Host gene expression signatures to identify infection type and organ dysfunction in children evaluated for sepsis: a multicentre cohort study

Luregn J Schlapbach, Devika Ganesamoorthy*, Clare Wilson*, Sainath Raman, Shane George, Peter J Snelling, Natalie Phillips, Adam Irwin, Natalie Sharp, Renate Le Marsney, Arjun Chavan, Allison Hempenstall, Seweryn Bialasiewicz, Anna D MacDonald, Keith Grimwood, Jessica C Kling, Stephen J McPherson, Antje Blumenthal, Myrsini Kaforou, Michael Levin, Jethro A Herberg, Kristen S Gibbons, Lachlan J M Coin*, for the EUCLIDS consortium† and the RAPIDS Study Group†*

Summary

Background Sepsis is defined as dysregulated host response to infection that leads to life-threatening organ dysfunction. Biomarkers characterising the dysregulated host response in sepsis are lacking. We aimed to develop host gene expression signatures to predict organ dysfunction in children with bacterial or viral infection.

Methods This cohort study was done in emergency departments and intensive care units of four hospitals in Queensland, Australia, and recruited children aged 1 month to 17 years who, upon admission, underwent a diagnostic test, including blood cultures, for suspected sepsis. Whole-blood RNA sequencing of blood was performed with Illumina NovaSeq (San Diego, CA, USA). Samples with completed phenotyping, monitoring, and RNA extraction by March 31, 2020, were included in the discovery cohort; samples collected or completed thereafter and by Oct 27, 2021, constituted the Rapid Paediatric Infection Diagnosis in Sepsis (RAPIDS) internal validation cohort. An external validation cohort was assembled from RNA sequencing gene expression count data from the observational European Childhood Life-threatening Infectious Disease Study (EUCLIDS), which recruited children with severe infection in nine European countries between 2012 and 2016. Feature selection approaches were applied to derive novel gene signatures for disease class (bacterial *vs* viral infection) and disease severity (presence *vs* absence of organ dysfunction 24 h post-sampling). The primary endpoint was the presence of organ dysfunction 24 h after blood sampling in the presence of confirmed bacterial versus viral infection. Gene signature performance is reported as area under the receiver operating characteristic curves (AUCs) and 95% CI.

Findings Between Sept 25, 2017, and Oct 27, 2021, 907 patients were enrolled. Blood samples from 595 patients were included in the discovery cohort, and samples from 312 children were included in the RAPIDS validation cohort. We derived a ten-gene disease class signature that achieved an AUC of 94·1% (95% CI 90·6–97·7) in distinguishing bacterial from viral infections in the RAPIDS validation cohort. A ten-gene disease severity signature achieved an AUC of 82·2% (95% CI 76·3–88·1) in predicting organ dysfunction within 24 h of sampling in the RAPIDS validation cohort. Used in tandem, the disease class and disease severity signatures predicted organ dysfunction within 24 h of sampling with an AUC of 90·5% (95% CI 83·3–97·6) for patients with predicted bacterial infection and 94·7% (87·8–100·0) for patients with predicted viral infection. In the external EUCLIDS validation dataset (n=362), the disease class and disease severity predicted organ dysfunction at time of sampling with an AUC of 70·1% (95% CI 44·1–96·2) for patients with predicted bacterial infection and 69·6% (53·1–86·0) for patients with predicted viral infection.

Interpretation In children evaluated for sepsis, novel host transcriptomic signatures specific for bacterial and viral infection can identify dysregulated host response leading to organ dysfunction.

Funding Australian Government Medical Research Future Fund Genomic Health Futures Mission, Children's Hospital Foundation Queensland, Brisbane Diamantina Health Partners, Emergency Medicine Foundation, Gold Coast Hospital Foundation, Far North Queensland Foundation, Townsville Hospital and Health Services SERTA Grant, and Australian Infectious Diseases Research Centre.

Introduction

Sepsis is defined as a dysregulated host response to infection that leads to life-threatening organ dysfunction.[1] Sepsis remains a leading cause of mortality in paediatric age groups, with more than 3 million annual deaths attributable to sepsis.[2] In the USA alone, paediatric sepsis was estimated to account for US$7·31 billion direct costs in 2016,[3] and one in five sepsis survivors will develop new or progressive medical conditions.[4] Most paediatric infections are viral, resulting in particular challenges in recognising sepsis in this age group. Campaigns incentivising early

Lancet Child Adolesc Health 2024; 8: 325–38

Published Online March 18, 2024 https://doi.org/10.1016/S2352-4642(24)00017-8

*Contributed equally

†Members are listed in the appendix (pp 3–10)

Children's Intensive Care Research Program, Child Health Research Centre (Prof L J Schlapbach PhD, D Ganesamoorthy PhD, S Raman MBBS PhD, S George MPH, N Phillips MPhil, N Sharp BSc, R Le Marsney MPH, A D MacDonald PhD, K S Gibbons PhD), **and Frazer Institute** (J C Kling PhD, S J McPherson PhD, Prof A Blumenthal Dr rer nat), **Faculty of Medicine, UQ Centre for Clinical Research** (A Irwin PhD), **School of Chemistry and Molecular Biosciences, The Australian Centre for Ecogenomics, and Queensland Paediatric Infectious Diseases Laboratory** (S Bialasiewicz PhD), **and Institute for Molecular Bioscience** (Prof L J M Coin PhD), **The University of Queensland, Brisbane, QLD, Australia; Department of Intensive Care and Neonatology, and Children's Research Center, University Children's Hospital Zurich, University of Zurich, Zurich, Switzerland** (Prof L J Schlapbach); **Paediatric Intensive Care Unit** (Prof L J Schlapbach, S Raman, N Sharp), **Emergency Department** (N Phillips), **and Infection Management and Prevention Services** (A Irwin), **Queensland Children's Hospital, Children's Health Queensland, Brisbane, QLD, Australia; Section of Paediatric Infectious Disease, Faculty of Medicine, Imperial College London, London, UK**

(C Wilson BChir, M Kaforou PhD, Prof M Levin PhD, J A Herberg PhD); **Department of Emergency Medicine, Gold Coast University Hospital, Southport, QLD, Australia** (S George, P J Snelling MPHTM); **School of Medicine and Dentistry and the Menzies Health Institute Queensland, Griffith University, Southport, QLD, Australia** (S George, P J Snelling, Prof K Grimwood MD); **Paediatric Intensive Care Unit, Townsville University Hospital, Townsville, QLD, Australia** (A Chavan MD); **Thursday Island Base Hospital, Thursday Island, QLD, Australia** (A Hempenstall MPH); **Department of Infectious Disease and Paediatrics, Gold Coast Health, Southport, QLD, Australia** (Prof K Grimwood); **Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, Australia** (Prof L J M Coin)

Correspondence to: Prof Luregn J Schlapbach, Department of Intensive Care and Neonatology, University Children's Hospital Zurich, University of Zurich, CH-8032 Zurich, Switzerland
luregn.schlapbach@kispi.uzh.ch

See **Online** for appendix

Research in context

Evidence before this study

We searched PubMed for research articles published in English between Jan 1, 2011, and Nov 30, 2023, with the terms "child OR paediatric", "sepsis OR septic shock", "infection", "bacterial", "viral" AND "transcriptomics OR multiarray OR RNAseq". There is an unmet need for point-of-care tests identifying host response patterns specific to bacterial versus viral infection leading to organ dysfunction in children. Whole-blood human transcriptomic analyses are a promising approach to characterise the host response, but most clinical sepsis studies to date included adults only, were limited to differentiating between bacterial and viral infections, or focused on mortality in the intensive care unit as an outcome. Furthermore, most previous studies used multi-array sequencing and relatively small cohorts.

Added value of this study

In this large multicentre, prospective Australian study with an external validation cohort recruited through a European consortium, we derived and validated two novel ten-gene expression signatures: one to discriminate between bacterial and viral infection (disease class); the second to identify patients at risk of developing organ dysfunction (disease severity). When used in tandem, this host gene expression signature could identify type of infection and organ dysfunction in children evaluated for sepsis with areas under the receiver operating characteristic curve above 90% (within 24 h of sampling) in the internal validation cohort and above 70% (at time of sampling) in the external validation cohort. Findings were robust across several severity outcomes such as need for organ support, need for vasopressors, multi-organ failure, and organ failure remote from the organ of infection. We compared the performance of these gene signature with that of existing infection-specific or severity-specific signatures and provide, for the first time, independent validation of those signatures.

Implications of all the available evidence

A novel host gene expression signature can identify type of infection and organ dysfunction in children evaluated for sepsis, thereby characterising trigger-specific and severity-specific host responses. Our findings confirm the feasibility of a precision approach to early sepsis diagnosis. Although the performance of the new signatures was high in a well characterised diverse paediatric cohort of children evaluated for sepsis with very early sampling upon presentation, further independent validation will be required. Whether the implementation of these novel sepsis signatures into point-of-care tests provides actionable information on treatable traits at the bedside, potentially leading to improved use of antibiotics, needs to be tested in future trials.

administration of antimicrobials have therefore been criticised for potentially encouraging unnecessary use of antibiotics.[5] Despite advances in microbiological diagnostics, their turnaround time and accuracy are inadequate for guiding initial empirical treatment, and they cannot predict disease severity. Rapid diagnostics have enormous potential to enhance timeliness and accuracy of sepsis treatment, and to reduce inappropriate antibiotic usage.

The mechanisms underpinning a dysregulated host response (or responses) that drives the progression from uncomplicated infection towards infection with organ dysfunction are poorly understood.[6–8] There is an unmet need for infection type-specific diagnostic markers characterising the progression from an initial infection that is simple to one with organ dysfunction. The ideal sepsis biomarker would yield information on the presence and type of underlying infection (to guide decisions on antibiotics), indicate the probability of organ dysfunction (to guide decisions on treatment escalation and resuscitation, such as fluids, inotropes, and intensive care unit [ICU] admission), and be translatable into a point-of-care platform. In the past decade, host transcriptomic biomarkers have shown great promise in differentiating between viral and bacterial infections.[9,10]

We hypothesised that RNA sequencing of whole blood from a child with suspected sepsis could be used to detect host response patterns that would differentiate between viral and bacterial infections and indicate whether a child would develop infection-associated organ dysfunction. The Rapid Paediatric Infection Diagnosis in Sepsis (RAPIDS) multicentre prospective cohort study was designed to develop and validate markers of the early host response in children evaluated for sepsis.

Methods

Study design and participants

This multicentre, prospective cohort study was performed at four hospitals in Queensland, Australia (appendix pp 3–10). Children aged 1 month to 17 years under evaluation for sepsis at participating hospital emergency departments and ICUs were eligible if upon admission they had a diagnostic test, including blood cultures, for suspected sepsis (appendix p 13). Neonates and patients with immune suppression were excluded.

The study reporting follows the Standards of Reporting of Diagnostic Accuracy Studies 2015 Update.[11] The institutional Human Research Ethics Committee approved the study on June 9, 2017 (HREC/17/QRCH/85; appendix pp 11–12). Written informed consent or consent to continue was obtained for all participants from their parents or carers (appendix p 13).

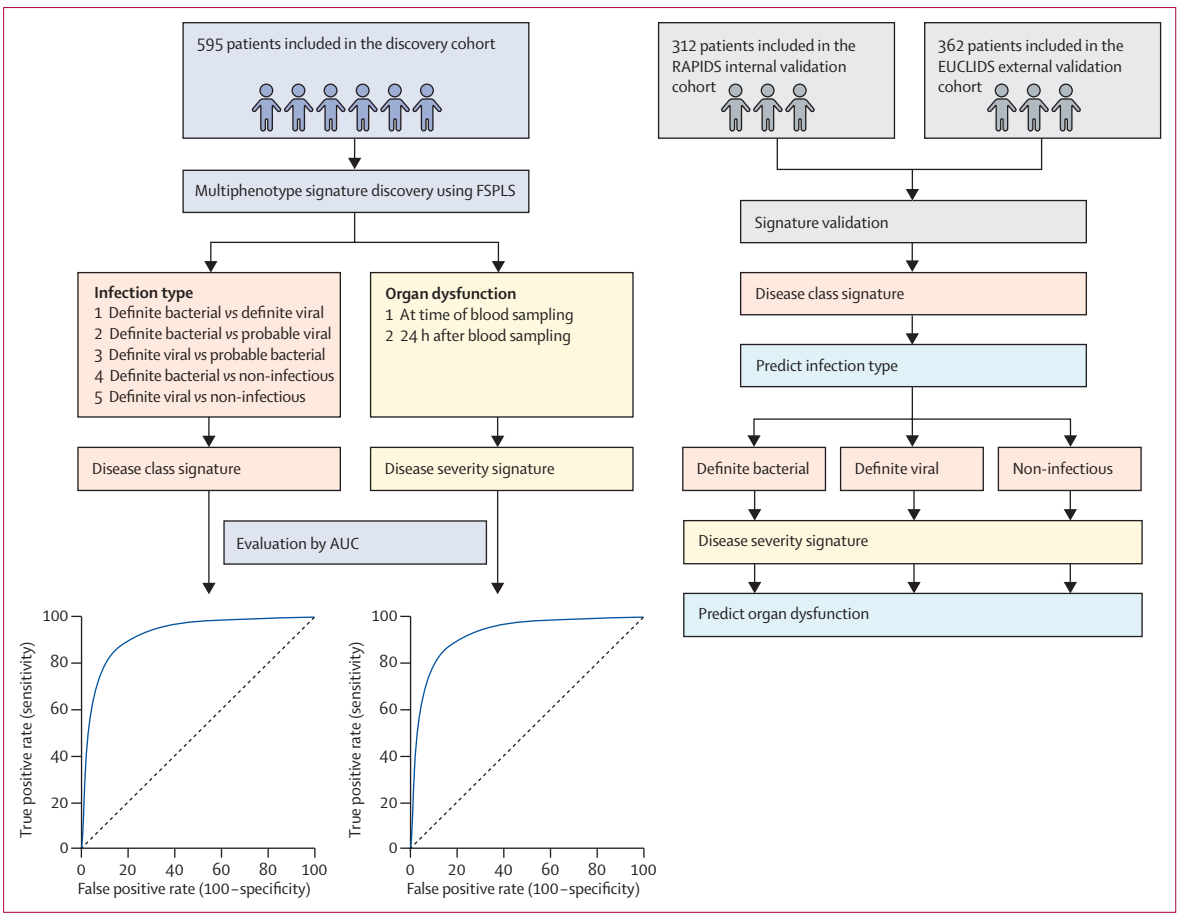

Figure 1: **Schematic workflow of the multiphenotype signature discovery using transcriptomics data**
The discovery cohort was used for the novel signature discovery. The disease class signature and disease severity signature were discovered using the FSPLS method. These signatures were validated on two independent validation cohorts. First, the infection type of the sample was predicted using the disease class signature, then the probability of developing organ dysfunction was predicted using the disease severity signature. AUC=area under the receiver operating characteristic curve. EUCLIDS=European Childhood Life-threatening Infectious Disease Study. FSPLS=forward selection partial least squares. RAPIDS=Rapid Paediatric Infection Diagnosis in Sepsis.

Procedures

2·5 mL blood was sampled in PAXgene RNA tubes (PreAnalytix, Hombrechtikon, Switzerland) in addition to routine clinical testing, including blood cultures, blood counts, C-reactive protein, and microbiological investigations, such as nasopharyngeal swabs, as indicated clinically. A REDCap study database[12] prospectively captured information on demographics, symptoms, comorbidities, microbiology results, antimicrobial treatment, and disease severity. Disease severity was assessed at baseline (time of blood sampling) and at 24 h using clinical, laboratory, and organ support criteria for organ dysfunction as defined by the 2005 International Pediatric Sepsis Definition Consensus Conference.[13,14] Accordingly, presence of organ dysfunction (any one or more of cardiovascular, respiratory, neurological, renal, hepatic, and haematological) was adjudicated. Infection status was categorised as either definite bacterial, definite viral, probable bacterial, probable viral, combined bacterial and viral, non-infectious, or unknown based on a validated approach (appendix pp 14–16).[9] Two assessors (NS, SR, or AI) experienced in paediatric critical care and infectious diseases independently verified the infection status using clinical records, microbiological results, laboratory data, and discharge reports. Adjudication of the final clinical phenotype required unanimous agreement of both assessors; any disagreement was resolved together with a third senior assessor (LJS) to ensure robust adjudication of clinical phenotypes.

Outcomes

The primary outcomes were the presence of organ dysfunction 24 h after blood sampling in children with definite bacterial infection and in children with definite viral infection. This outcome was constructed by combining the infection phenotype category (restricted to definite bacterial, definite viral, probable bacterial, probable viral, and non-infectious) with the adjudication by organ dysfunction at 24 h (ie, presence of any organ dysfunction *vs* no organ dysfunction at 24 h). Given the absence of

	Discovery cohort (n=595)	RAPIDS internal validation cohort (n=312)	EUCLIDS external validation cohort (n=362)
Sex			
Female	278 (47%)	124 (40%)	189 (52%)
Male	317 (53%)	188 (60%)	173 (48%)
Age group, years			
<1	156 (26%)	43 (14%)	99 (27%)
1 to <5	217 (36%)	159 (51%)	152 (42%)
5 to <10	119 (20%)	56 (18%)	60 (17%)
10 to <18	103 (17%)	54 (17%)	51 (14%)
Age, years	2·8 (1·0 to 7·7)	3·4 (1·4 to 7·3)	2·6 (0·8 to 5·7)
Chronic condition			
Any	132 (22%)	92 (29%)	..
Asthma	27 (5%)	12 (4%)	..
Congenital malformation	21 (4%)	16 (5%)	..
Congenital heart defect	20 (3%)	20 (6%)	..
Cerebral palsy, severe encephalopathy	18 (3%)	14 (4%)	..
Syndrome or genetic disorder	0	22 (7%)	..
Other chronic condition	80 (13%)	68 (22%)	..
Symptoms at presentation*			
Fever	464 (78%)	242/304 (80%)	..
Rash	71 (12%)	49/304 (16%)	..
Altered level of consciousness	56 (9%)	38/304 (13%)	..
Irritability	92 (15%)	35/304 (12%)	..
Seizures	34 (6%)	25/304 (8%)	..
Pain	155 (26%)	96/304 (32%)	..
Nausea or vomiting	182 (31%)	97/304 (32%)	..
Diarrhoea	72 (12%)	28/304 (9%)	..
Respiratory distress or apnoea	141 (24%)	63/304 (21%)	..
Cough	207 (35%)	97/304 (32%)	..
Pale or cyanotic episode	49 (8%)	27/304 (9%)	..
Cold extremities	13 (2%)	8/304 (3%)	..
Skin or wound infection	36 (6%)	12/304 (4%)	..
Other	156 (26%)	71/304 (23%)	..
Primary clinical focus			
Sepsis without a source	165 (28%)	94 (30%)	58 (16%)
Lower respiratory infection	183 (31%)	86 (28%)	105 (29%)
Upper respiratory infection	46 (8%)	32 (10%)	4 (1%)
Meningitis or encephalitis	18 (3%)	9 (3%)	66 (18%)
Urinary tract infection	47 (8%)	20 (6%)	13 (4%)
Arthritis or osteomyelitis	15 (3%)	3 (1%)	17 (5%)
Skin infection	28 (5%)	7 (2%)	25 (7%)
Wound infection	15 (3%)	5 (2%)	0
Toxic shock syndrome	6 (1%)	2 (1%)	11 (3%)
ENT infection or abscess	29 (5%)	11 (4%)	4 (1%)
Gastroenteritis	18 (3%)	18 (6%)	5 (1%)
Other	25 (4%)	25 (8%)	54 (15%)
Time from hospital admission to sampling, h	2·3 (1·4 to 4·1)	3·0 (1·8 to 7·5)	..
Admission to PICU			
Yes	173 (29%)	92 (29%)	212 (59%)
No	422 (71%)	220 (71%)	150 (41%)
Length of stay shorter than 24 h	137 (23%)	82 (26%)	..

(Table 1 continues on next page)

	Discovery cohort (n=595)	RAPIDS internal validation cohort (n=312)	EUCLIDS external validation cohort (n=362)
(Continued from previous page)			
Laboratory characteristics at baseline			
Base excess, mmol/L	−2·1 (−4·7 to −0·2); n=379	−1·8 (−4·3 to 0·2); n=178	−4·4 (−7·1 to −1·3); n=192)
PaO_2, mm Hg	99 (70 to 130); n=89	81 (69 to 106); n=55	97 (73 to 139); n=84
PCO_2, mm Hg	38 (34 to 45); n=89	40 (36 to 46); n=109	45 (35 to 56); n=183
Lactate, mmol/L	1·5 (1·1 to 2·3); n=394	1·4 (1·0 to 2·2); n=207	1·4 (0·9 to 2·3); n=185
Creatinine, µmol/L	30 (30 to 44); n=574	31 (30 to 42); n=283	36 (27 to 49); n=303
Bilirubin, µmol/L	7 (5 to 12); n=569	7 (5 to 12); n=287	6 (4 to 100); n=206
International normalised ratio	1·3 (1·1 to 1·6); n=140	1·3 (1·2 to 1·7); n=69	1·4 (1·2 to 1·9); n=90
Fibrinogen, g/L	3·4 (2·6 to 5·2); n=137)	3·3 (2·4 to 4·1); n=69)	..
Platelets, × 10^9 per µL	303 (219 to 378); n=564	270 (198 to 363); n=291	255 (163 to 347); n=338
White blood cell count, × 10^9 cells per µL	11·7 (7·9 to 16·5); n=583	10·7 (7·1 to 15·9); n=295	..
C-reactive protein, mg/L	25 (7 to 95); n=531	34 (10 to 89); n=267	..
Infection type			
Definite bacterial	172 (29%)	63 (20%)	190 (52%)
Probable bacterial	64 (11%)	39 (13%)	60 (17%)
Definite viral	110 (18%)	100 (32%)	39 (11%)
Probable viral	87 (15%)	32 (10%)	12 (3%)
Combined bacterial and viral infection	64 (11%)	30 (10%)	1 (<1%)
Non-infectious illness	45 (8%)	36 (12%)	..
Unknown	53 (9%)	9 (3%)	60 (17%)
Died	6 (1%)	4 (1%)	9 (2%)
At least one organ dysfunction			
Baseline	134 (23%)	76 (24%)	200 (55%)
24 h from blood sampling	87 (15%)	65 (21%)	..
Organ dysfunction remote from primary site of infection			
Baseline	132 (22%)	74 (24%)	179 (49%)
24 h from blood sampling	86 (14%)	61 (20%)	..
Any organ support			
Baseline	74 (12%)	51 (16%)	164 (45%)
24 h from blood sampling	69 (12%)	41 (13%)	..
Any inotropes			
Baseline	41 (7%)	28 (9%)	111 (31%)
24 h from blood sampling	46 (8%)	26 (8%)	..
Multi-organ dysfunction			
Baseline	81 (14%)	50 (16%)	136 (38%)
24 h from blood sampling	68 (11%)	40 (13%)	..

Data are n (%), n/N (%), or median (IQR). Percentages might not sum to 100 as a result of rounding. ENT=ear, nose, and throat. EUCLIDS=European Childhood Life-threatening Infectious Disease Study. PaO_2=arterial oxygen pressure. PCO_2=partial pressure of carbon dioxide. PICU=paediatric intensive care unit. RAPIDS=Rapid Paediatric Infection Diagnosis in Sepsis. *Percentages relate to the total number of patients with available data. Where data could not be extracted reliably from patient records, they were considered missing.

Table 1: **Clinical, microbiological, and disease severity characteristics of children evaluated for sepsis in the discovery and validation cohorts**

a gold standard for sepsis severity,[15] several secondary severity outcomes were defined: organ dysfunction remote from the primary focus of infection (as a proxy of organ dysfunction caused by a systemic process related to infection[16]); need for organ support (invasive or non-invasive respiratory support, inotropes or vasopressors, renal replacement, or extracorporeal membrane oxygenation); need for inotrope or vasopressors; multi-organ dysfunction; presence of cardiovascular, respiratory, or neurological dysfunction;[15,17] and type of organ dysfunction. These outcomes were assessed at the time of blood sampling and 24 h later; with the dynamics within the first 24 h (worsening or improving) defined as an additional secondary severity outcome appendix p 17).

RNA sequencing for discovery and validation cohorts

Blood samples were stored at –80°C until extraction. RNA was purified from samples using PAXgene Blood

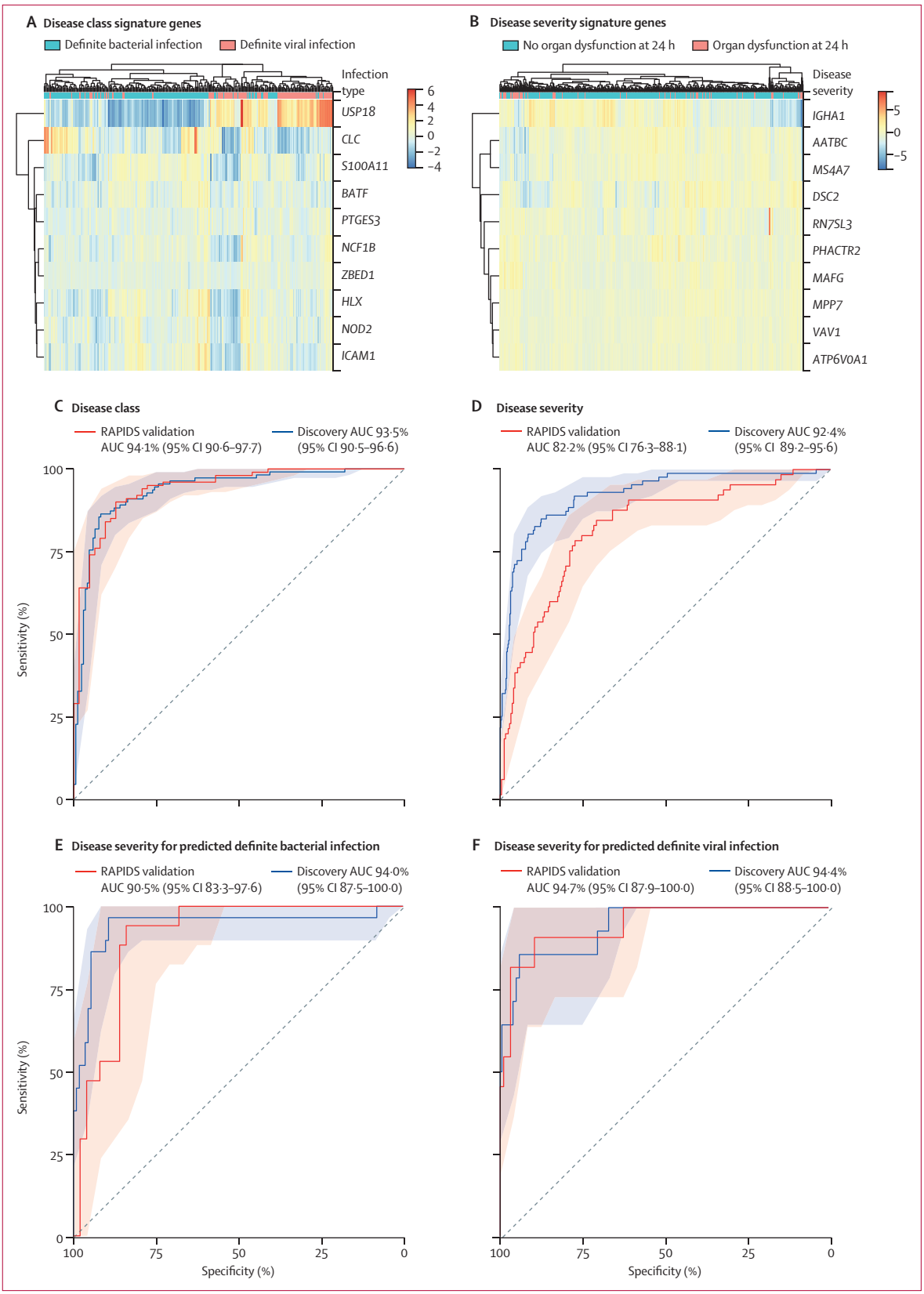

Figure 2: **Performance of disease class signature and disease severity signature in distinguishing infection type and identifying organ dysfunction**

Heat map showing the expression of disease class signature genes across patients in the discovery cohort with definite bacterial (n=172) and definite viral (n=110) infections (A); disease severity signature genes across patients in the discovery cohort with organ dysfunction (n=87) versus without organ dysfunction (n=508) at 24 h after sampling (B); ROC for the performance of the signature in the discovery and validation data to distinguish definite bacterial versus definite viral infections (C); with versus without organ dysfunction in all the patients (D); with versus without organ dysfunction in patients with predicted definite bacterial infections (E); and with versus without organ dysfunction in patients with predicted definite viral infections (F). AUC=area under the receiver operating characteristic curve. RAPIDS=Rapid Paediatric Infection Diagnosis in Sepsis. ROC=receiver operating characteristic curve.

miRNA kits (PreAnalytix). RNA library preparation and sequencing were conducted at the Institute for Molecular Biosciences Sequencing Facility (University of Queensland, Brisbane, QLD, Australia). The TruSeq RNA Ribo Zero Kit (Illumina, San Diego, CA, USA) was used for ribosomal RNA depletion and sequencing library preparation. Libraries were sequenced on a NovaSeq Sequencer (Illumina) to generate at least 20 million sequencing reads per sample. The RNA sequencing configuration was 75 bp single-end (50 samples), 100 bp single-end (545 samples), and 100 bp paired-end (316 samples). FastQC (version 1.10)[18] and MultiQC (version 0.11.9)[19] were used to assess the quality of sequencing reads. For the discovery cohort, the sample size was based on power to detect differential gene expression between conditions with 1·2-fold change, assuming 20 million reads per sample. According to the RNASeqPower package in R, we required at least 78 samples per condition to achieve 80% power. This was achieved for most comparisons, including organ dysfunction versus no organ dysfunction and definite bacterial versus definite viral. For the RAPIDS validation cohort, we used the methodology described by Burderer and colleagues[20] to estimate that a sample size of 315 would allow us to correctly estimate the sensitivity and specificity of the test within 0·05 at 95% confidence. Samples with completed phenotyping, monitoring, and RNA extraction by March 31, 2020, were included in the discovery cohort; samples collected thereafter and by Oct 31, 2021, were used to compile an internal validation cohort.

Sequencing reads were mapped to the human reference genome (version hg38) using STAR aligner (version 2.7.6a).[21] GENCODE version 35 gene transcript annotation was used for the alignment. HTSeq count (version 0.13.5)[22] was used to ascertain the number of reads mapped per gene. Principal component analysis was performed to identify any outliers (appendix pp 20–21).

Differential expression analysis

DESeq2[23] was used for differential expression analysis between different phenotypes (bacterial *vs* viral; with *vs* without organ dysfunction). Genes with less than ten read counts were excluded from analyses. Genes that had an absolute log2 fold-change of more than 1 and adjusted p value of less than 0·05 were considered as differentially expressed (appendix p 20).

External validation cohort

RNA sequencing gene expression count data were obtained from the European Childhood Life-threatening Infectious Disease Study (EUCLIDS).[24,25] This observational study recruited children with severe infection in nine European countries between 2012 and 2016.[24,25] Patients were phenotyped based on the likelihood of bacterial or viral infection[26] and considering severity at the time of sampling.

	Number of genes in signature	Discovery cohort (n=595)	RAPIDS internal validation cohort (n=312)	EUCLIDS external validation cohort (n=362)
Definite bacterial infection versus definite viral infection				
Novel disease class signature	10	0·935 (0·905–0·966)*	0·941 (0·906–0·977)*	0·909 (0·850–0·969)
Herberg et al (2016)[9]	2	0·861 (0·815–0·908)	0·900 (0·856–0·945)	0·923 (0·887–0·959)*
McHugh et al (2015)[10]	4	0·788 (0·733–0·843)	0·750 (0·673–0·827)	0·738 (0·646–0·831)
Tang et al (2017)[27]	1	0·894 (0·857–0·931)	0·883 (0·830–0·936)	0·895 (0·843–0·948)
Wong et al (2012)[28]	5	0·828 (0·779–0·876)	0·773 (0·699–0·848)	0·687 (0·591–0·784)
Sweeney et al (2016)[29]	7	0·924 (0·894–0·953)	0·911 (0·865–0·956)	0·911 (0·862–0·960)
Sampson et al (2017)[30]	4	0·894 (0·853–0·935)	0·894 (0·845–0·944)	0·921 (0·875–0·966)
Li et al (2017)[31]	4	0·691 (0·627–0·754)	0·668 (0·576–0·761)	0·800 (0·731–0·870)
Li et al (2021)[32]	3	0·881 (0·837–0·925)	0·907 (0·860–0·954)	0·906 (0·855–0·956)
Definite bacterial infection versus probable viral infection				
Novel disease class signature	10	0·912 (0·876–0·948)*	0·863 (0·771–0·954)	0·935 (0·899–0·972)*
Herberg et al (2016)[9]	2	0·794 (0·738–0·849)	0·872 (0·785–0·959)*	0·887 (0·811–0·963)
McHugh et al (2015)[10]	4	0·788 (0·730–0·846)	0·712 (0·596–0·829)	0·696 (0·526–0·866)
Tang et al (2017)[27]	1	0·722 (0·656–0·788)	0·788 (0·682–0·893)	0·726 (0·588–0·864)
Wong et al (2012)[28]	5	0·770 (0·711–0·830)	0·829 (0·739–0·919)	0·838 (0·750–0·926)
Sweeney et al (2016)[29]	7	0·862 (0·817–0·908)	0·850 (0·764–0·935)	0·770 (0·642–0·899)
Sampson et al (2017)[30]	4	0·760 (0·696–0·824)	0·852 (0·761–0·943)	0·831 (0·697–0·964)
Li et al (2017)[31]	4	0·708 (0·643–0·773)	0·779 (0·677–0·881)	0·754 (0·597–0·911)
Li et al (2021)[32]	3	0·810 (0·755–0·866)	0·851 (0·757–0·944)	0·844 (0·737–0·951)
Definite viral infection versus probable bacterial infection				
Novel disease class signature	10	0·909 (0·864–0·953)*	0·856 (0·784–0·929)*	0·793 (0·695–0·891)
Herberg et al (2016)[9]	2	0·805 (0·736–0·873)	0·827 (0·756–0·898)	0·762 (0·668–0·855)
McHugh et al (2015)[10]	4	0·741 (0·666–0·817)	0·765 (0·666–0·865)	0·668 (0·554–0·782)
Tang et al (2017)[27]	1	0·833 (0·768–0·897)	0·797 (0·707–0·887)	0·877 (0·810–0·944)
Wong et al[28]	5	0·780 (0·709–0·850)	0·686 (0·592–0·781)	0·544 (0·422–0·665)
Sweeney et al (2016)[29]	7	0·869 (0·810–0·927)	0·839 (0·763–0·915)	0·889 (0·824–0·954)*
Sampson et al (2017)[30]	4	0·838 (0·777–0·900)	0·836 (0·762–0·910)	0·807 (0·720–0·894)
Li et al (2017)[31]	4	0·603 (0·517–0·690)	0·564 (0·452–0·677)	0·578 (0·465–0·691)
Li et al (2021)[32]	3	0·828 (0·764–0·892)	0·841 (0·767–0·916)	0·774 (0·680–0·868)
Definite bacterial infection versus non-infectious				
Novel disease class signature	10	0·917 (0·879–0·954)*	0·654 (0·540–0·768)	NA
Herberg et al (2016)[9]	2	0·775 (0·704–0·847)	0·571 (0·454–0·689)	NA
McHugh et al (2015)[10]	4	0·830 (0·766–0·893)	0·663 (0·554–0·771)	NA
Tang et al (2017)[27]	1	0·598 (0·504–0·691)	0·462 (0·346–0·578)	NA
Wong et al (2012)[28]	5	0·702 (0·617–0·787)	0·712 (0·607–0·817)	NA
Sweeney et al (2016)[29]	7	0·841 (0·773–0·909)	0·678 (0·574–0·783)	NA
Sampson et al (2017)[30]	4	0·701 (0·612–0·790)	0·551 (0·434–0·668)	NA
Li et al (2017)[31]	4	0·786 (0·716–0·857)	0·713 (0·609–0·818)*	NA
Li et al (2021)[32]	3	0·615 (0·526–0·704)	0·479 (0·362–0·596)	NA

(Table 2 continues on next page)

	Number of genes in signature	Discovery cohort (n=595)	RAPIDS internal validation cohort (n=312)	EUCLIDS external validation cohort (n=362)
(Continued from previous page)				
Definite viral infection versus non-infection				
Novel disease class signature	10	0·945 (0·904–0·985)*	0·796 (0·709–0·882)*	NA
Herberg et al (2016)[9]	2	0·909 (0·862–0·957)	0·771 (0·677–0·865)	NA
McHugh et al (2015)[10]	4	0·850 (0·788–0·913)	0·690 (0·578–0·801)	NA
Tang et al (2017)[27]	1	0·914 (0·869–0·960)	0·749 (0·647–0·850)	NA
Wong et al (2012)[28]	5	0·821 (0·745–0·897)	0·694 (0·594–0·793)	NA
Sweeney et al (2016)[29]	7	0·933 (0·896–0·970)	0·766 (0·669–0·862)	NA
Sampson et al (2017)[30]	4	0·856 (0·785–0·927)	0·752 (0·656–0·849)	NA
Li et al (2017)[31]	4	0·862 (0·797–0·927)	0·726 (0·623–0·829)	NA
Li et al (2021)[32]	3	0·857 (0·795–0·919)	0·754 (0·659–0·848)	NA

Data are AUC (95% CI) unless specified otherwise. AUC=area under the receiver operating characteristic curve. EUCLIDS=European Childhood Life-threatening Infectious Disease Study. NA=not available. RAPIDS=Rapid Paediatric Infection Diagnosis in Sepsis. *For each tested phenotype, the best performing signature in terms of AUC.

Table 2: **Performance of the ten-gene expression disease class signature in distinguishing infection types, compared with existing host transcriptomic signatures**

Signature discovery and evaluation with FSPLS

In the discovery cohort, forward selection partial least squares (FSPLS; appendix p 22) was used to discover novel gene signatures to first distinguish infection types and to then predict presence of organ dysfunction (figure 1). The FSPLS approach enables simultaneous multiple comparisons to identify signatures that can be used to distinguish multiple phenotypes.

For disease class signature analysis, FSPLS was run with five different comparisons (definite bacterial *vs* definite viral, definite bacterial *vs* probable viral, definite viral *vs* probable bacterial, definite bacterial *vs* non-infectious, and definite viral *vs* non-infectious). Combined infections and unknown infections were not included in signature discovery. For severity signature analysis, FSPLS was run with participants with versus those without organ dysfunction at 24 h post sampling, and with those with versus those without organ dysfunction at the time of sampling. Severity weights stratified by disease class were obtained by running FSPLS on datasets stratified by predicted disease class (viral, bacterial, or non-infectious).

To predict sepsis in the validation cohorts, first, we used the novel disease class signature to predict the infection types as definite bacterial, definite viral, or non-infectious, because these groups have well defined phenotypes (appendix pp 16, 20). Then, we applied the novel disease severity signature for each infection type to identify organ dysfunction (figure 1).

To benchmark the novel signatures, we compared against previously published gene expression signatures for disease class[9,10,27–32] and disease severity[33–36] reported in patients with infection and sepsis (appendix pp 26–28). Because the weights of the genes in the signatures were not publicly available, we used our dataset to refit and generate the weights to use in the analysis. This allowed us to compare across all the signatures because they were all refitted similarly. We did not correct for multiple comparisons.

We assessed the enriched gene ontology terms in disease class and disease severity signature genes using ClusterProfiler (release 3.18).

All statistical analyses were performed with Stata/SE (version 17.0) and R (version 4.0.2). We used the pROC package[37] to calculate area under the receiver operating characteristic curves (AUCs) and associated 95% CIs to report the performance of signatures and the DeLong method[38] to compare the AUC values between signatures.

Role of the funding source

The funders of the study had no role in study design or conduct; data collection, analysis, or interpretation; writing of the manuscript; or the decision to submit.

Results

Between Sept 25, 2017, and Oct 27, 2021, 907 children evaluated for sepsis were enrolled: 595 in the discovery cohort and 312 in the RAPIDS validation cohort (table 1; appendix pp 34–36). Median time to blood sampling after hospital admission was 2·3 h (IQR 1·4–4·1) in the discovery cohort and 3·0 h (1·8–7·5) in the RAPIDS validation cohort. 362 children were included in the EUCLIDS external validation cohort.

Overall, 87 (15%) patients in the discovery cohort and 65 (21%) in the RAPIDS validation cohort had organ dysfunction 24 h after sampling (appendix p 17); of these patients, 76 (87%) in the discovery cohort and 57 (88%) in the validation cohort already had organ dysfunction at baseline. 24 (28%) patients in the discovery cohort and 22 (34%) patients in the validation cohort developed new or additional organ dysfunction within 24 h of sampling compared with sampling baseline. In the discovery cohort, 172 (29%) patients had definite bacterial infection, and 110 (18%) patients had definite viral infection, whereas in the RAPIDS validation cohort, 63 (20%) patients had definite bacterial infection, and 100 (32%) patients had definite viral infection (appendix pp 34–38).

We assessed differential gene expression in the discovery cohort: first for disease class (infection type), then for disease severity (organ dysfunction). Differential gene expression analysis based on the infection type identified 886 differentially expressed genes (adjusted $p<0·05$) between patients with definite viral and definite bacterial infections (appendix pp 21, 39). Comparing patients with and without organ dysfunction 24 h after sampling, 1028 genes were differentially expressed (appendix pp 21, 39). Among patients with organ dysfunction, differential gene expression patterns distinguished between definite bacterial and definite viral infections (appendix pp 21, 39).

Using FSPLS, we derived a novel ten-gene disease class signature, comprising *USP18*, *NCF1B*, *BATF*, *CLC*, *S100A11*, *ZBED1*, *PTGES3*, *HLX*, *NOD2*, and *ICAM1*, to distinguish between types of infection (figure 2A). In the discovery cohort, this disease class signature distinguished between definite bacterial sepsis and definite viral sepsis with an AUC of 93·5% (95% CI 90·5–96·6). AUCs of 94·1% (90·6–97·7) and 90·9% (85·0–96·9) were seen in the RAPIDS and EUCLIDS validation cohorts, respectively (figure 2C; table 2). Similar performances were achieved for other disease class phenotype comparisons. Compared with existing disease class signatures (appendix pp 26–28),

	Number of genes in signature	Discovery cohort (n=595)	RAPIDS internal validation cohort (n=312)	EUCLIDS external validation cohort (n=362)
Organ dysfunction 24 h after blood sampling				
Novel disease severity signature	10	0·924 (0·892–0·956)*	0·822 (0·763–0·881)	NA
Irwin et al (2017)[35]	3	0·755 (0·695–0·815)	0·772 (0·708–0·837)	NA
Lukaszewski et al (2022)[33]	27	0·870 (0·831–0·910)	0·802 (0·742–0·862)	NA
Pena et al (2014)[34]	31	0·865 (0·814–0·917)	0·750 (0·675–0·825)	NA
Baghela et al (2022)[36]	40	0·892 (0·851–0·933)	0·787 (0·724–0·849)	NA
Baghela et al (2022):[36] severity	8	0·851 (0·798–0·904)	0·724 (0·650–0·798)	NA
Baghela et al (2022):[36] severity	12	0·849 (0·800–0·898)	0·846 (0·790–0·902)*	NA
Baghela et al (2022):[36] mortality	10	0·844 (0·798–0·890)	0·822 (0·765–0·879)	NA
Organ dysfunction 24 h after blood sampling in patients with predicted bacterial infection				
Novel disease severity signature	10	0·940 (0·876–1·000)	0·905 (0·833–0·976)*	NA
Irwin et al (2017)[35]	3	0·792 (0·711–0·874)	0·735 (0·593–0·877)	NA
Lukaszewski et al (2022)[33]	27	0·887 (0·820–0·953)	0·761 (0·624–0·899)	NA
Pena et al (2014)[34]	31	0·942 (0·897–0·988)	0·709 (0·559–0·860)	NA
Baghela et al (2022)[36]	40	0·989 (0·977–1·000)*	0·788 (0·649–0·928)	NA
Baghela et al (2022):[36] severity	8	0·914 (0·844–0·985)	0·713 (0·574–0·852)	NA
Baghela et al (2022):[36] severity	12	0·909 (0·838–0·980)	0·859 (0·761–0·956)	NA
Baghela et al (2022):[36] mortality	10	0·858 (0·775–0·941)	0·611 (0·409–0·813)	NA
Organ dysfunction 24 h after blood sampling in patients with predicted viral infection				
Novel disease severity signature	10	0·944 (0·885–1·000)*	0·947 (0·879–1·000)*	NA
Irwin et al (2017)[35]	3	0·847 (0·751–0·943)	0·895 (0·776–1·000)	NA
Lukaszewski et al (2022)[33]	27	0·883 (0·795–0·971)	0·865 (0·706–1·000)	NA
Pena et al (2014)[34]	31	0·864 (0·741–0·988)	0·740 (0·541–0·938)	NA
Baghela et al (2022)[36]	40	0·917 (0·813–1·000)	0·877 (0·789–0·965)	NA
Baghela et al (2022):[36] severity	8	0·843 (0·704–0·982)	0·758 (0·563–0·952)	NA
Baghela et al (2022):[36] severity	12	0·920 (0·857–0·983)	0·837 (0·657–1·000)	NA
Baghela et al (2022):[36] mortality	10	0·852 (0·760–0·945)	0·824 (0·643–1·000)	NA
Organ dysfunction at baseline				
Novel disease severity signature	10	0·852 (0·809–0·895)*	0·775 (0·712–0·838)*	0·775 (0·727–0·823)
Irwin et al (2017)[35]	3	0·725 (0·672–0·777)	0·724 (0·656–0·791)	0·786 (0·739–0·832)
Lukaszewski et al (2022)[33]	27	0·806 (0·762–0·850)	0·771 (0·706–0·836)	0·791 (0·745–0·837)
Pena et al (2014)[34]	31	0·808 (0·763–0·852)	0·770 (0·708–0·831)	0·794 (0·748–0·840)
Baghela et al (2022)[36]	40	0·825 (0·782–0·868)	0·751 (0·686–0·816)	0·821 (0·777–0·864)*
Baghela et al (2022):[36] severity	8	0·785 (0·735–0·834)	0·725 (0·653–0·797)	0·811 (0·766–0·855)
Baghela et al (2022):[36] severity	12	0·772 (0·720–0·823)	0·715 (0·645–0·786)	0·792 (0·746–0·838)
Baghela et al (2022):[36] mortality	10	0·744 (0·692–0·796)	0·763 (0·699–0·828)	0·787 (0·740–0·833)
Organ dysfunction at baseline in patients with predicted bacterial infection				
Novel disease severity signature	10	0·924 (0·867–0·981)	0·886 (0·802–0·970)*	0·701 (0·441–0·962)
Irwin et al (2017)[35]	3	0·804 (0·721–0·888)	0·748 (0·594–0·901)	0·771 (0·572–0·970)
Lukaszewski et al (2022)[33]	27	0·898 (0·837–0·959)	0·824 (0·701–0·946)	0·667 (0·420–0·913)
Pena et al (2014)[34]	31	0·903 (0·841–0·964)	0·728 (0·572–0·884)	0·708 (0·453–0·964)
Baghela et al (2022)[36]	40	0·972 (0·944–1·000)*	0·728 (0·568–0·888)	0·743 (0·522–0·964)
Baghela et al (2022):[36] severity	8	0·852 (0·766–0·938)	0·782 (0·657–0·907)	0·611 (0·341–0·881)
Baghela et al (2022):[36] severity	12	0·863 (0·784–0·943)	0·740 (0·603–0·878)	0·660 (0·379–0·940)
Baghela et al (2022):[36] mortality	10	0·857 (0·789–0·926)	0·707 (0·533–0·881)	0·882 (0·722–1·000)*

(Table 3 continues on next page)

	Number of genes in signature	Discovery cohort (n=595)	RAPIDS internal validation cohort (n=312)	EUCLIDS external validation cohort (n=362)
(Continued from previous page)				
Organ dysfunction at baseline in patients with predicted viral infection				
Novel disease severity signature	10	0·828 (0·753–0·902)	0·757 (0·588–0·925)	0·696 (0·531–0·860)
Irwin et al (2017)[35]	3	0·729 (0·637–0·821)	0·844 (0·676–1·000)*	0·679 (0·509–0·849)
Lukaszewski et al (2022)[33]	27	0·803 (0·733–0·874)	0·779 (0·645–0·914)	0·776 (0·630–0·923)
Pena et al (2014)[34]	31	0·809 (0·721–0·897)	0·804 (0·703–0·905)	0·781 (0·644–0·917)
Baghela et al (2022)[36]	40	0·869 (0·802–0·935)*	0·596 (0·409–0·782)	0·925 (0·854–0·997)*
Baghela et al (2022):[36] severity	8	0·784 (0·704–0·864)	0·723 (0·550–0·896)	0·861 (0·752–0·970)
Baghela et al (2022):[36] severity	12	0·732 (0·640–0·824)	0·790 (0·639–0·941)	0·776 (0·636–0·917)
Baghela et al (2022):[36] mortality	10	0·741 (0·653–0·828)	0·622 (0·404–0·840)	0·820 (0·697–0·943)

Data are AUC (95% CI) unless specified otherwise. AUC=area under the receiver operating characteristic curve. EUCLIDS=European Childhood Life-threatening Infectious Disease Study. NA=not available. RAPIDS=Rapid Paediatric Infection Diagnosis in Sepsis. *For each tested phenotype, the best performing signature in terms of AUC.

Table 3: **Performance of the ten-gene expression disease severity signature in identifying organ dysfunction, by predicted infection type, compared to existing host transcriptomic signatures**

our ten-gene signature performed better for most classifications (table 2; appendix pp 40–46). The disease class signature also distinguished between combined bacterial and viral infection status and unknown infection status (appendix p 47). Gene ontology enrichment analysis of the signature genes showed enrichment of immune response gene ontology terms (appendix pp 29–32).

Using FSPLS, we also derived a novel ten-gene disease severity signature, comprising *AATBC*, *MAFG*, *VAV1*, *MS4A7*, *IGHA1*, *ATP6V0A1*, *RN7SL3*, *MPP7*, *DSC2*, and *PHACTR2*, to identify presence of organ dysfunction 24 h after sampling (figure 2B). In the discovery cohort, the disease severity signature distinguished between patients with and without organ dysfunction at 24 h with an AUC of 92·4% (95% CI 89·2–95·6) and an AUC of 82·2% (76·3–88·1) in the RAPIDS validation cohort (figure 2D; table 3). Compared with previously reported gene expression signatures for disease severity (appendix pp 26–28), the novel signature showed similar or superior performance (table 3; appendix pp 48–55). Gene ontology enrichment analysis of the novel disease severity signature genes showed enrichment of immunoglobulin complex, signal recognition, and proton transporting gene ontology terms, indicating biological pathways involved in the development of organ dysfunction (appendix pp 29–32).

Prediction of sepsis was tested in the RAPIDS validation cohort. Using severity weights stratified by disease class, the prediction of disease severity in patients with predicted definite bacterial infection achieved an AUC of 90·5% (95% CI 83·3–97·6; figure 2E; table 3). The prediction of disease severity in patients with predicted definite viral infection achieved an AUC of 94·7% (87·9–100·0; figure 2F; table 3).

We then assessed the disease class and disease severity signatures with EUCLIDS cohort data, which only included severity status recorded at the time of blood sampling. In the EUCLIDS cohort, prediction of disease severity in patients with definite bacterial infection achieved an AUC of 70·1% (95% CI 44·1–96·2), whereas the prediction of disease severity in patients with predicted definite viral infection achieved an AUC of 69·6% (53·1–86·0; table 3).

The novel severity signatures performed comparably in distinguishing secondary severity outcomes at baseline and within 24 h of sampling, including organ dysfunction remote from the site of infection, type of organ dysfunction, need for organ support, and need for inotrope support (table 4). In the RAPIDS validation cohort, the signatures identified progressive multi-organ dysfunction within 24 h of sampling with an AUC of 75·8% (95% CI 67·3–84·3; table 4).

Adding clinical information such as C-reactive protein concentrations and leukocyte counts to the gene signatures did not improve the prediction of disease class and disease severity, which was superior to routine clinical markers (appendix p 33). Both the disease class and disease severity signatures in the discovery and validation cohorts performed similarly across the age ranges included (appendix p 56).

Discussion

In this multicentre, prospective cohort study involving 907 children evaluated for suspected sepsis, we derived and validated novel gene expression signatures in order to rapidly distinguish between infections of bacterial versus viral aetiology and to predict life-threatening organ dysfunction within 24 h of hospital admission. The gene signatures provided actionable information about infection type and disease severity. The approach thus shows the potential of host transcriptomics to characterise sepsis in children.

In the past decade, several studies including adult and paediatric patients have used host gene expression analyses to differentiate between infections caused by

	Discovery cohort (n=595)	RAPIDS internal validation cohort (n=312)	EUCLIDS external validation cohort (n=362)
Organ dysfunction remote to infection site 24 h from blood sampling	0·927 (0·894–0·959)	0·820 (0·759–0·882)	NA
Organ dysfunction remote to infection site at baseline	0·854 (0·811–0·897)	0·778 (0·714–0·842)	0·770 (0·730–0·820)
Need for organ support 24 h from blood sampling	0·957 (0·940–0·974)	0·825 (0·751–0·899)	NA
Need for organ support at baseline	0·925 (0·895–0·954)	0·781 (0·709–0·853)	0·780 (0·740–0·830)
Administration of inotropes 24 h from blood sampling	0·963 (0·946–0·981)	0·846 (0·759–0·932)	NA
Administration of inotropes at baseline	0·940 (0·890–0·980)	0·820 (0·730–0·910)	0·830 (0·780–0·870)
Cardiovascular, respiratory, or neurological organ dysfunction 24 h from blood sampling*	0·962 (0·943–0·980)	0·828 (0·732–0·924)	NA
Cardiovascular, respiratory, or neurological organ dysfunction at baseline*	0·919 (0·874–0·963)	0·852 (0·757–0·947)	0·730 (0·660–0·814)
Multi-organ dysfunction 24 h from blood sampling	0·959 (0·942–0·976)	0·858 (0·792–0·925)	NA
Multi-organ dysfunction at baseline	0·902 (0·861–0·943)	0·801 (0·729–0·873)	0·800 (0·750–0·850)
Organ dysfunction better 24 h from blood sampling†	0·715 (0·648–0·783)	0·623 (0·515–0·731)	NA
Organ dysfunction worse 24 h from blood sampling†	0·872 (0·810–0·934)	0·721 (0·611–0·831)	NA
Multi-organ dysfunction better 24 h from blood sampling‡	0·798 (0·708–0·888)	0·645 (0·501–0·788)	NA
Multi-organ dysfunction worse 24 h from blood sampling‡	0·896 (0·851–0·941)	0·758 (0·673–0·843)	NA
Cardiovascular organ dysfunction 24 h from blood sampling	0·958 (0·937–0·978)	0·811 (0·736–0·886)	NA
Cardiovascular organ dysfunction at baseline	0·837 (0·783–0·892)	0·789 (0·705–0·872)	0·810 (0·770–0·860)
Respiratory organ dysfunction 24 h from blood sampling	0·953 (0·934–0·972)	0·840 (0·771–0·909)	NA
Respiratory organ dysfunction at baseline	0·914 (0·882–0·947)	0·782 (0·705–0·860)	0·760 (0·700–0·810)
Neurological organ dysfunction 24 h from blood sampling	0·908 (0·861–0·955)	0·820 (0·750–0·890)	NA
Neurological organ dysfunction at baseline	0·885 (0·840–0·930)	0·806 (0·737–0·875)	0·680 (0·610–0·750)
Renal organ dysfunction 24 h from blood sampling	0·890 (0·777–1·000)	0·692 (0·366–1·000)	NA
Renal organ dysfunction at baseline	0·864 (0·773–0·955)	0·692 (0·462–0·921)	0·760 (0·660–0·860)
Haematological organ dysfunction 24 h from blood sampling	0·945 (0·906–0·985)	0·735 (0·594–0·876)	NA
Haematological organ dysfunction at baseline	0·869 (0·800–0·938)	0·793 (0·650–0·935)	0·700 (0·580–0·820)
Hepatic organ dysfunction 24 h from blood sampling	0·950 (0·887–1·000)	0·678 (0·102–1·000)	NA
Hepatic organ dysfunction at baseline	0·749 (0·410–1·000)	0·600 (0·279–0·921)	0·720 (0·410–1·000)

Data are AUC (95% CI) for each outcome assessed 24 h after study blood sampling and at time of study blood sampling (baseline). Organ dysfunction was assessed using the 2005 International Pediatric Sepsis Consensus Conference criteria. AUC=area under the receiver operating characteristic curve. EUCLIDS=European Childhood Life-threatening Infectious Disease Study. NA=not available. RAPIDS=Rapid Paediatric Infection Diagnosis in Sepsis. *Presence of either cardiac, respiratory, or neurological organ dysfunction. †Compared with organ dysfunction at time of sampling, organ dysfunction increasing (worse) or decreasing (better) at 24 h post sampling. ‡Compared with multi-organ dysfunction at time of sampling, multi-organ dysfunction increasing to 2 or higher (worse) or decreasing to 0–1 (better) at 24 h post sampling.

Table 4: **Performance of novel severity signature in identifying various severity outcomes**

bacteria and virus.[9,32,39,40] ICU-based studies have derived pathways and associated differentially regulated genes with mortality in critically ill patients, which might identify patients who are more likely to suffer harm from specific interventions such as corticosteroids.[41,42] However, the integration of infection status and development of organ dysfunction—two key dimensions of sepsis—in a unifying measure of dysregulated host response has been lacking. Most gene expression studies to date have relied on microarray analysis. In this study, we used RNA sequencing to achieve substantially higher transcript resolution, and we included a larger number of patients to increase power. With the FSPLS approach, we derived minimal gene signatures for disease class and disease severity. FSPLS iteratively finds the next most explanatory feature after removing the projection of features onto the space spanned by variables previously selected. FSPLS tends to find smaller signatures than other common approaches such as the least absolute shrinkage and selection operator (LASSO). We enrolled children early and took blood samples within a median of 3 h of presentation to hospital. Nonetheless, 88% of patients who met criteria for organ dysfunction at 24 h after admission already manifested at least single organ dysfunction at the time of sampling.

Compared with eight existing infection type signatures,[9,10,27–32] our novel disease class signature performed similarly or better in both validation cohorts. Compared with seven existing disease severity signatures,[33–36] our novel signature showed superior performance for predicting organ dysfunction at 24 h after sampling in patients with predicted bacterial or viral infection within the RAPIDS validation cohort. Whereas

the best performing disease severity signature to date relies on 40 genes,[36] our ten-gene signatures are less complex, making implementation in rapid point-of-care platforms more feasible. When assessing other severity outcomes 24 h after sampling, such as organ dysfunction remote from the site of infection, multi-organ dysfunction, or need for inotropes, the novel disease severity signature performed well with AUCs above 80%. In the EUCLIDS validation cohort, however, information on organ dysfunction was only available from the time of sampling, and overall AUCs were above 70%. The relatively poorer performance of the disease severity signature in the EUCLIDS validation might further relate to later sampling, cohort differences such as recruitment bias towards severe bacterial infections, restriction to community-acquired infections, and differences in primary clinical focus (appendix pp 34–36).

Overall, we observed a wider range in diagnostic performance across the discovery and validation cohorts with disease severity signatures compared with disease class signatures. Notably, contrary to categorisation by microbiologically confirmed bacterial versus viral infection, concepts of disease severity such as organ dysfunction inherently lack a true gold standard against which to benchmark biomarkers. Accordingly, the clinical criteria used to define the primary outcome of disease severity might fall short of the underlying biological complexity, as seen in recent electronic health-record derived studies of sepsis phenotypes.[43–46] The heterogeneity of underlying causes, mechanisms, treatments, and trajectories characterising critical illness syndromes therefore fundamentally challenge the feasibility of a simple severity marker.

Both disease class and disease severity gene signatures derived in this study predicted multiple phenotypes successfully. The disease class signature identified infection type in patients evaluated for sepsis. The disease severity signature identified the presence of organ dysfunction and several other severity phenotypes, including whether the organ dysfunction was likely to worsen within 24 h of sampling. In combination, given the high negative predictive value (appendix pp 48–55), these novel sepsis signatures have the potential to guide clinical decision making on use of (or to rule out) antimicrobials and escalation of care.

Mortality and other severity outcomes in paediatric sepsis relate directly to delays between presentation and delivery of a sepsis treatment bundle.[47] Sepsis quality improvement programmes usually focus on presumed infection in the presence of clinical indicators of altered physiology. However, it is well recognised that clinical features of sepsis are often subtle and non-specific, in particular in children, in whom viral aetiologies predominate. Therefore, initiatives to promote early treatment with intravenous antibiotics have been met with criticism because they risk inappropriate use of antibiotics, potentially promoting antimicrobial resistance. In this context, a direct marker of a dysregulated host response to bacterial versus viral infection remains highly desirable and can serve to identify treatable traits early upon presentation.

Several limitations of this study need to be considered. First, although the findings were validated in an a-priori defined separate sequencing batch of patients, external validation using the EUCLIDS cohort was only partially feasible because it did not contain 24 h outcome data. Second, patients were recruited in a high-income setting with a low mortality rate, with a predominance of White patients, and almost complete absence of fungal and parasitic infections, which might not be representative of patients in less-resourced settings. Third, the study design excluded immunosuppressed patients, and most of the included patients had community-acquired infections, implying the need to validate the gene signatures in relatively more comorbid cohorts with hospital-acquired infections. Finally, we did not perform RT-PCR validation of the novel gene signatures, and future replication using a point-of-care device will be required.

In conclusion, in this large cohort of children evaluated for sepsis, encompassing a broad range of disease severity, pathogens, and comorbidities, we derived host transcriptomic signatures that discriminated, with high accuracy, between bacterial and viral infection and identified patients who were likely to manifest organ dysfunction within 24 h of sampling. Whether such actionable information can direct therapy to patients who are most likely to benefit from timely delivery of a sepsis bundle, while reducing unnecessary use of antibiotics, needs to be tested by interventional trials.

Contributors

LJS and LJMC contributed to concept and design of the study. LJS, DG, CW, SR, SG, PJS, NP, AI, NS, RLM, AC, AH, SB, ADM, KG, JCK, SJM, AB, MK, ML, JAH, KSG, and LJMC were responsible for acquisition, analysis, or interpretation of data. LJS, DG, CW, KSG, and LJMC drafted the manuscript. LJS, DG, CW, SR, SG, PJS, NP, AI, NS, RLM, AC, AH, SB, ADM, KG, JCK, SJM, AB, MK, ML, JAH, KSG, and LJMC contributed to critical revision of the manuscript for important intellectual content. LJS, DG, RLM, KSG, and LJMC did the statistical analysis. LJS, PJS, CW, SG, MK, ML, JAH, and LJMC obtained funding. LJS, ADM, JCK, SJM, and KSG provided administrative, technical, or material support. LJS, JAH, ML, KSG, and LJMC supervised the study. LJS, DG, and LJMC accessed and verified the data, and were responsible for the decision to submit the manuscript.

Declaration of interests

We declare no competing interests.

Data sharing

RNA sequencing data, processed gene counts, and patient metadata are available from https://doi.org/10.48610/7f1ad77.

Acknowledgments

We thank all the parents and children participating in this study and the medical, nursing, and research teams at the participating sites for their help in study setup, recruitment, data collection, and monitoring of study data for their invaluable support. We thank Angelika Christ from the UQ Sequencing Facility at The University of Queensland (Brisbane, QLD, Australia) for performing all the RNA sequencing work. Research reported in this publication was supported by Australian Government Medical Research Future Fund Genomic Health Futures Mission Project Grant GHFMPACI000004, with further support from Children's

Hospital Foundation Queensland Program Grant 50335, Australian Government Medical Research Future Fund administered by Brisbane Diamantina Health Partners MRFF9100000, Emergency Medicine Foundation EMPJ-378R27-2017-SNE, Gold Coast Hospital Foundation RGS2017-LG0020, Far North Queensland Foundation, Townsville Hospital and Health Services SERTA Grant, and Australian Infectious Diseases Research Centre. The EUCLIDS consortium was supported by EU's 7th Framework programme EC-GA 279185. AB was supported by Australian Research Council Future Fellowship FT220100487. DG was supported by Children's Hospital Foundation Queensland Mary McConnel Career Boost Program for Women in Paediatric Research WIS0092021, QLD Government Advance Queensland Women's Research Assistance Program WRAP182-2019RD1. JCK was supported by University of Queensland Postdoctoral Research Fellowship. MK was supported by Wellcome Trust 206508/Z/17/Z, UK Medical Research Foundation (MRF-160-0008-ELP-KAFO-C0801). SR was supported by Children's Hospital Foundation Queensland PCC0052021. LJS was supported by Australian National Health and Medical Research Council Practitioner Fellowship APP1161657, Children's Hospital Foundation Queensland 50282, NOMIS Foundation. CW was supported by Wellcome Trust 203928/Z/16/Z. JAH, MK, ML, and CW were supported by UK National Institute of Health and Care Research Imperial Biomedical Research Centre.

References

1 Schlapbach LJ, Watson RS, Sorce LR, et al. International Consensus Criteria for Pediatric Sepsis and Septic Shock. *JAMA* 2024; **331:** 665–74.
2 Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet* 2020; **395:** 200–11.
3 Carlton EF, Barbaro RP, Iwashyna TJ, Prescott HC. Cost of Pediatric Severe Sepsis Hospitalizations. *JAMA Pediatr* 2019; **173:** 986–87.
4 Maddux AB, Zimmerman JJ, Banks RK, et al. Health resource use in survivors of pediatric septic shock in the United States. *Pediatr Crit Care Med* 2022; **23:** e277–88.
5 Schlapbach LJ, Weiss SL, Wolf J. Reducing collateral damage from mandates for time to antibiotics in pediatric sepsis-primum non nocere. *JAMA Pediatr* 2019; **173:** 409–10.
6 Angus DC. Opening the debate on the new sepsis definition defining sepsis: a case of bounded rationality and fuzzy thinking? *Am J Respir Crit Care Med* 2016; **194:** 14–15.
7 Cohen J, Vincent JL, Adhikari NK, et al. Sepsis: a roadmap for future research. *Lancet Infect Dis* 2015; **15:** 581–614.
8 Reyes M, Filbin MR, Bhattacharyya RP, et al. An immune-cell signature of bacterial sepsis. *Nat Med* 2020; **26:** 333–40.
9 Herberg JA, Kaforou M, Wright VJ, et al. Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children. *JAMA* 2016; **316:** 835–45.
10 McHugh L, Seldon TA, Brandon RA, et al. A molecular host response assay to discriminate between sepsis and infection-negative systemic inflammation in critically ill patients: discovery and validation in independent cohorts. *PLoS Med* 2015; **12:** e1001916.
11 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; **351:** h5527.
12 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42:** 377–81.
13 Goldstein B, Giroir B, Randolph A. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med* 2005; **6:** 2–8.
14 Gebara BM. Values for systolic blood pressure. *Pediatr Crit Care Med* 2005; **6:** 500.
15 Schlapbach LJ, Goertz S, Hagenbuch N, et al. Organ dysfunction in children with blood culture-proven sepsis: comparative performance of four scores in a national cohort study. *Pediatr Crit Care Med* 2023; published online Oct 25. https://doi.org/10.1097/PCC.0000000000003388.
16 Morin L, Hall M, de Souza D, et al. The current and future state of pediatric sepsis definitions: an international survey. *Pediatrics* 2022; **149:** e2021052565.
17 Levin M, Cunnington AJ, Wilson C, et al. Effects of saline or albumin fluid bolus in resuscitation: evidence from re-analysis of the FEAST trial. *Lancet Respir Med* 2019; **7:** 581–93.
18 Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed Feb 5, 2023).
19 Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; **32:** 3047–48.
20 Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996; **3:** 895–900.
21 Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; **29:** 15–21.
22 Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; **31:** 166–69.
23 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15:** 550.
24 Boeddha NP, Schlapbach LJ, Driessen GJ, et al. Mortality and morbidity in community-acquired sepsis in European pediatric intensive care units: a prospective cohort study from the European Childhood Life-threatening Infectious Disease Study (EUCLIDS). *Crit Care* 2018; **22:** 143.
25 Martinón-Torres F, Salas A, Rivero-Calle I, et al. Life-threatening infections in children in Europe (the EUCLIDS Project): a prospective cohort study. *Lancet Child Adolesc Health* 2018; **2:** 404–14.
26 Nijman RG, Oostenbrink R, Moll HA, et al. A novel framework for phenotyping children with suspected or confirmed infection for future biomarker studies. *Front Pediatr* 2021; **9:** 688272.
27 Tang BM, Shojaei M, Parnell GP, et al. A novel immune biomarker *IF127* discriminates between influenza and bacteria in patients with suspected respiratory infection. *Eur Respir J* 2017; **49:** 1602098.
28 Wong HR, Salisbury S, Xiao Q, et al. The pediatric sepsis biomarker risk model. *Crit Care* 2012; **16:** R174.
29 Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci Transl Med* 2016; **8:** 346ra91.
30 Sampson DL, Fox BA, Yager TD, et al. A four-biomarker blood signature discriminates systemic inflammation due to viral infection versus other etiologies. *Sci Rep* 2017; **7:** 2914.
31 Li Y, Li Y, Bai Z, Pan J, Wang J, Fang F. Identification of potential transcriptomic markers in developing pediatric sepsis: a weighted gene co-expression network analysis and a case-control validation study. *J Transl Med* 2017; **15:** 254.
32 Li HK, Kaforou M, Rodriguez-Manzano J, et al. Discovery and validation of a three-gene signature to distinguish COVID-19 and other viral infections in emergency infectious disease presentations: a case-control and observational cohort study. *Lancet Microbe* 2021; **2:** e594–603.
33 Lukaszewski RA, Jones HE, Gersuk VH, et al. Presymptomatic diagnosis of postoperative infection and sepsis using gene expression signatures. *Intensive Care Med* 2022; **48:** 1133–43.
34 Pena OM, Hancock DG, Lyle NH, et al. an endotoxin tolerance signature predicts sepsis and organ dysfunction at initial clinical presentation. *EBioMedicine* 2014; **1:** 64–71.
35 Irwin AD, Grant A, Williams R, et al. Predicting risk of serious bacterial infections in febrile children in the emergency department. *Pediatrics* 2017; **140:** e20162853.
36 Baghela A, Pena OM, Lee AH, et al. Predicting sepsis severity at first clinical presentation: the role of endotypes and mechanistic signatures. *EBioMedicine* 2022; **75:** 103776.
37 Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; **12:** 77.
38 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44:** 837–45.
39 Tsalik EL, Henao R, Montgomery JL, et al. Discriminating bacterial and viral infection using a rapid host gene expression test. *Crit Care Med* 2021; **49:** 1651–63.

40 Sweeney TE, Azad TD, Donato M, et al. Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters. *Crit Care Med* 2018; **46:** 915–25.

41 Sweeney TE, Perumal TM, Henao R, et al. A community approach to mortality prediction in sepsis via gene expression analysis. *Nat Commun* 2018; **9:** 694.

42 Wong HR, Cvijanovich NZ, Anas N, et al. Developing a clinically feasible personalized medicine approach to pediatric septic shock. *Am J Respir Crit Care Med* 2015; **191:** 309–15.

43 Maslove DM, Tang B, Shankar-Hari M, et al. Redefining critical illness. *Nat Med* 2022; **28:** 1141–48.

44 Komorowski M, Green A, Tatham KC, Seymour C, Antcliffe D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine* 2022; **86:** 104394.

45 Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* 2019; **321:** 2003–17.

46 Sanchez-Pinto LN, Bennet TD, Stroup EK, et al. Derivation, validation, and clinical relevance of a pediatric sepsis phenotype with persistent hypoxemia, encephalopathy, and shock. *Pediatr Crit Care Med* 2023; **24:** 795–806.

47 Weiss SL, Peters MJ, Alhazzani W, et al. Surviving sepsis campaign international guidelines for the management of septic shock and sepsis-associated organ dysfunction in children. *Intensive Care Med* 2020; **46** (suppl 1)**:** 10–67.