JAMA | Preliminary Communication | INNOVATIONS IN HEALTH CARE DELIVERY

Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children

Jethro A. Herberg, PhD; Myrsini Kaforou, PhD; Victoria J. Wright, PhD; Hannah Shailes, BSc; Hariklia Eleftherohorinou, PhD; Clive J. Hoggart, PhD; Miriam Cebey-López, MSc; Michael J. Carter, MRCPCH; Victoria A. Janes, MD; Stuart Gormley, MRes; Chisato Shimizu, MD; Adriana H. Tremoulet, MD; Anouk M. Barendregt, BSc; Antonio Salas, PhD; John Kanegaye, MD; Andrew J. Pollard, PhD; Saul N. Faust, PhD; Sanjay Patel, FRCPCH; Taco Kuijpers, PhD; Federico Martinón-Torres, PhD; Jane C. Burns, MD; Lachlan J. M. Coin, PhD; Michael Levin, FRCPCH; for the IRIS Consortium

IMPORTANCE Because clinical features do not reliably distinguish bacterial from viral infection, many children worldwide receive unnecessary antibiotic treatment, while bacterial infection is missed in others.

OBJECTIVE To identify a blood RNA expression signature that distinguishes bacterial from viral infection in febrile children.

DESIGN, SETTING, AND PARTICIPANTS Febrile children presenting to participating hospitals in the United Kingdom, Spain, the Netherlands, and the United States between 2009-2013 were prospectively recruited, comprising a discovery group and validation group. Each group was classified after microbiological investigation as having definite bacterial infection, definite viral infection, or indeterminate infection. RNA expression signatures distinguishing definite bacterial from viral infection were identified in the discovery group and diagnostic performance assessed in the validation group. Additional validation was undertaken in separate studies of children with meningococcal disease (n = 24) and inflammatory diseases (n = 48) and on published gene expression datasets.

EXPOSURES A 2-transcript RNA expression signature distinguishing bacterial infection from viral infection was evaluated against clinical and microbiological diagnosis.

MAIN OUTCOMES AND MEASURES Definite bacterial and viral infection was confirmed by culture or molecular detection of the pathogens. Performance of the RNA signature was evaluated in the definite bacterial and viral group and in the indeterminate infection group.

RESULTS The discovery group of 240 children (median age, 19 months; 62% male) included 52 with definite bacterial infection, of whom 36 (69%) required intensive care, and 92 with definite viral infection, of whom 32 (35%) required intensive care. Ninety-six children had indeterminate infection. Analysis of RNA expression data identified a 38-transcript signature distinguishing bacterial from viral infection. A smaller (2-transcript) signature (*FAM89A* and *IFI44L*) was identified by removing highly correlated transcripts. When this 2-transcript signature was implemented as a disease risk score in the validation group (130 children, with 23 definite bacterial, 28 definite viral, and 79 indeterminate infections; median age, 17 months; 57% male), all 23 patients with microbiologically confirmed definite bacterial infection were classified as bacterial (sensitivity, 100% [95% CI, 85%-100%]) and 27 of 28 patients with definite viral infection were classified as viral (specificity, 96.4% [95% CI, 89.3%-100%]). When applied to additional validation datasets from patients with meningococcal and inflammatory diseases, bacterial infection was identified with a sensitivity of 91.7% (95% CI, 79.2%-100%) and 90.0% (95% CI, 70.0%-100%), respectively, and with specificity of 96.0% (95% CI, 88.0%-100%) and 95.8% (95% CI, 89.6%-100%). Of the children in the indeterminate groups, 46.3% (63/136) were classified as having bacterial infection, although 94.9% (129/136) received antibiotic treatment.

CONCLUSIONS AND RELEVANCE This study provides preliminary data regarding test accuracy of a 2-transcript host RNA signature discriminating bacterial from viral infection in febrile children. Further studies are needed in diverse groups of patients to assess accuracy and clinical utility of this test in different clinical settings.

JAMA. 2016;316(8):835-845. doi:10.1001/jama.2016.11236
Corrected on February 7, 2017.

Editorial page 824

Related article page 846

Supplemental content

CME Quiz at jamanetworkcme.com

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: Members of the IRIS Consortium are listed at the end of this article.

Corresponding Author: Michael Levin, FRCPCH, Section of Paediatrics, Division of Infectious Diseases, Department of Medicine, Imperial College London, Norfolk Place, London W2 1PG United Kingdom (m.levin@imperial.ac.uk).

The majority of febrile children have self-resolving viral infection, but a small proportion have life-threatening bacterial infections. Although culture of bacteria from normally sterile sites remains the gold standard for confirming bacterial infection, culture results may take several days and are frequently negative when infection resides in inaccessible sites or when antibiotics have been previously administered.[1-3] Current practice is to admit ill-appearing febrile children to the hospital and administer parenteral antibiotics while awaiting culture results.[4-6] Because only a minority of febrile children are ultimately proven to have bacterial infection, the process of ruling out bacterial infection results in a major burden on health care resources and in inappropriate antibiotic prescription.[7]

Molecular tests have the potential to identify bacterial and viral pathogens and improve distinction between bacterial and viral infection.[8] Rapid molecular viral diagnostics have increased the proportion of patients shown to carry respiratory pathogens,[9] but viruses are frequently identified in nasopharyngeal samples from healthy children.[10] Thus, detection of a virus in the nasopharynx does not rule out bacterial infection and is of little help in decisions on whether to administer antibiotics.

A number of studies have suggested that specific infections can be identified by the pattern of host genes activated during the inflammatory response.[11-15] This study investigated whether bacterial infection can be distinguished from other causes of fever in children by the pattern of host genes activated or suppressed in blood in response to the infection and whether a subset of these genes could be identified as the basis for a diagnostic test.

Key Points

Question Can febrile children with bacterial infection be distinguished from those with viral infection and other common causes of fever using whole-blood gene expression profiling?

Findings In this cross-sectional study that included 370 febrile children, those with bacterial infection were distinguished from those with viral infection with a sensitivity in the validation group of 100% (95% CI, 85%-100%) and specificity of 96.4% (95% CI, 89.3%-100%), using a 2-transcript signature.

Meaning This study provides preliminary data on the performance of a 2-transcript host RNA signature for discriminating bacterial from viral infection in febrile children. Further studies are needed in diverse groups of patients to assess accuracy and clinical utility of this test in different clinical settings.

Methods

Study Conduct and Oversight

Written informed consent was obtained from parents or guardians using locally approved research ethics committee permissions (St Mary's Research Ethics Committee (REC 09/H0712/58 and EC3263); Ethical Committee of Clinical Investigation of Galicia (CEIC ref 2010/015); UCSD Human Research Protection Program No. 140220; and Academic Medical Centre, University of Amsterdam (NL41846.018.12 and NL34230.018.10).

Discovery and Validation Groups

The overall design of the study is shown in **Figure 1**, **Figure 2**, and **Figure 3**.

Clinical data and samples were identified only by study number. Assignment of patients to clinical groups was made by consensus of 2 clinicians independent of those managing the patient, after review of investigation results using previously agreed-on definitions (Figure 2). Patients were recruited prospectively as part of a UK National Institute of Health Research–supported study (NIHR ID 8209), the Immunopathology of Respiratory, Inflammatory and Infectious Disease Study (IRIS), which recruited children at 3 UK hospitals; patients also were recruited in Spain (GENDRES network, Santiago de Compostela), and the United States (Rady Children's Hospital, San Diego). Inclusion criteria were fever (axillary temperature ≥38°C) and perceived illness of sufficient severity to warrant blood testing in children younger than 17 years. Patients with comorbidities likely to affect gene expression (bone marrow transplant, immunodeficiency, or immunosuppressive treatment) were excluded. Blood samples for RNA analysis were collected together with clinical blood tests at, or as close as possible to, presentation to hospital, irrespective of antibiotic use at the time of collection.

Additional Validation Groups

Additional validation groups (eMethods and eTable 1 in the Supplement) included children with meningococcal sepsis,[16] inflammatory diseases (juvenile idiopathic arthritis and Henoch-Schönlein purpura), and published gene expression datasets that compared bacterial infection with viral infection[12,15,17] or inflammatory disease.[18] Healthy children were recruited from outpatient departments. Data from healthy controls were not used in identification or validation of gene expression signatures and were used only for interpretation of direction of gene regulation.

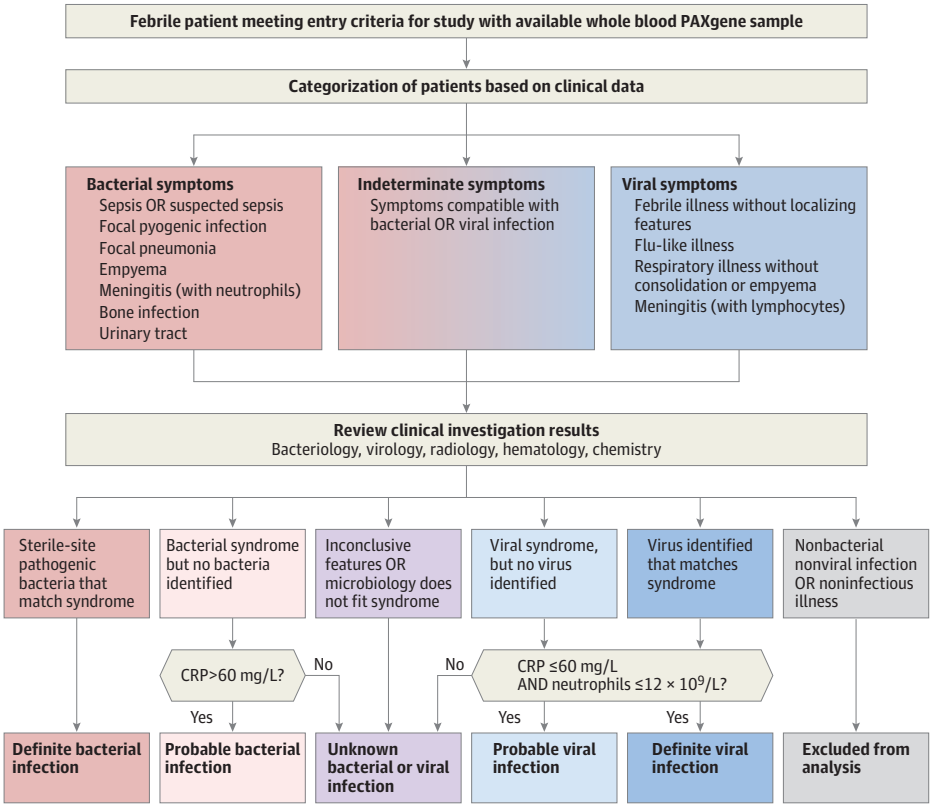
Diagnostic Process

All patients underwent routine investigations as part of clinical care, including complete blood cell count and differential, C-reactive protein level, blood chemistries, blood and urine cultures, and cerebrospinal fluid analysis where indicated. Throat swabs were cultured for bacteria, and viral diagnostics undertaken on nasopharyngeal aspirates using multiplex polymerase chain reaction for common respiratory viruses. Chest radiographs were undertaken as clinically indicated. Patients were assigned to diagnostic groups using predefined criteria (Figure 2). The definite bacterial infection group included only patients with culture-confirmed infection, and the definite viral infection group included only patients with culture, molecular, or immunofluorescent test–confirmed viral infection and no features of coexisting bacterial infection. Children in whom definitive diagnosis was not established (indeterminate infection) were categorized into probable bacterial infection, unknown bacterial or viral infection, and probable viral

Figure 1. Study Overview

Patient recruitment and subsequent selection for microarray analysis. DB indicates definite bacterial; DV, definite viral; HC, healthy control; HSP, Henoch-Schönlein purpura; JIA, juvenile idiopathic arthritis; PB, probable bacterial; PV, probable viral; SLE, systemic lupus erythematosus; U, unknown. [a]Controls not used for signature discovery. [b]Controls not used for signature validation. [c]Definite meningococcal disease.

Figure 2. Classification of Patients Into Diagnostic Groups



Febrile children with infections were recruited to the Immunopathology of Respiratory, Inflammatory and Infectious Disease Study and were classified into diagnostic groups as described in methods. To convert C-reactive protein (CRP) values to nmol/L, multiply by 0.9524.

infection groups based on level of clinical suspicion (Figure 2). Detection of virus did not prevent inclusion in the definite bacterial, probable bacterial, or unknown infection groups, because bacterial infection can occur in children co-infected with viruses.

Peripheral Blood Gene Expression by Microarray

Whole blood was collected into PAXgene blood RNA tubes (PreAnalytiX), frozen, and later extracted. Gene expression was analyzed on Illumina microarrays. Additional details of microarray method, quality control, and analysis are provided in the eAppendix (eMethods, eStatistical Methods, and eFigure 1 in the Supplement).

Statistical Analysis

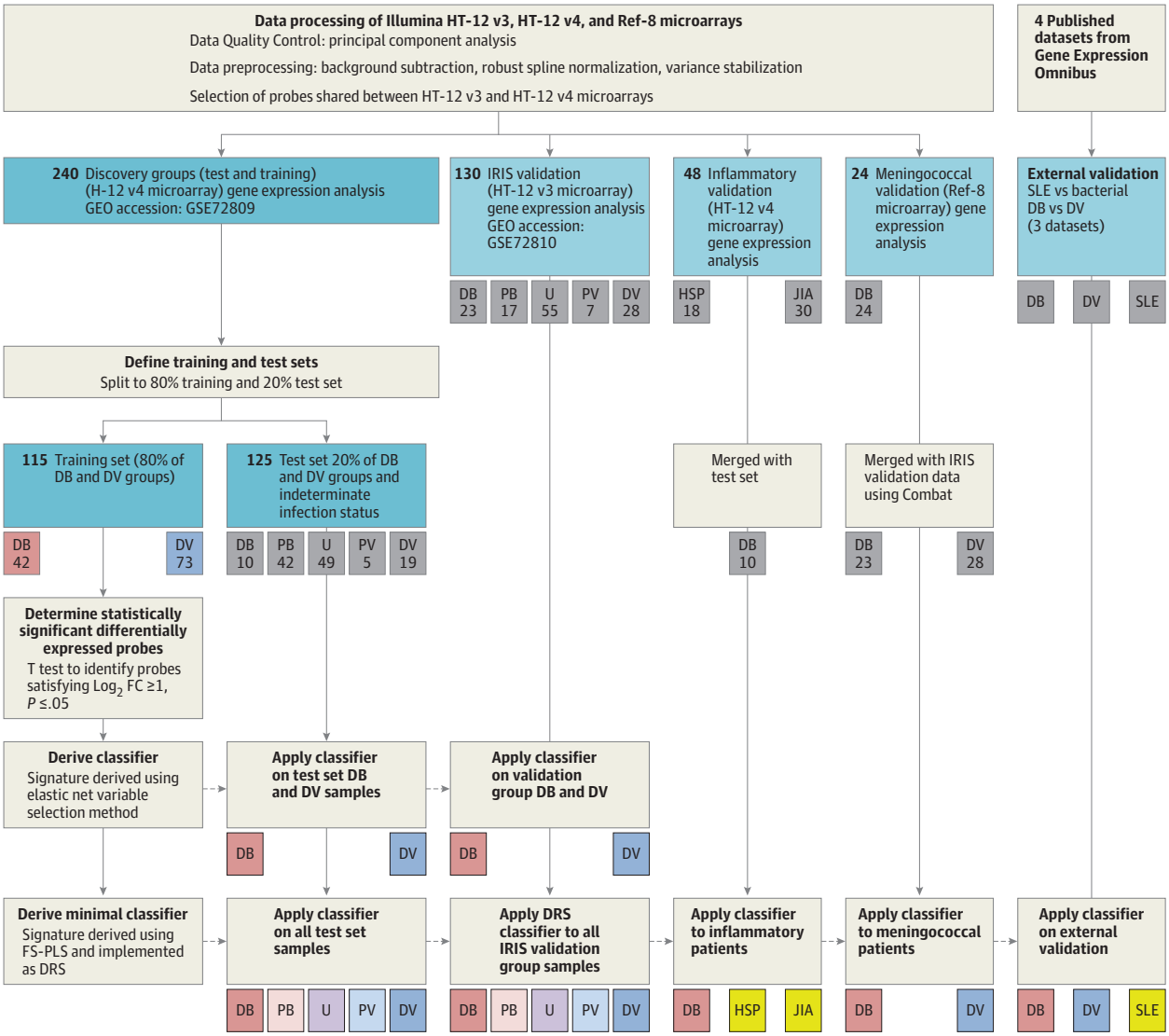
Transcript Signature Discovery

Expression data were analyzed using R version 3.1.2 (R Project for Statistical Computing). Patients with definite bacterial or viral infection in the discovery group were randomly assigned to training and test sets (80% and 20% of the patients, respectively), and significantly differentially expressed transcripts distinguishing definite bacterial infection from definite viral infection were identified in the training set (Figure 3). A linear model was fitted conditional on recruitment site, and moderated *t* statistics were calculated for each transcript. The *P* values obtained were corrected for multiple testing using the Benjamini-Hochberg false discovery rate method.[19] Logistic regression with variable selection was applied to the significantly differentially expressed transcripts (absolute $\log_2$-fold change >1 and 2-sided $P < .05$) using elastic net (a variable selection algorithm that selects sparse diagnostic transcript signatures—see eMethods and eFigure 2 in the Supplement).[20]

To further reduce the number of transcripts in the diagnostic signatures, a novel variable selection method was used that eliminates highly correlated transcripts: forward selection-partial least squares (see eAppendix in the Supplement). The disease risk score (DRS) method[21] was applied to the resulting minimal multitranscript signature to translate it into a single value that could be assigned to each individual, to form the basis of a simple diagnostic test.[11,21] The DRS method calculates a patient score by adding the total intensity of the upregulated transcripts (relative to comparator group) and subtracting the total intensity of the downregulated transcripts (relative to comparator group). The signatures identified in the discovery group were externally validated on previously published validation groups,[13] additional patient groups with meningococcal disease and inflammatory diseases, and published datasets (3 pediatric, 1 adult) (Figure 3).

To evaluate the predictive accuracy of the DRS and of models derived after variable selection analysis, point and interval metrics were calculated using the pROC package in R.[22] Results obtained using elastic net and DRS models were compared with reference-standard clinically assigned diagnoses (Figure 2). The area under the receiver operating characteristic

Figure 3. Analysis Workflow



Overall study pipeline showing sample handling, derivation of test and training sets, data processing, and analysis pipeline including application of 38-transcript elastic net classifier and 2-transcript disease risk score (DRS) classifier, to the group test set, the validation group data, and published (external) validation datasets. DB indicates definite bacterial; DV, definite viral; FC, fold change; FS-PLS, forward selection–partial least squares; HSP, Henoch-Schönlein purpura; JIA, juvenile idiopathic arthritis; PB, probable bacterial; PV, probable viral; SDE, significantly differentially expressed; SLE, systemic lupus erythematosus; U, unknown.

curve (AUC), sensitivity, and specificity were reported. Confidence intervals (95%) were calculated to measure the reliability of estimates.

Results

Two hundred forty patients (median age, 19 months; 62% male) were recruited to the discovery group between 2009-2013 in the United Kingdom (189 patients), Spain (16), and the United States (35). The definite bacterial infection group included 52 patients, of whom 36 (69%) required intensive care and 10 died. In the definite viral infection group of 92 patients, 32 (35%) required intensive care and none died (**Table 1**). The patients with bacterial and viral infection were subdivided into 80% (training set) and 20% (test set) (Figure 1 and Figure 3). The test set also included 96 children whose infection was not definitively diagnosed (indeterminate infection) (Figure 1 and Figure 3). The validation groups comprised 130 UK and Spanish children (median age, 17 months; 57% male) previously recruited[13] (IRIS validation; 23 with definite bacterial infection, 28 with definite viral infection, and 79 with indeterminate infection) and 72 additional validation children—25 from the United Kingdom, 30 from the Netherlands, and 17 from the United States (24 with meningococcal infection, 30 with juvenile idiopathic arthritis, and 18 with Henoch-Schönlein

Table 1. Demographic and Clinical Characteristics of the Study Groups

Characteristics	Discovery Group			IRIS Validation Group		
	Definite Bacterial Infection (n = 52)	Definite Viral Infection (n = 92)	Indeterminate Infection[a] (n = 96)	Definite Bacterial Infection (n = 23)	Definite Viral Infection (n = 28)	Indeterminate Infection[a] (n = 79)
Age, median (IQR), mo	22 (9-46)	14 (2-39)	27 (7-71)	22 (13-52)	18 (7-48)	15 (2-44)
Male, No. (%)	22 (42)	65 (71)	62 (65)	10 (43)	17 (61)	47 (59)
White race, No./total (%)[b]	35/48 (73)	46/87 (53)	47/85 (55)	12/22 (55)	14/27 (51)	42/71 (59)
Time from symptoms to blood sampling, median (IQR), d	5 (2-8.8)	4.5 (3.0-6.0)	5 (4.8-8)	4 (2.5-8)	3.5 (2.8-5.3)	4 (3-7)
Intensive care, No. (%)	36 (69)	32 (35)	57 (59)	13 (57)	7 (23)	42 (53)
Deaths, No.	10	0	2	1	1	8
C-reactive protein, median (IQR), mg/L[c]	176 (98-275)	16 (6-27)	102 (47-176)	217 (168-285)	7 (1-20)	67 (25-128)
Blood cell differential, median (IQR), %						
Neutrophils	75 (49-85)	50 (36-63)	63 (46-79)	82 (71-88)	53 (41-69)	64 (43-82)
Lymphocytes	19 (10-36)	34 (20-44)	22 (15-42)	15 (8-23)	32 (26-48)	30 (14-42)
Monocytes	5 (3-8)	10 (4-14)	6 (2-12)	3 (0-7)	7 (5-10)	5 (2-8)
Main clinical syndrome, No.						
Bone, joint, soft tissue infection	5	0	0	1	0	0
Fever without source/sepsis	21	7	9	5	2	6
Gastroenteritis	0	0	1	0	1	2
Meningitis/encephalitis	14	3	3	5	1	1
Respiratory tract, upper + lower	10	81	83	11	23	68
Other	2	1	0	1	1	2
Virus detected, No./total (%)[d]	22/34 (65)	92/92 (100)	62/87 (71)	8/13 (62)	28/28 (100)	52/77 (68)

Abbreviations: IQR, interquartile range; IRIS, Immunopathology of Respiratory, Inflammatory and Infectious Disease Study.

SI conversion factor: To convert C-reactive protein values to nmol/L, multiply by 0.9524.

[a] The indeterminate infection group in the discovery group comprised 42 probable bacterial, 49 unknown bacterial or viral, and 5 probable viral infections. The indeterminate infection group in the validation group comprised 17 probable bacterial, 55 unknown bacterial or viral, and 7 probable viral infections.

[b] Self-reported.

[c] Maximum value of C-reactive protein in illness is reported.

[d] Denominator denotes number of patients with viral investigations.

purpura) (Figure 1 and Figure 3). The numbers in each diagnostic category in the discovery, IRIS validation, and additional validation groups and their clinical features are shown in Table 1 and in eTable 1 in the Supplement. Details of the types of infection are shown in eTable 2 in the Supplement. Gene expression profiles of children in the discovery group clustered together on principal component analysis (eFigure 1 in the Supplement).

Identification of Minimal Transcript Signatures

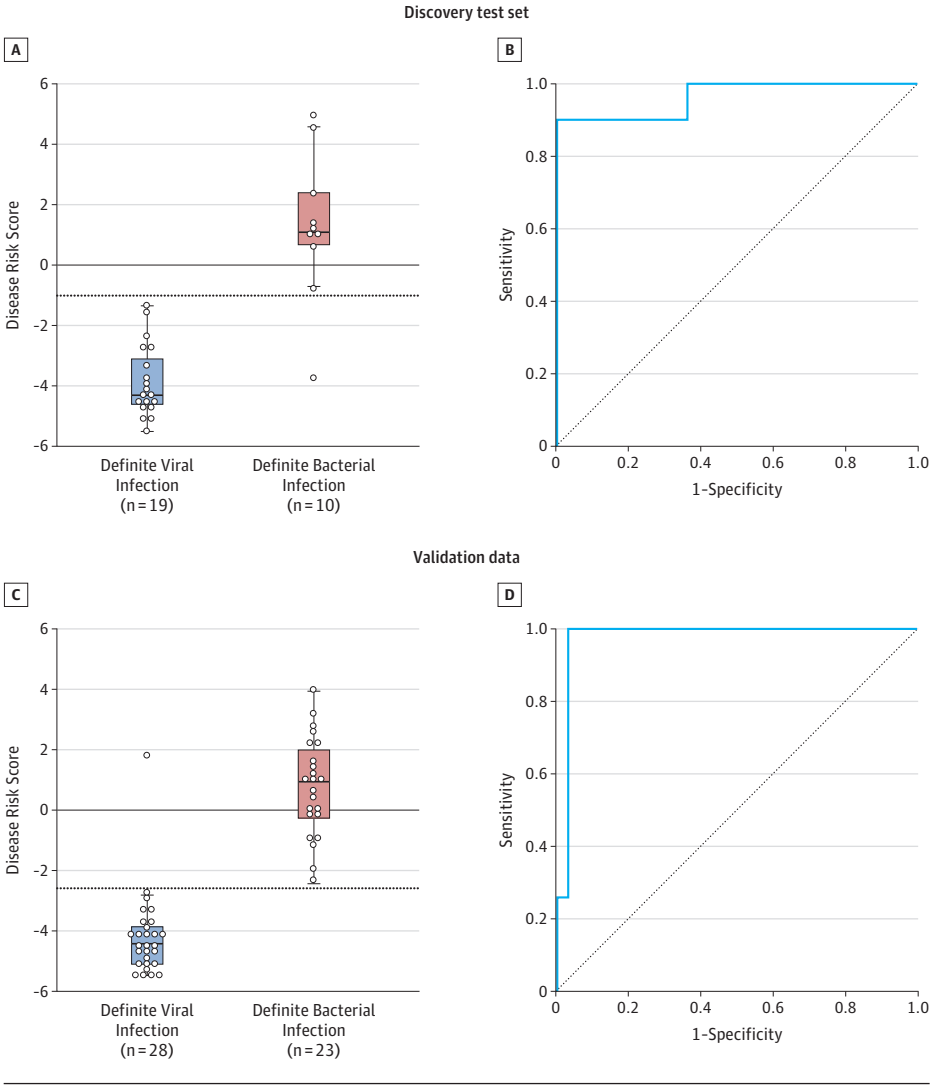
Of the 8565 transcripts differentially expressed between bacterial and viral infections, 285 were identified as potential biomarkers after applying filters based on log fold change and statistical significance (see Methods). Variable selection using elastic net identified 38 transcripts (eTable 3 in the Supplement) as best discriminators of bacterial and viral infection in the discovery test set, with sensitivity of 100% (95% CI, 69%-100%) and specificity of 95% (95% CI, 84%-100%) (eTable 4 in the Supplement). In the validation group, this signature had an AUC of 98% (95% CI, 94%-100%), sensitivity of 100% (95% CI, 85%-100%), and specificity of 86% (95% CI, 71%-96%) for distinguishing bacterial from viral infection (eTable 4 and eFigures 2 and 3 in the Supplement). The putative function of the 38 transcripts in our signature, as defined by gene ontology, is shown in eTable 5 in the Supplement.

After using the novel forward selection process to remove highly correlated transcripts, a 2-transcript signature was identified that distinguished bacterial from viral infections, including interferon-induced protein 44-like (*IFI44L*, RefSeq NM_006820.1), and family with sequence similarity 89, member A (*FAM89A*, RefSeq NM_198552.1). Both transcripts were included in the 38-transcript signature.

Implementation of a DRS

The expression data of both genes in the signature was combined into a single DRS for each patient, using the reported DRS method, which simplifies application of multitranscript signatures as a diagnostic test.[21] The sensitivity of the DRS was 86% (95% CI, 74%-95%) in the discovery group training set, 90% (95% CI, 70%-100%) in the discovery group test set, and 100% (95% CI, 85%-100%) in the validation data; specificity in the validation data was 96.4% (95% CI, 89.3%-100%) (**Figure 4**, panels A-D; eFigure 4 and eTable 4 in the Supplement). Expression of *IFI44L* was increased in patients with viral infection and *FAM89A* was increased in patients with bacterial infection, relative to healthy children (eTable 3 in

Figure 4. DRS and ROC Curves Based on the 2-Transcript Signature Applied to Definite Bacterial and Viral Infection Groups



Classification performance and receiver operating characteristic (ROC) curve based on the 2-transcript disease risk score (DRS) (the combined *IFI44L* and *FAM89A* expression values), in the definite bacterial and viral infection groups of the discovery group test set (20% of the total discovery group), and the IRIS validation group data. Horizontal lines in boxes indicate medians; lower and upper edges of boxes, interquartile ranges; whiskers, ≤1 times the interquartile range. Boxes and whiskers plotted using *boxplot* in R. Sensitivity, specificity, and area under the ROC curve are reported in eTable 4 in the Supplement. Horizontal dotted line indicates the DRS threshold that separates patients predicted as bacterial (above the line) or viral (below the line), as determined by the point on the ROC curve that maximized sensitivity and specificity.

the Supplement). The summary of diagnostic test accuracy, including STARD flow diagrams, is shown in eFigure 5 in the Supplement.

For additional validation, the 2-transcript signature was applied to patients with meningococcal disease (eFigure 6 in the Supplement) and inflammatory diseases (juvenile idiopathic arthritis and Henoch-Schönlein purpura). Bacterial infection was identified with a sensitivity of 91.7% (95% CI, 79.2%-100%) for patients with meningococcal disease and 90.0% (95% CI, 70.0%-100%) for those with inflammatory diseases and with a specificity of 96.0% (95% CI, 88.0%-100%) and 95.8% (95% CI, 89.6%-100%), respectively. When applied to 4 published datasets for children and adults with bacterial or viral infection and inflammatory disease (pediatric systemic lupus erythematosus),[12,15,17,18] the 2-transcript signature distinguished bacterial infection from viral infection and inflammatory disease in all these datasets, with AUCs ranging from 89.2% to 96.6% (eTable 6 and eFigures 7 and 8 in the Supplement).

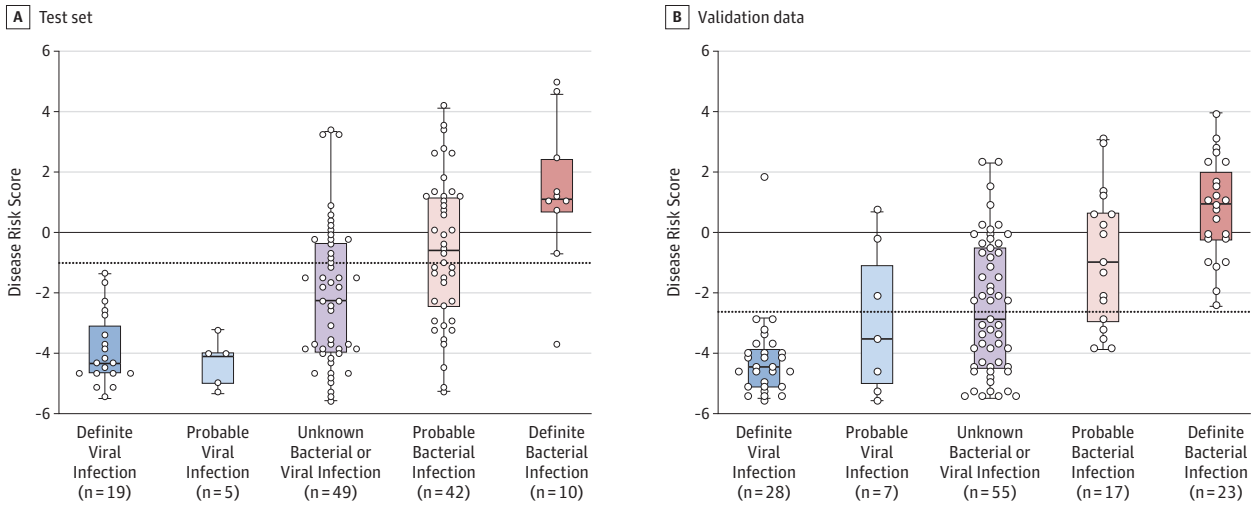
Effect of Viral and Bacterial Co-infection

The effect of viral co-infection on the signatures was investigated (Table 1). In the definite bacterial infection group, 30 of 47 patients tested (64%) had a virus isolated from nasopharyngeal samples. There was no significant difference in DRS between those with and without viral co-infection.

DRS in Patients With Indeterminate Infection Status

The classification performance of the DRS was investigated in patients with indeterminate viral or bacterial infection status. Patients were separated into those with clinical features strongly suggestive of bacterial infection (probable bacterial infection group), those with features consistent with either bacterial or viral infection (unknown infection group), and those with clinical features and results suggestive of viral infection (probable viral infection group) (Figure 2). The probable bacterial and unknown infection groups included patients with DRS values that indicated

Figure 5. Performance of the 2-Transcript DRS Signature in Indeterminate Infection Group



Classification performance of the 2-transcript disease risk score (DRS) (the combined *IFI44L* and *FAM89A* expression values) in the indeterminate groups of probable bacterial, probable viral, and unknown infection of the discovery test and IRIS validation sets. Horizontal lines in boxes indicate medians; lower and upper edges of boxes, interquartile ranges; whiskers, ≤1 times the interquartile range. Boxes and whiskers plotted using *boxplot* in R. Horizontal dotted line indicates the DRS threshold ($threshold_{test\_data\ set}$ = −1.03; $threshold_{validation\_data\ set}$ = −2.63) that separates patients predicted as bacterial (above the line) or viral (below the line), as determined by the point in the receiver operating characteristic curve that maximized sensitivity and specificity. For the discovery group test set, the training threshold was used.

Table 2. Comparison of Disease Risk Score Prediction With Antibiotic Treatment[a]

	No. (%)				
	Definite Bacterial Infection	Probable Bacterial Infection	Unknown Bacterial or Viral Infection	Probable Viral Infection	Definite Viral Infection
Patients with information on antibiotic use[b]	28 (84.8)	49 (83.1)	77 (74.0)	10 (83.3)	39 (83.0)
Patients receiving antibiotics	28 (100)	49 (100)	73 (94.8)	7 (70)	33 (84.6)
Patients receiving antibiotics and suggested by DRS to have bacterial infection	27 (96.4)	32 (65.3)	29 (39.7)	2 (28.6)	1 (3.0)

Abbreviation: DRS, disease risk score.

[a] Comparison of the proportion of patients in the combined discovery group test set and validation group receiving antibiotics, and the proportion of predicted bacterial, as predicted by the DRS (the combined *IFI44L* and *FAM89A* expression values).

[b] The denominator is 33 definite bacterial, 59 probable bacterial, 104 unknown bacterial or viral, 12 probable viral, and 47 definite viral infections.

viral infection, despite having clinical features that justified initiation of antibiotics by the clinical team. The median DRS showed a gradient of assignment that followed the degree of certainty in the clinical diagnosis, although many of the indeterminate infection group DRS values overlapped with those of the definite bacterial and definite viral infection groups (**Figure 5**A, Figure 4B).

DRS assignment as viral or bacterial was compared with clinical variables in the indeterminate infection group (eTable 7 in the Supplement). Measurement of C-reactive protein (CRP) levels is widely used to aid distinction of bacterial from viral infection and was included in the categorization of definite viral, probable bacterial, and probable viral infection in this study; patients categorized as having bacterial infection by DRS had higher CRP levels than those categorized as having viral infection (median, 101 [interquartile range {IQR}, 48-192] mg/L vs 71 [IQR, 27-120] mg/L; P = .02 [to convert values to nmol/L, multiply by 0.9524]). They also had increased incidence of shock (P = .006), requirement for ventilator support (P = .048), and intensive care admission (P = .07). There was a nonsignificant increase in white cell and neutrophil counts in patients assigned by DRS as having bacterial or viral infection (median, 14.1 [IQR, 8.3-19.4] and 11.1 [IQR, 7.3-16.0] for white cells [P = .08]; 8.7 [IQR, 5.0-13.8] and 6.8 [IQR, 3.5-11.4] for neutrophils [P = .11]).

Antibiotic Use

The number of children treated with antibiotics was compared with DRS prediction of bacterial or viral infection. There were high rates of antibiotic use in all groups, including greater than 90% in the unknown infection group. The high rate of antibiotic use in the indeterminate infection group contrasted with the low numbers predicted to have bacterial infection by both DRS and clinical assignment (**Table 2**).

Illness Severity and Duration

The study recruited a high proportion of seriously ill patients needing intensive care, thus raising concern that selection bias

might have influenced performance of the signature. To exclude bias based on severity or duration of illness, performance of the DRS was evaluated after stratification of patients into those with milder illness or severe illness requiring intensive care and by duration of reported illness before presentation. The DRS distinguished bacterial from viral infection in both severe and milder groups (eFigure 9 in the Supplement) and irrespective of day of illness (eFigure 10 in the Supplement).

Discussion

This study identified a host whole blood RNA transcriptomic signature that distinguished bacterial from viral infection with 2 gene transcripts. The signature also distinguished bacterial infection from childhood inflammatory diseases, systemic lupus erythematosus, juvenile idiopathic arthritis, and Henoch-Schönlein purpura and discriminated bacterial from viral infection in published adult studies.[12,15,17,18] The results extend previous studies suggesting that bacterial and viral infections have different signatures.[12,13,17,23]

The transcripts identified in the 38-transcript elastic net signature comprise a combination of transcripts up-regulated by viruses or by bacteria. The 2 transcripts *IFI44L* and *FAM89A* in the smaller signature show reciprocal expression in viral and bacterial infection. *IFI44L* has been reported to be up-regulated in antiviral responses mediated by type I interferons,[24] and *FAM89A* was reported as elevated in children with septic shock.[25]

An obstacle in the development of improved tests to distinguish bacterial from viral infection is the lack of a reference standard. Some studies include patients with clinically diagnosed bacterial infection who have features of bacterial infection but whose cultures remain negative. Negative cultures may reflect prior antibiotic use, low numbers of bacteria, or inaccessible sites of infection. If patients with indeterminate status are included in biomarker discovery, there is a risk that the identified biomarker will not be specific for true infection. This study adopted the rigorous approach of identifying the signature in culture-confirmed cases and using the signature to explore likely proportions of true infection in the indeterminate infection group.

The proportion of children predicted by DRS signature to have bacterial infection follows the level of clinical suspicion (greater in the probable bacterial infection group and less in the probable viral infection group), thus supporting the hypothesis that the signatures may be an indication of the true proportion of bacterial infection in each group. Furthermore, a higher proportion of patients in the indeterminate infection group, assigned as bacterial by the signature (probable infection and unknown infection groups), had clinical features normally associated with severe bacterial infection, including increased need for intensive care, higher neutrophil counts, and higher CRP levels, suggesting that the signature may be providing additional clues to the presence of bacterial infection.

The decision to initiate antibiotics in febrile children is largely driven by concern about missing bacterial infection. A test that correctly distinguishes children with bacterial infection from those with viral infections would reduce inappropriate antibiotic prescription and investigation. The DRS suggests that many children who were prescribed antibiotics did not have a bacterial illness. If the score reflects the true likelihood of bacterial infection, its implementation could reduce unnecessary investigation, hospitalization, and treatment with antibiotics. Confirmation that the DRS provides an accurate estimate of bacterial infection in the large group of patients with negative cultures, for whom there is no reference standard, can only be achieved in prospective clinical trials. Careful consideration will be needed to design an ethically acceptable and safe trial in which observation without antibiotic administration is undertaken for febrile children suggested by DRS to be at low risk of bacterial infection.

In comparison with the high frequency of common viral infections in febrile children presenting to health care, inflammatory and vasculitic illness are very rare.[26-29] However, children presenting with inflammatory or vasculitic conditions commonly undergo extensive investigation to exclude bacterial infection and treatment with antibiotics before the correct diagnosis is made. Although children with inflammatory conditions were not included in the discovery process, the 2-transcript signature was able to distinguish bacterial infection from systemic lupus erythematosus, juvenile idiopathic arthritis, and Henoch-Schönlein purpura. Additional studies including a wider range of inflammatory diseases are needed to assess use of the signature for excluding bacterial infection in inflammatory diseases.

This study has a number of important limitations. The cross-sectional design aimed to recruit equal numbers of children with bacterial and viral infections. The numbers of children recruited thus did not reflect the usual frequency of bacterial infection in febrile children presenting to health care facilities. Further studies of a test based on the 2-transcript signature in unselected febrile children will be needed to provide information on positive and negative predictive performance of the test.

A second limitation is that validation of the signatures was undertaken in groups that included a high proportion of patients requiring intensive care, and with a relatively narrow spectrum of pathogens, which may not reflect the spectrum of infection in other settings. The signature performed well, both in patients with less severe infection and those admitted to intensive care, and performance was not influenced by duration of illness. However, further studies will be needed to evaluate the DRS signature in less severely ill patients with a wider range of infections or in settings such as emergency departments or outpatient offices. Another limitation is the use in validation of published datasets and data obtained using different microarray platforms. Although batch effects were minimized computationally, additional studies are needed in which gene expression is measured on identical platforms.

A major challenge in using transcriptomic signatures for diagnosis is the translation of multitranscript signatures into

clinical tests suitable for use in hospital laboratories or at the bedside. The DRS signature, distinguishing viral from bacterial infections with only 2 transcripts, has potential to be translated into a clinically applicable test using current technology such as polymerase chain reaction.[30] Furthermore, new methods for rapid detection of nucleic acids, including nanoparticles and electrical impedance, have potential for low-cost, rapid analysis of multitranscript signatures.

Conclusions

This study provides preliminary data regarding test accuracy of a 2-transcript host RNA signature discriminating bacterial from viral infection in febrile children. Further studies are needed in diverse groups of patients to assess accuracy and clinical utility of this test in different clinical settings.

ARTICLE INFORMATION

Correction: This article was corrected online on February 7, 2017, for errors in the text, Figure 1, and the online Supplement.

Author Affiliations: Section of Paediatrics, Division of Infectious Diseases, Department of Medicine, Imperial College London, London, United Kingdom (Herberg, Kaforou, Wright, Shailes, Eleftherohorinou, Hoggart, Carter, Janes, Gormley, Levin); Translational Paediatrics and Infectious Diseases Section, Department of Paediatrics, Hospital Clínico Universitario de Santiago, Santiago de Compostela, Galicia, Spain (Cebey-López, Salas, Martinón-Torres); Grupo de Investigación en Genética, Vacunas, Infecciones y Pediatría (GENVIP), Healthcare Research Institute of Santiago de Compostela and Universidade de Santiago de Compostela, Spain (Cebey-López, Salas, Martinón-Torres); Department of Pediatrics, University of California San Diego, La Jolla (Shimizu, Tremoulet, Kanegaye, Burns); Rady Children's Hospital San Diego, San Diego, California (Shimizu, Tremoulet, Kanegaye, Burns); Emma Children's Hospital, Department of Paediatric Haematology, Immunology and Infectious Diseases, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands (Barendregt, Kuijpers); Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses, and Instituto de Ciencias Forenses, Grupo de Medicina Xenómica, Spain (Salas); Department of Paediatrics, University of Oxford, Oxford, United Kingdom (Pollard); National Institute of Health Research Oxford Biomedical Research Centre, Oxford, United Kingdom (Pollard); National Institute of Health Research Wellcome Trust Clinical Research Facility, University of Southampton, Southhampton, United Kingdom (Faust); University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom (Faust, Patel); Sanquin Research and Landsteiner Laboratory, Amsterdam Medical Centre, Department of Blood Cell Research, University of Amsterdam, Amsterdam, the Netherlands (Kuijpers); Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland, Australia (Coin).

Author Contributions: Drs Herberg and Levin had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Herberg, Kaforou, and Wright contributed equally to this work.
Concept and design: Herberg, Kaforou, Wright, Levin.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Herberg, Kaforou, Wright, Levin.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Herberg, Kaforou, Wright, Eleftherohorinou, Hoggart, Carter, Janes, Barendregt, Levin.
Obtaining funding: Herberg, Kaforou, Wright, Tremoulet, Barendregt, Kuijpers, Burns, Levin.
Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Kanegaye reported receiving a grant from the Gordon and Marilyn Macklin Foundation. Dr Faust reported receiving research nursing support from the National Institute of Health Research Clinical Research Network; and grants from Pfizer, Sanofi, GlaxoSmithKline, Novartis, Alios, and Ablynx. Dr Coin reported holding a patent pending for disease risk score methodology. Dr Levin reported preparing a patent application for the 2-gene signature used in this study. No other authors reported disclosures.
Funding/Support: This work was supported by the Imperial College Comprehensive Biomedical Research Centre (DMPED P26077); National Institute of Health Research (NIHR) Senior Investigator award (Dr Levin); Great Ormond St Hospital Charity (V1401) (Dr Wright); European Union's Seventh Framework Program (EC-GA 279185) (EUCLIDS) (Dr Herberg); Imperial College-Wellcome Trust Antimicrobial Research Collaborative (ARC) Early Career Fellowship (RSRO 54990) (Dr Kaforou); Spanish Research Program (FIS; PI10/00540 and Intensificación actividad investigadora of National Plan I + D + I and FEDER funds) and Regional Galician funds (Promotion of Research Project 10 PXIB 918 184 PR) (Dr Martinón-Torres); Southampton NIHR Wellcome Trust Clinical Research Facility and NIHR Wessex Local Clinical Research Network; and Academic Medical Centre Amsterdam MD/PhD program 2013 (Ms Barendregt). The UK meningococcal disease cohort was established with grant support from the Meningitis Research Foundation (United Kingdom); the inflammatory disease cohort was supported by a Macklin Foundation grant (Dr Burns), National Institutes of Health grant U54-HL108460 (Dr Burns); and The Hartwell Foundation and The Harold Amos Medical Faculty Development Program/Robert Wood Johnson Foundation (Dr Tremoulet).
Role of the Funders/Sponsors: The funders/sponsors had no role in the design or conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.
Group Information: Members of the IRIS Consortium: *Imperial College London (United Kingdom):* Michael J. Carter, Lachlan J. M. Coin, Hariklia Eleftherohorinou, Erin Fitzgerald, Stuart Gormley, Jethro A. Herberg, Clive Hoggart, David Inwald, Victoria A. Janes, Kelsey D. J. Jones, Myrsini Kaforou, Sobia Mustafa, Simon Nadel, Stéphane Paulus, Nazima Pathan, Joanna Reid, Hannah Shailes, Victoria J. Wright, Michael Levin. *Southampton (United Kingdom):* Saul Faust, Jenni McCorkill, Sanjay Patel. *Oxford (United Kingdom):* Andrew J. Pollard, Louise Willis, Zoe Young. *Micropathology Ltd (United Kingdom):* Colin Fink, Ed Sumner. *University of California, San Diego and Rady Children's Hospital, La Jolla, California (United States):* John T. Kanegaye, Chisato Shimizu, Adriana Tremoulet, Jane Burns. *Hospital Clínico Universitario de Santiago (Spain):* Miriam Cebey-López, Antonio Salas, Antonio Justicia Grande, Irene Rivero, Alberto Gómez Carballa, Jacobo Pardo Seco, José María Martinón Sánchez, Lorenzo Redondo Collazo, Carmen Rodríguez-Tenreiro, Lucia Vilanova Trillo, Federico Martinón-Torres. *Emma Children's Hospital, Department of Paediatric Haematology, Immunology & Infectious Disease, Academic Medical Centre, University of Amsterdam (the Netherlands):* Anouk M. Barendregt, Merlijn van den Berg, Dieneke Schonenberg, Taco Kuijpers. *London School of Hygiene and Tropical Medicine, London (United Kingdom):* Martin L. Hibberd. *Newcastle upon Tyne Foundation Trust Hospitals and Newcastle University, Newcastle upon Tyne (United Kingdom):* Marieke Emonts.
Additional Information: People linked to the IRIS consortium through membership of the GENDRES consortium (http://www.gendres.org) are listed in the eAppendix in the Supplement.
Reproducible Research Statement: The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar R, Domrachev M, Lash AE. et al. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30 (1):207-210) and are accessible through GEO Series accession number GSE72829 (http://www.ncbi.nlm.nih.gov/geo/).
Additional Contributions: We wish to thank the children and families who have participated in the study, Jayne Dennis, PhD (St George's University of London), for her support and advice for the microarray experiments, Chenxi Zhou, MSc (University of Queensland), for his help with the feature selection software, and the clinical teams for their assistance in recruiting patients to the study. These persons did not receive direct compensation for their work.

REFERENCES

1. Iroh Tam PY, Bernstein E, Ma X, Ferrieri P. Blood culture in evaluation of pediatric community-acquired pneumonia: a systematic review and meta-analysis. *Hosp Pediatr*. 2015;5(6):324-336.

2. Martin NG, Sadarangani M, Pollard AJ, Goldacre MJ. Hospital admission rates for meningitis and septicaemia caused by *Haemophilus influenzae*, *Neisseria meningitidis*, and *Streptococcus pneumoniae* in children in England over five

decades: a population-based observational study. *Lancet Infect Dis*. 2014;14(5):397-405.

3. Colvin JM, Muenzer JT, Jaffe DM, et al. Detection of viruses in young children with fever without an apparent source. *Pediatrics*. 2012;130(6):e1455-e1462.

4. Craig JC, Williams GJ, Jones M, et al. The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: prospective cohort study of 15 781 febrile illnesses. *BMJ*. 2010;340:c1594.

5. Nijman RG, Vergouwe Y, Thompson M, et al. Clinical prediction model to aid emergency doctors managing febrile children at risk of serious bacterial infections: diagnostic study. *BMJ*. 2013;346:f1706.

6. Sadarangani M, Willis L, Kadambari S, et al. Childhood meningitis in the conjugate vaccine era: a prospective cohort study. *Arch Dis Child*. 2015;100(3):292-294.

7. Le Doare K, Nichols AL, Payne H, et al; CABIN Network. Very low rates of culture-confirmed invasive bacterial infections in a prospective 3-year population-based surveillance in Southwest London. *Arch Dis Child*. 2014;99(6):526-531.

8. Warhurst G, Dunn G, Chadwick P, et al. Rapid detection of health-care–associated bloodstream infection in critical care using multipathogen real-time polymerase chain reaction technology: a diagnostic accuracy study and systematic review. *Health Technol Assess*. 2015;19(35):1-142.

9. Cebey-López M, Herberg J, Pardo-Seco J, et al; GENDRES Network. Viral co-infections in pediatric patients hospitalized with lower tract acute respiratory infections. *PLoS One*. 2015;10(9):e0136526.

10. Berkley JA, Munywoki P, Ngama M, et al. Viral etiology of severe pneumonia among Kenyan infants and children. *JAMA*. 2010;303(20):2051-2057.

11. Anderson ST, Kaforou M, Brent AJ, et al; ILULU Consortium; KIDS TB Study Group. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med*. 2014;370(18):1712-1723.

12. Hu X, Yu J, Crosby SD, Storch GA. Gene expression profiles in febrile children with defined viral and bacterial infection. *Proc Natl Acad Sci U S A*. 2013;110(31):12792-12797.

13. Herberg JA, Kaforou M, Gormley S, et al. Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. *J Infect Dis*. 2013;208(10):1664-1668.

14. Mejias A, Dimo B, Suarez NM, et al. Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. *PLoS Med*. 2013;10(11):e1001549.

15. Ramilo O, Allman W, Chung W, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood*. 2007;109(5):2066-2077.

16. Pathan N, Hemingway CA, Alizadeh AA, et al. Role of interleukin 6 in myocardial dysfunction of meningococcal septic shock. *Lancet*. 2004;363(9404):203-209.

17. Suarez NM, Bunsow E, Falsey AR, Walsh EE, Mejias A, Ramilo O. Superiority of transcriptional profiling over procalcitonin for distinguishing bacterial from viral lower respiratory tract infections in hospitalized adults. *J Infect Dis*. 2015;212(2):213-222.

18. Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010;466(7309):973-977.

19. Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289-300.

20. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67:301-320.

21. Kaforou M, Wright VJ, Oni T, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med*. 2013;10(10):e1001538.

22. 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.

23. Zaas AK, Burke T, Chen M, et al. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. *Sci Transl Med*. 2013;5(203):203ra126.

24. Schoggins JW, Wilson SJ, Panis M, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*. 2011;472(7344):481-485.

25. Wong HR, Cvijanovich N, Lin R, et al. Identification of pediatric septic shock subclasses based on genome-wide expression profiling. *BMC Med*. 2009;7:34.

26. Gurion R, Lehman TJ, Moorthy LN. Systemic arthritis in children: a review of clinical presentation and treatment [published online December 25, 2011]. *Int J Inflammation*. doi:10.1155/2012/271569.

27. Rabizadeh S, Dubinsky M. Update in pediatric inflammatory bowel disease. *Rheum Dis Clin North Am*. 2013;39(4):789-799.

28. Thierry S, Fautrel B, Lemelle I, Guillemin F. Prevalence and incidence of juvenile idiopathic arthritis: a systematic review. *Joint Bone Spine*. 2014;81(2):112-117.

29. Gardner-Medwin JM, Dolezalova P, Cummins C, Southwood TR. Incidence of Henoch-Schönlein purpura, Kawasaki disease, and rare vasculitides in children of different ethnic origins. *Lancet*. 2002;360(9341):1197-1202.

30. van den Kieboom CH, Ferwerda G, de Baere I, et al. Assessment of a molecular diagnostic platform for integrated isolation and quantification of mRNA in whole blood. *Eur J Clin Microbiol Infect Dis*. 2015;34(11):2209-2212.