deep, Eniola Nsirim, Andrew Pollard, Louise Willis, Zoe Young, C. Royad, Sonia White, PM. Fortune, Phil Hudnott, Federico Martinón Torres, Fernando Álvez González, Ruth Barral-Arca, Miriam Cebey-López, María José Curras-Tuala, Natalia García, Luisa García Vicente, Alberto Gómez-Carballa, Jose Gómez Rial, Andrea Grela Beiroa, Antonio Justicia Grande, Pilar Leboráns Iglesias, Alba Elena Martínez Santos, Nazareth Martinón-Torres, José María Martinón Sánchez, Beatriz Morillo Gutiérrez, Belén Mosquera Pérez, Pablo Obando Pacheco, Jacobo Pardo-Seco, Sara Pischedda, Irene Rivero Calle, Carmen Rodríguez-Tenreiro, Lorenzo Redondo-Collazo, Sonia Serén Fernández, María del Sol Porto Silva, Ana Vega, Jose Manuel Fernández García, María Elena Gamborino Caramés, María Sol Rodríguez Calvo, Marta Aldonza Torres, Vanesa Álvarez Iglesias, Carmen Curros Novo, Isabel Ferreirós Vidal, Narmeen Mallah, Laura Navarro Ramón, Isabel Rego Lijo, Alba Camino Mera, Lúa Castelo Martínez, Ana Isabel Dacosta Urbieta, Wiktor Dominik Nowak, Julia García Currás, Nour El Zaharaa Mallah, Julián Montoto Louzao, Sara Rey Vázquez, Sandra Viz Lasheras, Patricia Regueiro Casuso, Miriam Taboada Puga, Lucía Vilanova Trillo, Susana Beatriz Reyes, María Cruz León León, Álvaro Navarro Mingorance, Xavier Gabaldó Barrios, Eider Oñate Vergara, Andrés Concha Torre, Ana Vivanco, Reyes Fernández, Francisco Giménez Sánchez, Miguel Sánchez Forte, Pablo Rojo, J. Ruiz Contreras, Alba Palacios, Cristina Epalza Ibarrondo, Elizabeth Fernandez Cooke, Marisa Navarro, Cristina Álvarez Álvarez, María José Lozano, Eduardo Carreras, Sonia Brió Sanagustín, Olaf Neth, M[a] del Carmen Martínez Padilla, Luis Manuel Prieto Tato, Sara Guillén, Laura Fernández Silveira, David Moreno, Ronald de Groot, A. Marceline Tutu van Furth, Michiel van der Flier, Navin P. Boeddha, Gertjan J. A. Driessen, Jan A. Hazelzet, Taco W. Kuijpers, Dasja Pajkrt, Elisabeth A. M. Sanders, Diederik van de Beek, A. van der Ende, H.L.A. Philipsen, A.O.A. Adeel, M.A. Breukels, D.M.C. Brinkman, C.C.M.M. de Korte, E. de Vries, W.J. de Waal, R. Dekkers, A. Dings-Lammertink, R.A. Doedens, A.E. Donker, M. Dousma, T.E. Faber, G.P.J.M. Gerrits, J.A.M. Gerver, J. Heidema, J. Homan-van der Veen, M.A.M. Jacobs, N.J.G. Jansen, P. Kawczynski, K. Klucovska, M.C.J. Kneyber, Y. Koopman-Keemink, V.J. Langenhorst, J. Leusink, B.F. Loza, I.T. Merth, C.J. Miedema, C. Neeleman, J.G. Noordzij, C.C. Obihara, A.L.T. van Overbeek – van Gils, G.H. Poortman, S.T. Potgieter, J. Potjewijd, P.P.R. Rosias, T. Sprong, G.W. ten Tussher, B.J. Thio, G.A. Tramper-Stranders, M. van Deuren, H. van der Meer, A.J.M. van Kuppevelt, A.M. van Wermeskerken, W.A. Verwijs, T.F.W. Wolfs, Luregn J Schlapbach, Philipp Agyeman, Christoph Aebi, Christoph Berger, Eric Giannoni, Martin Stocker, Klara M Posfay-Barbe, Ulrich Heininger, Sara Bernhard-Stirnemann, Anita Niederer-Loher, Christian Kahlert, Paul Hasters, Christa Relly, Walter Baer, Enitan Carrol, Stéphane Paulus, Hannah Frederick, Rebecca Jennings, Joanne Johnston, Rhian Kenwright, Colin G Fink, Elli Pinnock, Marieke Emonts, Rachel Agbeko, Suzanne Anderson, Fatou Secka, Kalifa Bojang, Isatou Sarr, Ngane Kebbeh, Gibbi Sey, Momodou Saidykhan, Fatoumatta Cole, Gilleh Thomas, Martin Antonio, Werner Zenz, Daniela S. Klobassa, Alexander Binder, Nina A. Schweintzger, Manfred Sagmeister, Hinrich Baumgart, Markus Baumgartner, Uta Behrends, Ariane Biebl, Robert Birnbacher, Jan-Gerd Blanke, Carsten Boelke, Kai Breuling, Jürgen Brunner, Maria Buller, Peter Dahlem, Beate Dietrich, Ernst Eber, Johannes Elias, Josef Emhofer, Rosa Etschmaier, Sebastian Farr, Ylenia Girtler, Irina Grigorow, Konrad Heimann, Ulrike Ihm, Zdenek Jaros, Hermann Kalhoff, Wilhelm Kaulfersch, Christoph Kemen, Nina Klocker, Bernhard Köster, Benno Kohlmaier, Eleni Komini, Lydia Kramer, Antje Neubert, Daniel Ortner, Lydia Pescollderungg, Klaus Pfurtscheller, Karl Reiter, Goran Ristic, Siegfried Rödl, Andrea Sellner, Astrid Sonnleitner, Matthias Sperl, Wolfgang Stelzl, Holger Till, Andreas Trobisch, Anne Vierzig, Ulrich Vogel, Christina Weingarten, Stefanie Welke, Andreas Wimmer, Uwe Wintergerst, Daniel Wüller, Andrew Zaunschirm, Ieva Ziuraite, and Veslava Žukovskaja.

PERFORM consortium

Michael Levin, Aubrey Cunnington, Tisham De, Jethro Herberg, Myrsini Kaforou, Victoria Wright, Lucas Baumard, Evangelos Bellos, Giselle D'Souza, Rachel Galassini, Dominic Habgood-Coote, Shea Hamilton, Clive Hoggart, Sara Hourmat, Heather Jackson, Ian Maconochie, Stephanie Menikou, Naomi Lin, Samuel Nichols, Ruud Nijman, Ivonne Pena Paz, Oliver Powell, Priyen Shah, Ching-Fen Shen, Clare Wilson, Amina Abdulla, Ladan Ali, Sarah Darnell, Rikke Jorgensen, Sobia Mustafa, Salina Persand, Molly Stevens, Eunjung Kim, Benjamin Pierce, Katy Fidler, Julia Dudley, Vivien Richmond, Emma Tavliavini, Ching-Chuan Liu, Shih-Min Wang, Federico Martinón-Torres, Antonio Salas, Fernando Álvez González, Cristina Balo Farto, Ruth Barral-Arca, María Barreiro Castro, Xabier Bello, Mirian Ben García, Sandra Carnota, Miriam Cebey-López, María José Curras-Tuala, Carlos Durán Suárez, Luisa García Vicente, Alberto Gómez-Carballa, Jose Gómez Rial, Pilar Leboráns Iglesias, Nazareth Martinón-Torres, José María Martinón Sánchez, Belén Mosquera Pérez, Jacobo Pardo-Seco, Lidia Piñeiro Rodríguez, Sara Pischedda, Sara Rey Vázquez, Irene Rivero Calle, Carmen Rodríguez-Tenreiro, Lorenzo Redondo-Collazo, Miguel Sadiki Ora, Sonia Serén Fernández, Cristina Serén Trasorras, Marisol Vilas Iglesias, Dace Zavadska, Anda Balode, Arta Bārzdiņa, Dārta Deksne, Dace Gardovska, Dagne Grāvele, Ilze Grope, Anija Meiere, Ieva Nokalna, Jana Pavāre, Zanda Pučuka, Katrīna Selecka, Aleksandra Sidorova, Dace Svile, Urzula Nora Urbāne,

Effua Usuf, Kalifa Bojang, Syed M. A. Zaman, Fatou Secka, Suzanne Anderson, Anna Rocalsatou Sarr, Momodou Saidykhan, Saffiatou Darboe, Samba Ceesay, Umberto D'alessandro, Henriëtte A. Moll, Dorine M. Borensztajn, Nienke N. Hagedoorn, Chantal Tan, Clementien L. Vermont, Joany Zachariasse, W Dik, Philipp Agyeman, Luregn J Schlapbach, Eric Giannoni, Martin Stocker, Klara M Posfay-Barbe, Ulrich Heininger, Sara Bernhard-Stirnemann, Anita Niederer-Loher, Christian Kahlert, Giancarlo Natalucci, Christa Relly, Thomas Riedel, Christoph Aebi, Christoph Berger, Enitan D Carrol, Stéphane Paulus, Elizabeth Cocklin, Rebecca Jennings, Joanne Johnston, Simon Leigh, Karen Newall, Sam Romaine, Andrew J. Pollard, Rama Kandasamy, Michael J. Carter, Daniel O'Connor, Sagida Bibi, Dominic F. Kelly, Meeru Gurung, Stephen Thorson, Imran Ansari, David R. Murdoch, Shrijana Shrestha, Marieke Emonts, Emma Lim, Lucille Valentine, Karen Allen, Kathryn Bell, Adora Chan, Stephen Crulley, Kirsty Devine, Daniel Fabian, Sharon King, Paul McAlinden, Sam McDonald, Anne McDonnell, Ailsa Pickering, Evelyn Thomson, Amanda Wood, Diane Wallia, Frances Baxter, Ashley Bell, Mathew Rhodes, Rachel Agbeko, Christine Mackerness, Bryan Baas, Lieke Kloosterhuis, Wilma Oosthoek, Tasnim Arif, Joshua Bennet, Kalvin Collings, Ilona van der Giessen, Alex Martin, Aqeela Rashid, Emily Rowlands, Gabriella de Vries, Fabian van der Velden, Mike Martin, Ravi Mistry, Ulrich von Both, Laura Kolberg, Manuela Zwerenz, Judith Buschbeck, Christoph Bidlingmaier, Vera Binder, Katharina Danhauser, Nikolaus Haas, Matthias Griese, Tobias Feuchtinger, Julia Keil, Matthias Kappler, Eberhard Lurz, Georg Muench, Karl Reiter, Carola Schoen, François Mallet, Karen Brengel-Pesce, Alexandre Pachot, Marine Mommert, Marko Pokorn, Mojca Kolnik, Katarina Vincek, Tina Plankar Srovin, Natalija Bahovec, Petra Prunk, Veronika Osterman, Tanja Avramoska, Taco Kuijpers, Ilse Jongerius, J. Merlijn van den Berg, Dieneke Schonenberg, Anouk M. Barendregt, Dasja Pajkrt, Martijn van der Kuip, A. Marceline Tutu van Furth, Evelien Sprenkeler, Judith Zandstra, Guido van Mierlo, Jan Geissler, Sebastián J. Vastert, Benjamin Evans, Jake Stevens, Peter Matthews, Kyle Billing, Werner Zenz, Alexander Binder, Benno Kohlmaier, Daniela S. Kohlfürst, Nina A. Schweintzger, Christoph Zurl, Susanne Hösele, Manuel Leitner, Lena Pölz, Glorija Rajic, Bianca Stoiser, Martina Strempfl, Manfred G. Sagmeister, Sebastian Bauchinger, Martin Benesch, Astrid Ceolotto, Ernst Eber, Siegfried Gallistl, Harald Haidl, Almuthe Hauer, Christa Hude, Andreas Kapper, Markus Keldorfer, Sabine Löffler, Tobias Niedrist, Heidemarie Pilch, Andreas Pfleger, Klaus Pfurtscheller, Siegfried Rödl, Andrea Skrabl-Baumgartner, Volker Strenger, Elmar Wallner, Dennie Tempel, Danielle van Keulen, Annelieke M Strijbosch, Maike K. Tauchert, Patricia Schmied, Irene Alba-Alejandre, Florian Hoffmann, Sabrina Juranek, Esther Maier, Karl Reiter, Sebastian Schroepf, Manuel Dewez, David Bath, Elizabeth Fitchett, and Fiona Cresswell.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Antonio Salas (antonio.salas@usc.es).

Materials availability

This study did not generate new unique reagents, and all materials utilized for the study are described in the STAR Methods section.

Data and code availability

- RNA-seq data supporting the findings of the present study are publicly available at Gene Expression Omnibus (GEO) under accession number GSE261482 and EBI ArrayExpress under accession number E-MTAB-11671,[105] E-MTAB-12793,[106] and E-MTAB-14564. Additional details are provided in Table S7.
- Original R code has been deposited at GitHub (https://github.com/Sandraviz/Pneumonia_Transcriptomic-Analysis).
- Any additional information required to analyze the data reported in this paper is available upon request.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to the study investigators of GENDRES network (www.gendres.org) (Annex), as well as the nursery and laboratory service at the Hospital Clínico Universitario de Santiago de Compostela, for their invaluable dedication and support. This research project was made possible through the access granted by the Galician Supercomputing Center (CESGA) to its supercomputing infrastructure. The supercomputer FinisTerrae III and its permanent data storage system have been funded by the Spanish Ministry of Science and Innovation, the Galician Government, and the European Regional Development Fund (ERDF). This work was supported by the European Seventh Framework Programme for Research and Technological Development (FP7) under the EUCLIDS project (grant agreement number 279185) and the European Union's Horizon 2020 research and innovation program under grant agreement nos. 668303 (PERFORM) and 848196 (DIAMONDS). This study also received support by (1) ISCIII: TRINEO: PI22/00162, DIAVIR: DTS19/00049, Resvi-Omics: PI19/01039 (to A.S.), ReSVinext: PI16/01569, Enterogen: PI19/01090, OMI-COVI-VAC: PI22/00406 (to F.M.-T.), and cofinanciados FEDER; (2) GAIN: IN607B 2020/08 and IN607A 2023/02 (to A.S.) and GEN-COVID (IN845D 2020/23 [to F.M.-T.], IIN607A2021/05 [to F.M.-T.], and IN677D 2024/06 [to A.G.-C.]; (3) ACIS: BI-BACVIR (PRIS-3, to A.S.) and CovidPhy (SA 304 C, to A.S.); (4) Spanish Ministry of Science and Innovation (MCIN)/Spanish Research Agency (AEI) (PID2022-142156OB-I00, to A.G.-C.); and (5) consorcio Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CB21/06/00103, to A.S. and F.M.-T.). A.G.-C. is supported by the Miguel Servet programme (CP23/00080) contract, funded by the Instituto de Salud Carlos III (ISCIII) and co-funded by the European Union. L.J.S. was supported by the NOMIS Foundation. The funders were not involved in the study design; collection, analysis, and interpretation of data; the writing of this article; or the decision to submit it for publication.

AUTHOR CONTRIBUTIONS

A.S., F.M.-T., A.G.-C., and S.V.-L. conceptualized and designed the study. A.S., F.M.-T., M.L., M.V.d.F., C.F., M.E., W.Z., A.J.P., T.T.K., S.A., M.P., H.A.M., and R.d.G. provided key resources. I.R.-C., A.I.D., J.A.H., A.J.C., and E.D.C. were responsible for clinical classification of the patients and collection of clinical metadata. C.L.V., U.v.B., D.S.K., M.T., D.Z., S.C.P., L.J.S., P.A., C.C., M.d.C.M.-P., A.P.-A., E.G.-S., J.V.-R., F.G.-S., P.A.-Q., and L.M.-G. contributed to patients' recruitment. V.J.W. coordinated sample processing and biobanking. S.V.-L., X.B., J.P.-S., A.G.-C., M.K., and D.H.-C. conducted bioinformatics and computational biology analysis. Computational models and machine learning approaches used to analyze the data were developed by A.S., S.V.-L., and A.G.-C.. A.S., S.V.-L., and A.G.-C. wrote the first draft of the manuscript. All authors contributed to discussions regarding the interpretation of results and reviewed and approved the final manuscript.

DECLARATION OF INTERESTS

The authors have a European patent application related to this work under the identification number EP24382240.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - RNA-seq analysis
  - Differential expression analysis
  - Pathway analysis
  - Signature discovery using LASSO
  - Signature validation
  - Weighted Gene Correlation Network Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.111747.

Received: February 26, 2024
Revised: October 24, 2024
Accepted: January 2, 2025
Published: January 4, 2025

REFERENCES

1. Vos, T., Lim, S.S., Abbafati, C., Abbas, K.M., Abbasi, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., and Abdelalim, A. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet *396*, 1204–1222.
2. World Health Organization (WHO) (2022). Pneumonia in children.
3. Haq, I.J., Battersby, A.C., Eastham, K., and McKean, M. (2017). Community acquired pneumonia in children. Br. Med. J. *356*, j686.
4. Shoar, S., and Musher, D.M. (2020). Etiology of community-acquired pneumonia in adults: a systematic review. Pneumonia *12*, 11.
5. Lynch, J.P., III. (2001). Hospital-acquired pneumonia: risk factors, microbiology, and treatment. Chest *119*, 373S–384S.
6. Torres, A., Catia, C., Niederman, M.S., Rosario, M., Chalmers, J.D., and Wunderink, R.G. (2021). Pneumonia (Primer). Nat. Rev. Dis. Prim. *7*, 25.
7. Jain, S., Williams, D.J., Arnold, S.R., Ampofo, K., Bramley, A.M., Reed, C., Stockmann, C., Anderson, E.J., Grijalva, C.G., Self, W.H., et al. (2015). Community-acquired pneumonia requiring hospitalization among US children. N. Engl. J. Med. *372*, 835–845.
8. Le Roux, D.M., and Zar, H.J. (2017). Community-acquired pneumonia in children—a changing spectrum of disease. Pediatr. Radiol. *47*, 1392–1398.
9. Schot, M.J.C., Dekker, A.R.J., Giorgi, W.G., Hopstaken, R.M., de Wit, N.J., Verheij, T.J.M., and Cals, J.W.L. (2018). Diagnostic value of signs, symptoms and diagnostic tests for diagnosing pneumonia in ambulant children in developed countries: a systematic review. NPJ Prim. Care Respir. Med. *28*, 40.
10. Rambaud-Althaus, C., Althaus, F., Genton, B., and D'Acremont, V. (2015). Clinical features for diagnosis of pneumonia in children younger than 5 years: a systematic review and meta-analysis. Lancet Infect. Dis. *15*, 439–450.
11. Scott, J.A.G., Wonodi, C., Moïsi, J.C., Deloria-Knoll, M., DeLuca, A.N., Karron, R.A., Bhat, N., Murdoch, D.R., Crawley, J., Levine, O.S., et al. (2012). The definition of pneumonia, the assessment of severity, and clinical standardization in the Pneumonia Etiology Research for Child Health study. Clin. Infect. Dis. *54*, S109–S116.
12. Eslamy, H.K., and Newman, B. (2011). Pneumonia in normal and immunocompromised children: an overview and update. Radiol. Clin. *49*, 895–920.
13. Prina, E., Ceccato, A., and Torres, A. (2016). New aspects in the management of pneumonia. Crit. Care *20*, 267.
14. Cillóniz, C., Torres, A., and Niederman, M.S. (2021). Management of pneumonia in critically ill patients. Br. Med. J. *375*, e065871.
15. Thomas, J., Pociute, A., Kevalas, R., Malinauskas, M., and Jankauskaite, L. (2020). Blood biomarkers differentiating viral versus bacterial pneumonia aetiology: a literature review. Ital. J. Pediatr. *46*, 4–10.
16. Christ-Crain, M., and Opal, S.M. (2010). Clinical review: the role of biomarkers in the diagnosis and management of community-acquired pneumonia. Crit. Care *14*, 203.
17. van Aerde, K.J., de Haan, L., van Leur, M., Gerrits, G.P., Schers, H., Moll, H.A., Hagedoorn, N.N., Herberg, J.A., Levin, M., Rivero-Calle, I., et al. (2021). Respiratory tract infection management and antibiotic prescription in children: a unique study comparing three levels of healthcare in the Netherlands. Pediatr. Infect. Dis. J. *40*, e100–e105.
18. Berti, E., Galli, L., de Martino, M., and Chiappini, E. (2013). International guidelines on tackling community-acquired pneumonia show major discrepancies between developed and developing countries. Acta Paediatr. *102*, 4–16.
19. Izadnegahdar, R., Cohen, A.L., Klugman, K.P., and Qazi, S.A. (2013). Childhood pneumonia in developing countries. Lancet Respir. Med. *1*, 574–584.
20. Lynch, T., Platt, R., Gouin, S., Larson, C., and Patenaude, Y. (2004). Can we predict which children with clinically suspected pneumonia will have the presence of focal infiltrates on chest radiographs? Pediatrics *113*, e186–e189.
21. Grief, S.N., and Loza, J.K. (2018). Guidelines for the Evaluation and Treatment of Pneumonia. Prim. Care *45*, 485–503.
22. Torres, A., Niederman, M.S., Chastre, J., Ewig, S., Fernandez-Vandellos, P., Hanberger, H., Kollef, M., Li Bassi, G., Luna, C.M., Martin-Loeches, I., et al. (2018). Summary of the international clinical guidelines for the management of hospital-acquired and ventilator-acquired pneumonia. ERJ Open Res. *4*, 00028-2018.
23. Cilloniz, C., Liapikou, A., and Torres, A. (2020). Advances in molecular diagnostic tests for pneumonia. Curr. Opin. Pulm. Med. *26*, 241–248.
24. Sungurlu, S., and Balk, R.A. (2018). The role of biomarkers in the diagnosis and management of pneumonia. Clin. Chest Med. *39*, 691–701.
25. Zar, H.J., Andronikou, S., and Nicol, M.P. (2017). Advances in the diagnosis of pneumonia in children. Br. Med. J. *358*, j2739.
26. Bhuiyan, M.U., Blyth, C.C., West, R., Lang, J., Rahman, T., Granland, C., de Gier, C., Borland, M.L., Thornton, R.B., Kirkham, L.-A.S., et al. (2019). Combination of clinical symptoms and blood biomarkers can improve discrimination between bacterial or viral community-acquired pneumonia in children. BMC Pulm. Med. *19*, 71–79.
27. Scicluna, B.P., Klein Klouwenberg, P.M.C., van Vught, L.A., Wiewel, M.A., Ong, D.S.Y., Zwinderman, A.H., Franitza, M., Toliat, M.R., Nürnberg, P., Hoogendijk, A.J., et al. (2015). A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. Am. J. Respir. Crit. Care Med. *192*, 826–835.
28. Verbakel, J.Y., Lemiengre, M.B., De Burghgraeve, T., De Sutter, A., Aertgeerts, B., Bullens, D.M.A., Shinkins, B., Van den Bruel, A., and Buntinx, F. (2018). Point-of-care C reactive protein to identify serious infection in acutely ill children presenting to hospital: prospective cohort study. Arch. Dis. Child. *103*, 420–426.
29. Eccles, S., Pincus, C., Higgins, B., and Woodhead, M.; Guideline Development Group (2014). Diagnosis and management of community and hospital acquired pneumonia in adults: summary of NICE guidance. Br. Med. J. *349*, g6722.
30. Herberg, J.A., Kaforou, M., Wright, V.J., Shailes, H., Eleftherohorinou, H., Hoggart, C.J., Cebey-López, M., Carter, M.J., Janes, V.A., Gormley, S.,

et al. (2016). Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children. JAMA *316*, 835–845.

31. Barral-Arca, R., Gómez-Carballa, A., Cebey-López, M., Currás-Tuala, M.J., Pischedda, S., Viz-Lasheras, S., Bello, X., Martinón-Torres, F., and Salas, A. (2020). Rna-seq data-mining allows the discovery of two long non-coding rna biomarkers of viral infection in humans. Int. J. Mol. Sci. *21*, 2748.
32. Mejias, A., Dimo, B., Suarez, N.M., Garcia, C., Suarez-Arrabal, M.C., Jartti, T., Blankenship, D., Jordan-Villegas, A., Ardura, M.I., Xu, Z., et al. (2013). Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. PLoS Med. *10*, e1001549.
33. Bloom, C.I., Graham, C.M., Berry, M.P.R., Rozakeas, F., Redford, P.S., Wang, Y., Xu, Z., Wilkinson, K.A., Wilkinson, R.J., Kendrick, Y., et al. (2013). Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. PLoS One *8*, e70630.
34. Barral-Arca, R., Pardo-Seco, J., Martinón-Torres, F., and Salas, A. (2018). A 2-transcript host cell signature distinguishes viral from bacterial diarrhea and it is influenced by the severity of symptoms. Sci. Rep. *8*, 8043.
35. Gómez-Carballa, A., Cebey-López, M., Pardo-Seco, J., Barral-Arca, R., Rivero-Calle, I., Pischedda, S., Currás-Tuala, M.J., Gómez-Rial, J., Barros, F., and Martinón-Torres, F. (2019). A qPCR expression assay of IFI44L gene differentiates viral from bacterial infections in febrile children. Sci. Rep. *9*, 1–12.
36. Kaforou, M., Wright, V.J., Oni, T., French, N., Anderson, S.T., Bangani, N., Banwell, C.M., Brent, A.J., Crampin, A.C., Dockrell, H.M., et al. (2013). Detection of tuberculosis in HIV-infected and-uninfected African adults using whole blood RNA expression signatures: a case-control study. PLoS Med. *10*, e1001538.
37. Sweeney, T.E., Braviak, L., Tato, C.M., and Khatri, P. (2016). Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir. Med. *4*, 213–224.
38. Anderson, S.T., Kaforou, M., Brent, A.J., Wright, V.J., Banwell, C.M., Chagaluka, G., Crampin, A.C., Dockrell, H.M., French, N., Hamilton, M.S., et al. (2014). Diagnosis of childhood tuberculosis and host RNA expression in Africa. N. Engl. J. Med. *370*, 1712–1723.
39. Herberg, J.A., Kaforou, M., Gormley, S., Sumner, E.R., Patel, S., Jones, K.D.J., Paulus, S., Fink, C., Martinon-Torres, F., Montana, G., et al. (2013). Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. J. Infect. Dis. *208*, 1664–1668.
40. Barral-Arca, R., Gómez-Carballa, A., Cebey-López, M., Bello, X., Martinón-Torres, F., and Salas, A. (2020). A meta-analysis of multiple whole blood gene expression data unveils a diagnostic host-response transcript signature for respiratory syncytial virus. Int. J. Mol. Sci. *21*, 1831.
41. DeBerg, H.A., Zaidi, M.B., Altman, M.C., Khaenam, P., Gersuk, V.H., Campos, F.D., Perez-Martinez, I., Meza-Segura, M., Chaussabel, D., Banchereau, J., et al. (2018). Shared and organism-specific host responses to childhood diarrheal diseases revealed by whole blood transcript profiling. PLoS One *13*, e0192082.
42. Mahajan, P., Kuppermann, N., Suarez, N., Mejias, A., Casper, C., Dean, J.M., and Ramilo, O.; Febrile Infant Working Group for the Pediatric Emergency Care Applied Research Network PECARN (2015). RNA transcriptional biosignature analysis for identifying febrile infants with serious bacterial infections in the emergency department: a feasibility study. Pediatr. Emerg. Care *31*, 1–5.
43. Gómez-Carballa, A., Barral-Arca, R., Cebey-López, M., Bello, X., Pardo-Seco, J., Martinón-Torres, F., and Salas, A. (2021). Identification of a minimal 3-transcript signature to differentiate viral from bacterial infection from best genome-wide host RNA biomarkers: a multi-cohort analysis. Int. J. Mol. Sci. *22*, 3148.
44. Parnell, G.P., McLean, A.S., Booth, D.R., Armstrong, N.J., Nalos, M., Huang, S.J., Manak, J., Tang, W., Tam, O.-Y., Chan, S., and Tang, B.M. (2012). A distinct influenza infection signature in the blood transcriptome of patients with severe community-acquired pneumonia. Crit. Care *16*, R157.
45. Viasus, D., Simonetti, A.F., Nonell, L., Vidal, O., Meije, Y., Ortega, L., Arnal, M., Bódalo-Torruella, M., Sierra, M., Rombauts, A., et al. (2023). Whole-Blood Gene Expression Profiles Associated with Mortality in Community-Acquired Pneumonia. Biomedicines *11*, 429.
46. Zhao, J., He, X., Min, J., Yao, R.S.Y., Chen, Y., Chen, Z., Huang, Y., Zhu, Z., Gong, Y., Xie, Y., et al. (2023). A multicenter prospective study of comprehensive metagenomic and transcriptomic signatures for predicting outcomes of patients with severe community-acquired pneumonia. EBioMedicine *96*, 104790.
47. Hopp, L., Loeffler-Wirth, H., Nersisyan, L., Arakelyan, A., and Binder, H. (2018). Footprints of sepsis framed within community acquired pneumonia in the blood transcriptome. Front. Immunol. *9*, 1620.
48. Anahtar, M., Chan, L.W., Ko, H., Rao, A., Soleimany, A.P., Khatri, P., and Bhatia, S.N. (2022). Host protease activity classifies pneumonia etiology. Proc. Natl. Acad. Sci. USA *119*, e2121778119.
49. Siljan, W.W., Sivakumaran, D., Ritz, C., Jenum, S., Ottenhoff, T.H., Ulvestad, E., Holter, J.C., Heggelund, L., and Grewal, H.M. (2022). Host transcriptional signatures predict etiology in community-acquired pneumonia: Potential antibiotic stewardship tools. Biomark. Insights *17*, 11772719221099130.
50. Principi, N., and Esposito, S. (2017). Biomarkers in pediatric community-acquired pneumonia. Int. J. Mol. Sci. *18*, 447.
51. Virkki, R., Juven, T., Rikalainen, H., Svedström, E., Mertsola, J., and Ruuskanen, O. (2002). Differentiation of bacterial and viral pneumonia in children. Thorax *57*, 438–441.
52. Silterra, J., Gillette, M.A., Lanaspa, M., Pellé, K.G., Valim, C., Ahmad, R., Acácio, S., Almendinger, K.D., Tan, Y., Madrid, L., et al. (2017). Transcriptional categorization of the etiology of pneumonia syndrome in pediatric patients in malaria-endemic areas. J. Infect. Dis. *215*, 312–320.
53. Wallihan, R.G., Suárez, N.M., Cohen, D.M., Marcon, M., Moore-Clingenpeel, M., Mejias, A., and Ramilo, O. (2018). Molecular distance to health transcriptional score and disease severity in children hospitalized with community-acquired pneumonia. Front. Cell. Infect. Microbiol. *8*, 382.
54. Kolberg, L., Khanijau, A., van der Velden, F.J., Herberg, J., De, T., Galassini, R., Cunnington, A.J., Wright, V.J., Shah, P., and Kaforou, M. (2023). Raising AWaRe-ness of antimicrobial stewardship challenges in pediatric emergency care: results from the PERFORM study assessing consistency and appropriateness of antibiotic prescribing across Europe. Clin. Infect. Dis. *78*, 526–534.
55. Juss, J.K., House, D., Amour, A., Begg, M., Herre, J., Storisteanu, D.M.L., Hoenderdos, K., Bradley, G., Lennon, M., Summers, C., et al. (2016). Acute respiratory distress syndrome neutrophils have a distinct phenotype and are resistant to phosphoinositide 3-kinase inhibition. Am. J. Respir. Crit. Care Med. *194*, 961–973.
56. Potey, P.M., Rossi, A.G., Lucas, C.D., and Dorward, D.A. (2019). Neutrophils in the initiation and resolution of acute pulmonary inflammation: understanding biological function and therapeutic potential. J. Pathol. *247*, 672–685.
57. Soehnlein, O., Oehmcke, S., Rothfuchs, A., Rothfuchs, A.G., Frithiof, R., Mörgelin, M., Mörgelin, M., Herwald, H., and Lindbom, L. (2008). Neutrophil degranulation mediates severe lung damage triggered by streptococcal M1 protein. Eur. Respir. J. *32*, 405–412.
58. Sande, C.J., Njunge, J.M., Mwongeli Ngoi, J., Mutunga, M.N., Chege, T., Gicheru, E.T., Gardiner, E.M., Gwela, A., Green, C.A., Drysdale, S.B., et al. (2019). Airway response to respiratory syncytial virus has incidental antibacterial effects. Nat. Commun. *10*, 2218.

59. Yao, Y., Zhao, J., Hu, J., Song, H., Wang, S., and Wang, Y. (2022). Identification of a Four-Gene Signature for Diagnosing Paediatric Sepsis. BioMed Res. Int. *2022*, 5217885.

60. Lu, J., Li, Q., Wu, Z., Zhong, Z., Ji, P., Li, H., He, C., Feng, J., and Zhang, J. (2020). Two gene set variation indexes as potential diagnostic tool for sepsis. Am. J. Transl. Res. *12*, 2749–2759.

61. Banerjee, S., Mohammed, A., Wong, H.R., Palaniyar, N., and Kamaleswaran, R. (2021). 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. *12*, 592303.

62. Cabral-Pacheco, G.A., Garza-Veloz, I., Castruita-De la Rosa, C., Ramirez-Acuña, J.M., Perez-Romero, B.A., Guerrero-Rodriguez, J.F., Martinez-Avila, N., and Martinez-Fierro, M.L. (2020). The roles of matrix metalloproteinases and their inhibitors in human diseases. Int. J. Mol. Sci. *21*, 9739.

63. Uhel, F., Scicluna, B.P., van Vught, L.A., Cremer, O.L., Bonten, M.J., Schultz, M.J., and van der Poll, T. (2019). Matrix metalloproteinase-8: a useful biomarker to refine the diagnosis of community-acquired pneumonia upon intensive care unit admission? Crit. Care *23*, 226.

64. Lauhio, A., Hästbacka, J., Pettilä, V., Tervahartiala, T., Karlsson, S., Varpula, T., Varpula, M., Ruokonen, E., Sorsa, T., and Kolho, E. (2011). Serum MMP-8,-9 and TIMP-1 in sepsis: high serum levels of MMP-8 and TIMP-1 are associated with fatal outcome in a multicentre, prospective cohort study. Hypothetical impact of tetracyclines. Pharmacol. Res. *64*, 590–594.

65. Fang, X., Duan, S.-F., Hu, Z.-Y., Wang, J.-J., Qiu, L., Wang, F., and Chen, X.-L. (2022). Inhibition of Matrix Metalloproteinase-8 Protects Against Sepsis Serum Mediated Leukocyte Adhesion. Front. Med. *9*, 814890.

66. Herberg, J., Huang, H., Thezenas, M.L., Janes, V., Carter, M., Gormley, S., Hamilton, M.S., Kessler, B., Levin, M., and Casals-Pascual, C. (2019). Lipocalin-2 is a sensitive and specific marker of bacterial infection in children. Preprint at bioRxiv. https://doi.org/10.1016/j.cmi.2020.07.006.

67. Liu, W., and Rodgers, G.P. (2022). Olfactomedin 4 is a biomarker for the severity of infectious diseases (Oxford University Press US), p. ofac061.

68. Alder, M.N., Opoka, A.M., Lahni, P., Hildeman, D.A., and Wong, H.R. (2017). Olfactomedin 4 is a candidate marker for a pathogenic neutrophil subset in septic shock. Crit. Care Med. *45*, e426–e432.

69. Demaret, J., Venet, F., Plassais, J., Cazalis, M.-A., Vallin, H., Friggeri, A., Lepape, A., Rimmelé, T., Textoris, J., and Monneret, G. (2016). Identification of CD177 as the most dysregulated parameter in a microarray study of purified neutrophils from septic shock patients. Immunol. Lett. *178*, 122–130.

70. Levy, Y., Wiedemann, A., Hejblum, B.P., Durand, M., Lefebvre, C., Surenaud, M., Lacabaratz, C., Perreau, M., Foucat, E., and Dechenaud, M. (2021). CD177, a specific marker of neutrophil activation, is associated with coronavirus disease 2019 severity and death. iScience *24*, 102711.

71. Wang, Z., Zhang, L., Li, S., Xu, F., Han, D., Wang, H., Huang, T., Yin, H., and Lyu, J. (2022). The relationship between hematocrit and serum albumin levels difference and mortality in elderly sepsis patients in intensive care units—a retrospective study based on two large database. BMC Infect. Dis. *22*, 629.

72. Takegawa, R., Kabata, D., Shimizu, K., Hisano, S., Ogura, H., Shintani, A., and Shimazu, T. (2019). Serum albumin as a risk factor for death in patients with prolonged sepsis: an observational study. J. Crit. Care *51*, 139–144.

73. Dai, D.-M., Wang, D., Hu, D., Wan, W.-L., Su, Y., Yang, J.-L., Wang, Y.-P., Wang, F., Yang, L., Sun, H.-M., et al. (2020). Difference in hematocrit and plasma albumin levels as an additional biomarker in the diagnosis of infectious disease. Arch. Med. Sci. *16*, 522–530.

74. Turcato, G., Zaboli, A., Kostic, I., Melchioretto, B., Ciccariello, L., Zaccaria, E., Olivato, A., Maccagnani, A., Pfeifer, N., and Bonora, A. (2022). Severity of SARS-CoV-2 infection and albumin levels recorded at the first emergency department evaluation: a multicentre retrospective observational study. Emerg. Med. J. *39*, 63–69.

75. Chen, C., Zhang, Y., Zhao, X., Tao, M., Yan, W., and Fu, Y. (2021). Hypoalbuminemia–An indicator of the severity and prognosis of COVID-19 patients: a multicentre retrospective analysis. Infect. Drug Resist. *14*, 3699–3710.

76. Cheng, J., Ji, D., Yin, Y., Wang, S., Song, K., Pan, Q., Zhang, Q., and Yang, L. (2022). Proteomic profiling of serum small extracellular vesicles reveals immune signatures of children with pneumonia. Transl. Pediatr. *11*, 891–908.

77. Kellum, J.A., Kong, L., Fink, M.P., Weissfeld, L.A., Yealy, D.M., Pinsky, M.R., Fine, J., Krichevsky, A., Delude, R.L., and Angus, D.C.; GenIMS Investigators (2007). Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the Genetic and Inflammatory Markers of Sepsis (GenIMS) Study. Arch. Intern. Med. *167*, 1655–1663.

78. Nguyen Thi Dieu, T., Pham Nhat, A., Craig, T.J., and Duong-Quy, S. (2017). Clinical characteristics and cytokine changes in children with pneumonia requiring mechanical ventilation. J. Int. Med. Res. *45*, 1805–1817.

79. Feikin, D.R., Hammitt, L.L., Murdoch, D.R., O'Brien, K.L., and Scott, J.A.G. (2017). The enduring challenge of determining pneumonia etiology in children: considerations for future research priorities. Clin. Infect. Dis. *64*, S188–S196.

80. Murdoch, D.R., O'Brien, K.L., Driscoll, A.J., Karron, R.A., and Bhat, N.Pneumonia Methods Working Group; PERCH Core Team (2012). Laboratory methods for determining pneumonia etiology in children. Clin. Infect. Dis. *54*, S146–S152.

81. Shah, P., Voice, M., Calvo-Bado, L., Rivero-Calle, I., Morris, S., Nijman, R., Broderick, C., De, T., Eleftheriou, I., Galassini, R., et al. (2023). Relationship between molecular pathogen detection and clinical disease in febrile children across Europe: a multicentre, prospective observational study. Lancet Reg. Health. Eur. *32*, 100682.

82. Arora, S., Ahmad, S., Irshad, R., Goyal, Y., Rafat, S., Siddiqui, N., Dev, K., Husain, M., Ali, S., Mohan, A., and Syed, M.A. (2019). TLRs in pulmonary diseases. Life Sci. *233*, 116671.

83. Chollet-Martin, S., Montravers, P., Gibert, C., Elbim, C., Desmonts, J.M., Fagon, J.Y., and Gougerot-Pocidalo, M.A. (1992). Subpopulation of hyperresponsive polymorphonuclear neutrophils in patients with adult respiratory distress syndrome. Am. Rev. Respir. Dis. *146*, 990–996.

84. Fitzgerald, K.A., and Kagan, J.C. (2020). Toll-like receptors and the control of immunity. Cell *180*, 1044–1066.

85. Medvedev, A.E., Lentschat, A., Kuhns, D.B., Blanco, J.C.G., Salkowski, C., Zhang, S., Arditi, M., Gallin, J.I., and Vogel, S.N. (2003). Distinct mutations in IRAK-4 confer hyporesponsiveness to lipopolysaccharide and interleukin-1 in a patient with recurrent bacterial infections. J. Exp. Med. *198*, 521–531.

86. Ku, C.-L., Picard, C., Erdös, M., Jeurissen, A., Bustamante, J., Puel, A., von Bernuth, H., Filipe-Santos, O., Chang, H.-H., Lawrence, T., et al. (2007). IRAK4 and NEMO mutations in otherwise healthy children with recurrent invasive pneumococcal disease. J. Med. Genet. *44*, 16–23.

87. Picard, C., Casanova, J.-L., and Puel, A. (2011). Infectious diseases in patients with IRAK-4, MyD88, NEMO, or IκBα deficiency. Clin. Microbiol. Rev. *24*, 490–497.

88. Von Bernuth, H., Picard, C., Jin, Z., Pankla, R., Xiao, H., Ku, C.-L., Chrabieh, M., Mustapha, I.B., Ghandil, P., Camcioglu, Y., et al. (2008). Pyogenic bacterial infections in humans with MyD88 deficiency. Science *321*, 691–696.

89. Xu, X., Yuan, B., Liang, Q., Huang, H., Yin, X., Sheng, X., Nie, N., and Fang, H. (2015). Gene expression profile analysis of ventilator-associated pneumonia. Mol. Med. Rep. *12*, 7455–7462.

90. Jagadeesh, A., Prathyusha, A., Sheela, G.M., and Bramhachari, P.V. (2020). T cells in viral infections: the myriad flavours of antiviral immunity.

Dynam. Immune Activation Viral Dis., 139–148. https://doi.org/10.1007/978-981-15-1045-8_9.

91. Kervevan, J., and Chakrabarti, L.A. (2021). Role of CD4+ T cells in the control of viral infections: Recent advances and open questions. Int. J. Mol. Sci. *22*, 523.
92. Schmidt, M.E., and Varga, S.M. (2018). The CD8 T cell response to respiratory virus infections. Front. Immunol. *9*, 678.
93. Swain, S.L., McKinstry, K.K., and Strutt, T.M. (2012). Expanding roles for CD4+ T cells in immunity to viruses. Nat. Rev. Immunol. *12*, 136–148.
94. Santiago, R.P., Figueiredo, C.V.B., Fiuza, L.M., Yahouédéhou, S.C.M.A., Oliveira, R.M., Aleluia, M.M., Carvalho, S.P., Fonseca, C.A., Nascimento, V.M.L., Rocha, L.C., et al. (2020). Transforming growth factor beta receptor 3 haplotypes in sickle cell disease are associated with lipid profile and clinical manifestations. Mediat. Inflamm. *2020*, 3185015.
95. Ning, J., Sun, K., Wang, X., Fan, X., Jia, K., Cui, J., and Ma, C. (2023). Use of machine learning-based integration to develop a monocyte differentiation-related signature for improving prognosis in patients with sepsis. Mol. Med. *29*, 37.
96. Basu, S., Naha, A., Veeraraghavan, B., Ramaiah, S., and Anbarasu, A. (2022). In silico structure evaluation of BAG3 and elucidating its association with bacterial infections through protein–protein and host-pathogen interaction analysis. J. Cell. Biochem. *123*, 115–127.
97. Nalbant, D., Youn, H., Nalbant, S.I., Sharma, S., Cobos, E., Beale, E.G., Du, Y., and Williams, S.C. (2005). FAM20: an evolutionarily conserved family of secreted proteins expressed in hematopoietic cells. BMC Genom. *6*, 11–21.
98. Lu, X., Xue, L., Sun, W., Ye, J., Zhu, Z., and Mei, H. (2018). Identification of key pathogenic genes of sepsis based on the Gene Expression Omnibus database. Mol. Med. Rep. *17*, 3042–3054.
99. Davenport, E.E., Burnham, K.L., Radhakrishnan, J., Humburg, P., Hutton, P., Mills, T.C., Rautanen, A., Gordon, A.C., Garrard, C., Hill, A.V.S., et al. (2016). Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. Lancet Respir. Med. *4*, 259–271.
100. Jia, C., Zhang, F., Zhu, Y., Qi, X., and Wang, Y. (2017). Public data mining plus domestic experimental study defined involvement of the old-yet-uncharacterized gene matrix-remodeling associated 7 (MXRA7) in physiopathology of the eye. Gene *632*, 43–49.
101. Wang, H., Guo, Y., Lu, H., Luo, Y., Hu, W., Liang, W., Garcia-Barrio, M.T., Chang, L., Schwendeman, A., Zhang, J., and Chen, Y.E. (2022). Krüppel-like factor 14 deletion in myeloid cells accelerates atherosclerotic lesion development. Cardiovasc. Res. *118*, 475–488.
102. Yuan, Y., Fan, G., Liu, Y., Liu, L., Zhang, T., Liu, P., Tu, Q., Zhang, X., Luo, S., Yao, L., et al. (2022). The transcription factor KLF14 regulates macrophage glycolysis and immune function by inhibiting HK2 in sepsis. Cell. Mol. Immunol. *19*, 504–515.
103. Larcher, R., Boudet, A., Roger, C., Villa, F., and Loubet, P. (2024). Mycoplasma pneumoniae is back! Is it the next pandemic? Anaesth. Crit. Care Pain Med. *43*, 101338.
104. Sauteur, P.M.M., Beeton, M.L., Pereyre, S., Bébéar, C., Gardette, M., Hénin, N., Wagner, N., Fischer, A., Vitale, A., and Lemaire, B. (2023). Mycoplasma pneumoniae: delayed re-emergence after COVID-19 pandemic restrictions. Lancet Microbe *5*, e100.
105. Habgood-Coote, D., Wilson, C., Shimizu, C., Barendregt, A.M., Philipsen, R., Galassini, R., Calle, I.R., Workman, L., Agyeman, P.K.A., Ferwerda, G., et al. (2023). Diagnosis of childhood febrile illness using a multi-class blood RNA molecular signature. Med *4*, 635–654. https://doi.org/10.1016/j.medj.2023.06.007.
106. Jackson, H.R., Miglietta, L., Habgood-Coote, D., D'Souza, G., Shah, P., Nichols, S., Vito, O., Powell, O., Davidson, M.S., Shimizu, C., et al. (2023). Diagnosis of Multisystem Inflammatory Syndrome in Children by a Whole-Blood Transcriptional Signature. J. Pediatric Infect. Dis. Soc. *12*, 322–331. https://doi.org/10.1093/jpids/piad035.
107. Martinón-Torres, F., Salas, A., Rivero-Calle, I., Cebey-López, M., Pardo-Seco, J., Herberg, J.A., Boeddha, N.P., Klobassa, D.S., Secka, F., Paulus, S., et al. (2018). Life-threatening infections in children in Europe (the EUCLIDS Project): a prospective cohort study. Lancet. Child Adolesc. Health *2*, 404–414.
108. Habgood-Coote, D., Wilson, C., Shimizu, C., Barendregt, A.M., Philipsen, R., Galassini, R., Calle, I.R., Workman, L., Agyeman, P.K.A., Ferwerda, G., et al. (2023). Diagnosis of childhood febrile illness using a multi-class blood RNA molecular signature. Med *4*, 635–654.
109. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data.
110. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
111. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.
112. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
113. Zhang, Y., Parmigiani, G., and Johnson, W.E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom. Bioinform. *2*, lqaa078. https://doi.org/10.1093/nargab/lqaa078.
114. Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.
115. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.
116. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.
117. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS A J. Integr. Biol. *16*, 284–287.
118. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Software *33*, 1–22.
119. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. *12*, 77–78.
120. López-Ratón, M., Rodríguez-Álvarez, M.X., Cadarso-Suárez, C., and Gude-Sampedro, F. (2014). OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. J. Stat. Software *61*, 1–36.
121. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human whole blood samples	This paper	Available in Table S7
Critical commercial assays		
TruSeq RNA Sample Preparation Kit	Illumina	# RS-122-2001
Illumina® Ribo-Zero Gold kit	Illumina	# 20037135
Deposited data		
RNA-Seq data	This paper	GEO database: GSE261482
RNA-Seq data	This paper	ArrayExpress database: E-MTAB-14564
RNA-Seq data	Jackson et al.[106]	ArrayExpress database: E-MTAB-12793
RNA-Seq data	Habgood-Coote et al.[105]	ArrayExpress database: E-MTAB-11671
Microarray data	Wallihan et al.[68]	GEO accession number: GSE103119
R code	This paper	https://github.com/Sandraviz/Pneumonia_Transcriptomic-Analysis
Software and algorithms		
R software	https://www.r-project.org/	Version 4.2.0
FastQC	Andrews et al.[55]	https://github.com/s-andrews/FastQC
STAR	Dobin et al.[56]	https://github.com/alexdobin/STAR
FeatureCounts	Liao et al.[58]	https://www.bioconductor.org/packages/release/bioc/html/Rsubread.html
DESeq2	Love et al.[59]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
RUVSeq	Risso et al.[60]	https://bioconductor.org/packages/release/bioc/html/RUVSeq.html
limma	Ritchie et al.[62]	https://bioconductor.org/packages/release/bioc/html/limma.html
Clusterprofiler	Yu et al.[63]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
glmnet	Friedman et al.[64]	https://cran.r-project.org/web/packages/glmnet/index.html
pROC	Robin et al.[65]	https://cran.r-project.org/web/packages/pROC/index.html
OptimalCutPoints	López-Ratón et al.[66]	https://cran.r-project.org/web/packages/OptimalCutpoints/index.html
WGCNA	Langfelder et al.[67]	https://cran.r-project.org/web/packages/WGCNA/index.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study utilized blood samples from paediatric patients recruited in the European Union Childhood Life-threatening Infectious Diseases Study (EUCLIDS- https://www.euclids-project.eu/).[107] The EUCLIDS study enrolled children and young adolescents aged 1 month to 18 years who were admitted to hospitals with suspected sepsis or severe focal infection. The study encompassed a network of 194 hospitals across nine European countries, as well as two hospitals located in The Gambia, Africa. Patient recruitment occurred between July 1, 2012, and December 31, 2015.

Pneumonia phenotype was defined as clinical symptoms compatible with acute respiratory infection and inflammation of one or both lungs with lobar or segmental or multilobar collapse/consolidation on chest X-ray. More specifically, following radiological findings of consolidation/pleural effusion must be met: alveolar consolidation (defined as a dense or fluffy opacity that occupies a portion or whole of a lobe or of the entire lung that may or may not contain air-bronchograms) or pleural effusion (defined as fluid

in the lateral pleural space and not just in the minor or oblique fissure) that was spatially associated with a pulmonary parenchymal infiltrate (including other infiltrate) or obliterated enough of the hemithorax to obscure an opacity. It does not include just perihilar consolidation or patchy consolidation. Children with pneumonia (n = 154) and age and sex matched healthy controls (n = 38; Table 1) were selected from the EUCLIDS database[108] for transcriptome analysis in whole blood samples collected in PAXgene™ and Tempus™ RNA tubes. In a subsequent analysis, pneumonia samples were further classified using the algorithm developed by Herberg et al.[30] C-reactive protein (CRP) and neutrophil threshold criteria were employed where available, with 72.5% and 33.3% of missing CRP values in definitive bacterial (DB) and definitive viral (DV), respectively (see[30] for details on clinical characteristics of the cohort and definitions on DB and DV). Thus, DB (n = 40) and DV (n = 9) pneumonias were selected to study the differences in transcriptome depending on the etiology of the infection. Pneumonias categorized as Probable Bacterial (PB) and Probable Viral (PV) were excluded from this specific analysis to minimize possible confounding factors in stablishing a transcriptomic signature for differentiating bacterial from viral pneumonia in children but were used post-hoc to test the signature in a more realistic clinical scenario (Figure 1).

In order to validate the proposed signature, a new independent cohort comprising blood samples obtained from children with viral (n = 10) and bacterial (n = 22) pneumonias was employed (Figure 1). Clinical and demographic information are provided in Table 2. These samples were recruited by the PErsonalised Risk assessment in Febrile illness to Optimise Real-life Management across the European Union (PERFORM - https://www.perform2020.org/) and the Diagnosis and Management of Febrile Illness using RNA Personalised Molecular Signature Diagnosis (DIAMONDS - https://www.diamonds2020.eu) consortiums.

All newly generated data from human patients were obtained through harmonized procedures for patient recruitment, classification, clinical data collection, and sample acquisition, processing, and storage across the participating centers. All necessary ethical approvals were obtained from each participating country's Ethics Committee (EC).

METHOD DETAILS

RNA-seq analysis

For the discovery cohort, RNA was isolated from blood samples collected in PAXgene™ and in Tempus™ RNA tubes. RNA sequencing was conducted on a HiSeq 4000 (Illumina) platform, with library preparation and sequencing of 30 million 75 or 100 bp paired-end reads. The Illumina's TruSeq RNA Sample Preparation Kit was used for library preparation, and ribosomal and globin RNA depletion was performed using the Illumina® Ribo-Zero Gold kit. For the validation cohort, RNA was isolated using PAXgene blood miRNA isolation kit according to the manufacturer's instructions (Qiagen). A DNAse treatment was carried out with the RNA clean & concentrator kit (Zymo Research) prior to sequencing. RNA was quantified using RiboGreen (Invitrogen) on the FLUOstar OPTIMA plate reader (BMG Labtech) and the integrity was analyzed on the TapeStation 2200(Agilent, RNA ScreenTape). After a normalization step, a strand specific library preparation was completed using NEBNext® Ultra™ II mRNA kit (NEB) and NEB rRNA/globin depletion probes following manufacturer's recommendations. Individual libraries were normalized using Qubit, pooled together and diluted. The sequencing was performed using a 150 paired-end configuration in a Novaseq6000 platform.

Quality control of all raw data was carried out using *FastQC*,[109] alignment and read counting were performed using *STAR*,[110] alignment filtering was done with *SAMtools*[111] and read counting was carried out using *FeatureCounts*.[112]

Differential expression analysis

RNAseq data was processed for batch correction using control samples and *COMBAT-Seq* R package.[113] The *RUVg* method implemented in the *RUVSeq* package of R[114] was employed to identified and adjust for other factors of undesirable variation present in the dataset, selecting a $k = 2$ as a number of variables to include. After that, data normalization, differential expression (DE) analysis and covariates correction were carried out using *DESeq* package.[115] Differentially expressed genes (DEGs) were identified through the Negative Binomial distribution implemented in the *DESeq* package. The fitted model included known covariates such as sex, batch, and type of collection tube as well as the factors of undesirable variation, to correct the differences in gene expression related to these factors. A generalized linear model was fitted, and a t-statistic was calculated for each gene. P values were corrected for multiple testing using the Benjamini-Hochberg False Discovery Rate (FDR).

As a validation cohort for the pneumonia *vs*. control study, we downloaded the GSE103119 microarray dataset from the Gene Expression Omnibus (GEO) database. This dataset comprises whole blood gene expression data from children with pneumonia (n = 152) as well as healthy controls (n = 20). Data were normalized, processed and analyzed using the *limma* R package.[116] A linear model was fitted, considering the sex as a categorical covariate. Multiple testing correction was performed using the FDR.

Principal Component Analysis (PCA) was used to visualize the global transcriptome patterns of both RNA-seq and microarray data. A Spearman test was computed for the correlation indices (ρ) and the associated p values. Wilcoxon test was performed to assess statistical significance between groups.

Pathway analysis

We used the Reactome and GO (Gene Ontology) databases as references to examine biological pathways and processes associated with the DEGs in pneumonia patients. We followed both an over-representation analysis (ORA) and a gene-set enrichment (GSEA) approaches, considering DEGs with a $\log_2 FC \geq |1.5|$ and an adjusted p value < 0.05. The analysis was performed using the

Clusterprofiler[117] R package. To account for multiple test correction, we applied the FDR procedure, establishing a significant threshold of 0.05.

Signature discovery using LASSO

We employed a Least Absolute Shrinkage and Selection Operator (LASSO) regression model to identify a subset of genes that could serve as a predictive transcriptomic signature differentiating viral from bacterial pneumonia in the RNA-seq cohort. A predictive transcriptomic signature was computed using the R package *glmnet*[118] and the *cv.glmnet* function. A logistic LASSO regression model was fitted with the alpha parameter set to 1. We selected 36 DEGs filtered based on $|\log_2 FC| > 1.5$, adjusted p value < 0.05, and a Base Mean > 50 as input for the model. To determine the optimal parameter selection for the LASSO regression model, we conducted a 10-fold cross-validation, which helps evaluate the model's performance and effectively tune the parameters.

To assess the accuracy of the predictive transcriptomic signature, we calculated the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals (CI) using the *pROC* package[119] in R. Receiver Operating Characteristic (ROC) curves were generated to graphically represent the model's true positive rate against the false positive rate. The optimal cut-point value, maximizing sensitivity and specificity, was calculated using the Youden method included in the *OptimalCutPoints* R package.[120]

Signature validation

We assessed the performance and accuracy of the 5-transcript signatures using RNAseq data from an additional pediatric cohort (Table 2). Coefficients and intercepts from the LASSO model were applied to the new data to perform ROC analysis and calculate the AUC. We tested different data comparisons: DB vs. DV pneumonias, a disaggregation of different DB pneumonias by the causal pathogen (*S.pneumoniae*, *S.aureus* and *S.pyogenes*) *vs*. DV pneumonias and a disaggregation by severity groups based on PICU admission of the patients (mild/moderate *vs.* severe disease).

Weighted Gene Correlation Network Analysis

We conducted a co-expression network analysis using the Weighted Gene Correlation Network Analysis (*WGCNA)* package[121] to detect clusters of co-expressed genes specifically correlated to viral or bacterial pneumonia. We selected all patients with DV ($n = 9$) and DB ($n = 40$) pneumonia for the analysis. Only protein-coding genes exhibiting the highest expression variance among samples (the top 25%) were included, totaling 4,831 genes. Normalized gene expression data were used to construct a signed weighted correlation network. A matrix of correlations between all pairs of selected genes was generated from the expression values and converted into an adjacency matrix with a power function. We determined a soft-thresholding power of 20 (maximum model fitting index), selected on the criterion of scale-free topology after evaluating several candidate powers (Figure S5).

We calculated the Topological Overlap Matrix (TOM) and the corresponding dissimilarity (1–TOM) values. For module detection and merging, we showed a minimum module size of 30, and a cut height threshold of 20, respectively. Correlation between module eigengenes and the viral/bacterial pneumonia was calculated to identify modules of co-expressed genes significantly associated with the phenotype (gene significance, or GS) and using the viral category as reference. Module Membership (MM) measured intramodular connectivity, and the most interconnected gene (hub gene) named the modules.

The biological pathways represented by each of the significantly correlated modules were investigated through an over-representation analysis with the *Clusterprofiler* R package,[117] with terms from GO and Reactome databases as references.

All graphics were created using R software v.4.2.0 (www-r-project.org).

QUANTIFICATION AND STATISTICAL ANALYSIS

Wilcoxon and Fisher exact test were used to assess statistical significance on clinical and demographic features between groups in numerical and categorical variables respectively. (*) indicates statistical significance. "*n*" represent number of patients or controls in each case and median values are presented with either proportion on their group or interquartile ranges with the observed minimum and maximum values.