

Published in final edited form as:

JAMA. 2016 August 23; 316(8): 835–845. doi:10.1001/jama.2016.11236.

Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children

Jethro A Herberg, PhD^{#1}, Myrsini Kaforou, PhD^{#1}, Victoria J Wright, PhD^{#1}, Hannah Shailes, BSc¹, Hariklia Eleftherohorinou, PhD¹, Clive J Hoggart, PhD¹, Miriam Cebey-Lopez, MSc², Michael J Carter, MRCPCH¹, Victoria A Janes, MD¹, Stuart Gormley, MRes¹, Chisato Shimizu, MD^{3,4}, Adriana H Tremoulet, MD^{3,4}, Anouk M Barendregt, BSc⁵, Antonio Salas, PhD^{2,6}, John Kanegaye, MD^{3,4}, Andrew J Pollard, PhD⁷, Saul N Faust, PhD^{8,9}, Sanjay Patel, FRCPCH⁹, Taco Kuijpers, PhD^{5,10}, Federico Martinon-Torres, PhD², Jane C Burns, MD^{3,4}, Lachlan JM Coin, PhD¹¹, and Michael Levin, FRCPCH¹

¹Section of Paediatrics, Division of Infectious Diseases, Department of Medicine, Imperial College London, London W2 1PG, UK ²Translational Paediatrics and Infectious Diseases section,

Corresponding Author: Prof. Michael Levin, Section of Paediatrics, Division of Infectious Diseases, Department of Medicine, Imperial College London, Norfolk Place, London, W2 1PG. Tel: +44 (0)20 7594 3990; Fax: +44 (0)20 7594 3984; m.levin@imperial.ac.uk.

Authors' contributions

Criteria 1: Study conception & design: JAH, MK, VJW, ML; Data acquisition: JAH, VJW, HS, MC-L, SG, CS, AHT, AMB, JK, SP; Data analysis: JAH, MK, VJW, HE, CJH, MJC, VAJ, ML; Data interpretation: JAH, MK, VJW, CJH, HE, AS, AJP, SNF, SP, TK, FMT, JCB, LJMC, ML

Criteria 2: Manuscript draft: JAH, MK, VJW, ML; Manuscript revision: All

Criteria 3: Final approval of manuscript: All

Criteria 4: Accountable for all aspects of the work: All.

Data access and responsibility: Jethro A Herberg and Michael Levin had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Analysis: Data analysis was carried out by Jethro A Herberg PhD, Myrsini Kaforou PhD¹, Victoria J Wright PhD¹, Clive J Hoggart PhD¹, Michael J Carter MRCPCH¹, Victoria A Janes MD¹, Hariklia Eleftherohorinou PhD¹, Anouk M Barendregt BSc^{6,7} and Michael Levin¹ FRCPCH (¹Section of Paediatrics, Division of Infectious Diseases, Department of Medicine, Imperial College London, London W2 1PG, UK).

Conflicts of interest: None

Collaborators: the study was carried out on behalf of the **IRIS Consortium** which includes the following members:

Imperial College London (UK) Michael J Carter, Lachlan JM Coin, Hariklia Eleftherohorinou, Erin Fitzgerald, Stuart Gormley, Jethro A. Herberg, Clive Hoggart, David Inwald, Victoria A. Janes, Kelsey DJ Jones, Myrsini Kaforou, Sobia Mustafa, Simon Nadel, Stéphane Paulus, Nazima Pathan, Joanna Reid, Hannah Shailes, Victoria J. Wright, Michael Levin; **Southampton (UK)** Saul Faust, Jenni McCorkill, Sanjay Patel; **Oxford (UK)** Andrew J Pollard, Louise Willis, Zoe Young; **Micropathology Ltd (UK)** Colin Fink, Ed Sumner; **University of California San Diego and Rady Children's Hospital, La Jolla, California (USA)** John T. Kanegaye, Chisato Shimizu, Adriana Tremoulet, Jane Burns; **Hospital Clínico Universitario de Santiago (Spain)** Miriam Cebey-Lopez, Antonio Salas, Antonio Justicia Grande, Irene Rivero, Alberto Gómez Carballa, Jacobo Pardo Seco, José María Martínón Sánchez, Lorenzo Redondo Collazo, Carmen Rodríguez-Tenreiro, Lucia Vilanova Trillo, Federico Martínón-Torres, **Emma Children's Hospital, Department of Paediatric Haematology, Immunology & Infectious Disease, Academic Medical Centre, University of Amsterdam, Amsterdam** Anouk M Barendregt, Merlijn van den Berg, Taco Kuijpers and Dienneke Schonenberg; **London School of Hygiene and Tropical Medicine, London (UK)** Martin L Hibberd; **Newcastle upon Tyne Foundation Trust Hospitals and Newcastle University, Newcastle upon Tyne, UK** Marieke Emonts.

Other disclaimers: none

Previous presentations of the data: the full data have not been presented

Compensation of other individuals involved: none

Reproducible Research Statement: The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO Series accession number GSE72829 (<http://www.ncbi.nlm.nih.gov/geo/>).

Department of Paediatrics, Hospital Clínico Universitario de Santiago, Santiago de Compostela, Galicia, Spain, and Grupo de Investigación en Genética, Vacunas, Infecciones y Pediatría (GENVIP), Healthcare research Institute of Santiago de Compostela and Universidade de Santiago de Compostela, Spain ³Department of Pediatrics, University of California San Diego, La Jolla, California, USA ⁴Rady Children's Hospital San Diego, San Diego, California, USA ⁵Emma Children's Hospital, Department of Paediatric Haematology, Immunology & Infectious Disease, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands ⁶Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses, and Instituto de Ciencias Forenses, Grupo de Medicina Xenómica, Spain ⁷Department of Paediatrics, University of Oxford and the NIHR Oxford Biomedical Research Centre, Oxford, UK ⁸NIHR Wellcome Trust Clinical Research Facility, University of Southampton UK ⁹University Hospital Southampton NHS Foundation Trust, Southampton, UK ¹⁰Sanquin Research and Landsteiner Laboratory, Department of Blood Cell Research, Amsterdam Medical Centre, University of Amsterdam, Amsterdam, The Netherlands ¹¹Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland, Australia

These authors contributed equally to this work.

Abstract

Importance—As clinical features do not reliably distinguish bacterial from viral infection, many children worldwide receive unnecessary antibiotic treatment whilst bacterial infection is missed in others.

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

Design—Febrile children presenting to participating hospitals in UK, Spain, Netherlands and USA between 2009-2013 were prospectively recruited, comprising a discovery group and validation group. Each group was classified after microbiological investigation into definite bacterial, 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) 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—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 the indeterminate group.

Results—The discovery cohort of 240 children (median age 19 months, 62% males) 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. 96 children had indeterminate infection. Bioinformatic 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, including 23 bacterial, 28 viral, 79 indeterminate; median age 17 months, 57% males), bacterial infection was identified in all 23 microbiologically-confirmed definite bacterial patients, with a sensitivity of 100% (95% confidence interval [CI], 100 - 100), and in 1 of 28 definite viral patients, with specificity of 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% (79.2-100) and 90.0% (70.0-100) respectively, and with specificity of 96.0% (88.0-100) and 95.8% (89.6-100). A minority of children in the indeterminate group were classified as having bacterial infection (63 of 136, 46.3%), although most received antibiotic treatment (129 of 136, 94.9%).

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.

Background

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 hospital and administer parenteral antibiotics while awaiting culture results.^{4–6} As 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 healthcare resources and in inappropriate antibiotic prescription.⁷

Molecular tests have the potential to identify bacterial and viral pathogens 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 investigates 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.

Methods

Patient groups – Discovery and Validation groups

Overall design of the study is shown in Figures 1 - 3. 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 three UK hospitals; patients were also recruited in Spain (GENDRES network, Santiago de Compostela), and USA (Rady Children's Hospital, San Diego). Inclusion criteria were fever (axillary temperature $\geq 38^{\circ}\text{C}$) and perceived illness of sufficient severity to warrant blood testing in children <17 years of age. Patients with co-morbidities 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 (Supplementary Methods and eTable 1 in the Supplemental Appendix) included children with meningococcal sepsis, 16 inflammatory diseases (Juvenile Idiopathic Arthritis and Henoch-Schönlein purpura) and published gene expression datasets which compared bacterial infection with viral infection, 12, 15, 17 or inflammatory disease.¹⁸ Healthy children were recruited from out-patient departments. Data from healthy controls were not utilized in identification or validation of gene expression signatures, and were only used for interpretation of direction of gene regulation.

Diagnostic process

All patients underwent routine investigations as part of clinical care including blood count and differential, C-reactive protein, blood chemistry, 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 PCR for common respiratory viruses. Chest radiographs were undertaken as clinically indicated. Patients were assigned to diagnostic groups using predefined criteria (Fig. 2). The Definite Bacterial group included only patients with culture confirmed infection, and the Definite Viral group included only patients with culture, molecular or immunofluorescent test-confirmed viral infection and no features of co-existing bacterial infection. Children in whom definitive diagnosis was not established (indeterminate infection) were categorized into Probable Bacterial, Unknown Bacterial or Viral, and Probable Viral groups based on level of clinical suspicion (Fig. 2). Detection of virus did not prevent inclusion in the Definite, Probable Bacterial, or Unknown groups, as bacterial infection can occur in children co-infected with viruses.

Study conduct and oversight

Clinical data and samples were identified only by study number. Assignment of patients to clinical groups was made by consensus of two clinicians independent of those managing the patient, after review of investigation results using previously agreed definitions (Fig. 2).

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 #140220; and Academic Medical Centre, University of Amsterdam (NL41846.018.12 and NL34230.018.10).

Peripheral blood gene expression by microarray

Whole blood was collected into PAXgene blood RNA tubes (PreAnalytiX, Germany), frozen, and later extracted. Gene expression was analyzed on Illumina microarrays. Additional details of microarray method, quality control, and analysis are provided in the Supplemental Appendix (Methods, Statistical Methods and eFig. 1).

Statistical Analysis

Transcript signature discovery

Expression data were analyzed using '*R*' *Language and Environment for Statistical Computing (R) 3.1.2*. 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 from Definite Viral infection were identified in the training set (Fig. 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 ($|\log_2$ fold change $| > 1$ and 2-sided $p < 0.05$) using *elastic net* (a variable selection algorithm that selects sparse diagnostic transcript signatures - see Supplemental Appendix Methods; eFig. 2).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 Supplemental Appendix). The Disease Risk Score (DRS) method²¹ was applied to the resulting minimal multi-transcript signature, in order 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 up-regulated transcripts (relative to comparator group) and subtracting the total intensity of the down-regulated 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 three pediatric and one adult published data sets (Fig. 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 to "gold-standard" clinically-assigned diagnoses (Fig 2). The area under the characteristic curve (AUC), the sensitivity and the specificity were reported. Confidence intervals at 95% were calculated to measure the reliability of our estimates (CI_{95%}).

Results

240 patients were recruited to the discovery group between 2009-2013 in the UK (189 patients), Spain (16), and USA (35). The Definite Bacterial group included 52 patients, of whom 36 (69%) required intensive care, and 10 died. In the Definite Viral group of 92 patients, 32 (35%) required intensive care, and none died (Table 1). The bacterial and viral patients were subdivided into 80% and 20% - forming a training set and test set respectively (Figs. 1, 3). The test set (20%) also included 96 children whose infection was not definitively diagnosed (indeterminate) (Figs. 1, 3). The validation groups comprised 130 UK and Spanish children previously recruited (IRIS validation – with 23 Definite Bacterial, 28 Definite Viral patients and 79 patients with indeterminate infection) and 72 additional validation children from the UK (25 children), Netherlands (30), and USA (17) – with 24 meningococcal infection, 30 juvenile idiopathic arthritis, and 18 patients with Henoch-Schönlein purpura) (Figs. 1 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 eTable 1. Details of the types of infection are shown in eTable 2. Gene expression profiles of children in the discovery group clustered together on Principal Component Analysis (eFig. 1).

Identification of minimal transcript signatures

Of the 8565 transcripts differentially expressed between bacterial and viral infections, 285 transcripts 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) as best discriminators of bacterial and viral infection in the discovery test set with sensitivity of 100% (95% CI, 100-100) and specificity of 95% (95% CI, 84 -100) (eTable 4). In the validation group, this signature had an AUC of 98% (95% CI, 94-100), sensitivity of 100% (95% CI, 100-100), and specificity of 86% (95% CI, 71-96) for distinguishing bacterial from viral infection (eTable 4, eFigs. 2, 3). The putative function of the 38 transcripts in our signature, as defined by Gene Ontology is shown in eTable 5.

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

Implementation of a Disease Risk Score (DRS)

The expression data of both genes in the signature was combined into a single Disease Risk Score for each patient, using the reported DRS method which simplifies application of multi transcript signatures as a diagnostic test.²¹ The sensitivity (95% CI) of the DRS in the training, test and validation sets respectively was: 86% (74-95), 90% (70-100), and 100% (100-100) (Fig. 4A-D, eFig. 4 and eTable 4). Expression of *IFI44L* was increased in viral patients and *FAM89A* was increased in bacterial patients relative to healthy children (eTable

3). The summary of diagnostic test accuracy including STARD flow diagrams are shown in eFig. 5.

For additional validation the 2-transcript signature was applied to patients with meningococcal disease (eFig. 6) and inflammatory diseases (Juvenile Idiopathic arthritis and Henoch-Schönlein purpura). Bacterial infection was identified with a sensitivity (95% CI) of 91.7% (79.2-100) and 90.0% (70.0-100) respectively, and with a specificity (95% CI) of 96.0% (88.0-100) and 95.8% (89.6-100). When applied to four published datasets for children and adults with bacterial or viral infection, and inflammatory disease (pediatric SLE),^{12,15,17,18} the 2-transcript signature distinguished bacterial infection from viral infection and inflammatory disease in all these datasets with AUC ranging from 89.2% to 96.6% (eTable 6 and eFigure 7-8).

Effect of viral and bacterial co-infection

The effect of viral co-infection on the signatures was investigated (Table 1). 30/47 (64%) of the definite bacterial infection group who were tested 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), those with features consistent with either bacterial or viral infection (Unknown), and those with clinical features and results suggestive of viral infection (Probable Viral) as in Fig. 2. The Probable Bacterial and Unknown groups included patients with DRS values that indicated 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 group DRS values overlapped with those of Definite Bacterial and Definite Viral groups (Fig. 5A, 4B).

DRS assignment as 'viral' or 'bacterial' was compared to clinical variables in the indeterminate group (eTable 7). CRP 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 assigned as bacterial by DRS had higher CRP levels than those assigned as viral infection (median [IQR]: 101 [48-192] and 71 [27-120] mg/l; $p=0.015$ respectively). They also had increased incidence of shock ($p=0.006$), requirement for ventilator support ($p=0.048$) and intensive care admission ($p=0.066$). There was a non-significant increase in white cell and neutrophil counts in patients assigned by DRS as bacterial or viral respectively: (median [IQR] 14.1 [8.3-19.4] and 11.1 [7.3-16.0] for white cells; 8.7 [5.0-13.8] and 6.8 [3.5-11.4] for neutrophils), ($p=0.079$ and 0.114 respectively).

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 >90% in the Unknown group. The high rate of antibiotic use in the indeterminate groups 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 (eFig. 9), and irrespective of day of illness (eFig. 10).

Discussion

This study identified a host whole blood RNA transcriptomic signature that distinguished bacterial from viral infection with two gene transcripts. The signature also distinguished bacterial infection from childhood inflammatory diseases, SLE, JIA and HSP and discriminated bacterial from viral infection in published adult studies.^{12,15,17,18} The results extend previous studies that suggest 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 two 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 interferons²⁴ 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 gold standard. Some studies include patients with “clinically diagnosed bacterial infection” who have features of bacterial infection but 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 groups.

The proportion of children predicted by DRS signature to have bacterial infection follows the level of clinical suspicion (greater in Probable Bacterial and less in the Probable Viral groups), 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 group, assigned as bacterial by the signature (Probable and Unknown

groups) had clinical features normally associated with severe bacterial infection, including increased need for intensive care, and higher neutrophil counts, and CRP, 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 gold 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 healthcare, 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 SLE, JIA and HSP. 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 in both 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 multi-transcript signatures into clinical tests suitable for use in hospital laboratories or at the bedside. The DRS signature, distinguishing viral from bacterial infections with only two transcripts, has potential to be translated into a clinically applicable test using current technology such as PCR.³⁰ Furthermore, new methods for rapid detection of nucleic acids including nanoparticles, and electrical impedance have potential for low-cost rapid analysis of multi-transcript 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

The authors wish to thank the children and families who have participated in the study, Dr. Jayne Dennis PhD, St. George's University of London, 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. Collaborators did not receive direct compensation for their work.

Funding: This work was supported by the Imperial College Comprehensive Biomedical Research Centre [DMPED P26077]; NIHR Senior Investigator award to ML; Great Ormond St Hospital Charity [V1401] to VJW; JH is supported by European Union's seventh Framework program [EC-GA 279185] (EUCLIDS); MK is supported by the Imperial College-Wellcome Trust Antimicrobial Research Collaborative (ARC) Early Career Fellowship [RSRO 54990]; 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) to FMT; Southampton NIHR Wellcome Trust Clinical Research Facility and NIHR Wessex Local Clinical Research Network; Academic Medical Centre Amsterdam MD/PhD program 2013 to AMB; the UK meningococcal disease cohort was established with grant support from the Meningitis Research Foundation (UK); the inflammatory disease cohort was supported by the Macklin Foundation grant to JCB; the National Institutes of Health U54-HL108460 to JCB; The Hartwell Foundation and The Harold Amos Medical Faculty Development Program/Robert Wood Johnson Foundation to AHT.

Role of the funders: the funders had no role in any of the following “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”.

Additional acknowledgements: people linked to the IRIS consortium through membership of the GENDRES consortium (www.gendres.org) are listed in the Supplemental Appendix.

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. [PubMed: 26034164]
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. *The Lancet Infectious diseases*. 2014; 14(5):397–405. [PubMed: 24631222]

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):e1455–1462. [PubMed: 23129086]
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 (Clinical research ed.)*. 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. [PubMed: 23550046]
6. Sadarangani M, Willis L, Kadambari S, et al. Childhood meningitis in the conjugate vaccine era: a prospective cohort study. *Archives of disease in childhood*. 2015; 100(3):292–294. [PubMed: 25256088]
7. Le Doare K, Nichols AL, Payne H, et al. Very low rates of culture-confirmed invasive bacterial infections in a prospective 3-year population-based surveillance in Southwest London. *Archives of disease in childhood*. 2014; 99(6):526–531. [PubMed: 24554055]
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 technology assessment (Winchester, England)*. 2015; 19(35):1–142.
9. Cebey Lopez M, Herberg J, Pardo-Seco J, et al. Viral Co-Infections in Pediatric Patients Hospitalized with Lower Tract Acute Respiratory Infections. *PloS one*. 2015; 10(9):e0136526. [PubMed: 26332375]
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. [PubMed: 20501927]
11. Anderson ST, Kaforou M, Brent AJ, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med*. 2014; 370(18):1712–1723. [PubMed: 24785206]
12. Hu X, Yu J, Crosby SD, Storch GA. Gene expression profiles in febrile children with defined viral and bacterial infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110(31):12792–12797. [PubMed: 23858444]
13. Herberg JA, Kaforou M, Gormley S, et al. Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. *The Journal of infectious diseases*. 2013; 208(10):1664–1668. [PubMed: 23901082]
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 medicine*. 2013; 10(11):e1001549. [PubMed: 24265599]
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. [PubMed: 17105821]
16. Pathan N, Hemingway CA, Alizadeh AA, et al. Role of interleukin 6 in myocardial dysfunction of meningococcal septic shock. *Lancet, The*. 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. *The Journal of infectious diseases*. 2015; 212(2):213–222. [PubMed: 25637350]
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. [PubMed: 20725040]
19. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met*. 1995; 57(1):289–300.
20. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Statist Soc B*. 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 medicine*. 2013; 10(10):e1001538. [PubMed: 24167453]
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. [PubMed: 21414208]

23. Zaas AK, Burke T, Chen M, et al. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. *Science translational medicine*. 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. [PubMed: 21478870]
25. Wong HR, Cvijanovich N, Lin R, et al. Identification of pediatric septic shock subclasses based on genome-wide expression profiling. *BMC medicine*. 2009; 7:34. [PubMed: 19624809]
26. Gurion R, Lehman TJ, Moorthy LN. Systemic arthritis in children: a review of clinical presentation and treatment. *International journal of inflammation*. 2012; 2012 271569.
27. Rabizadeh S, Dubinsky M. Update in pediatric inflammatory bowel disease. *Rheumatic diseases clinics of North America*. 2013; 39(4):789–799. [PubMed: 24182855]
28. Thierry S, Fautrel B, Lemelle I, Guillemin F. Prevalence and incidence of juvenile idiopathic arthritis: a systematic review. *Joint, bone, spine : revue du rhumatisme*. 2014; 81(2):112–117.
29. Gardner-Medwin JM, Dolezalova P, Cummins C, Southwood TR. Incidence of Henoch-Schonlein purpura, Kawasaki disease, and rare vasculitides in children of different ethnic origins. *Lancet (London, England)*. 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. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. 2015; 34(11):2209–2212.

Key Points

Question

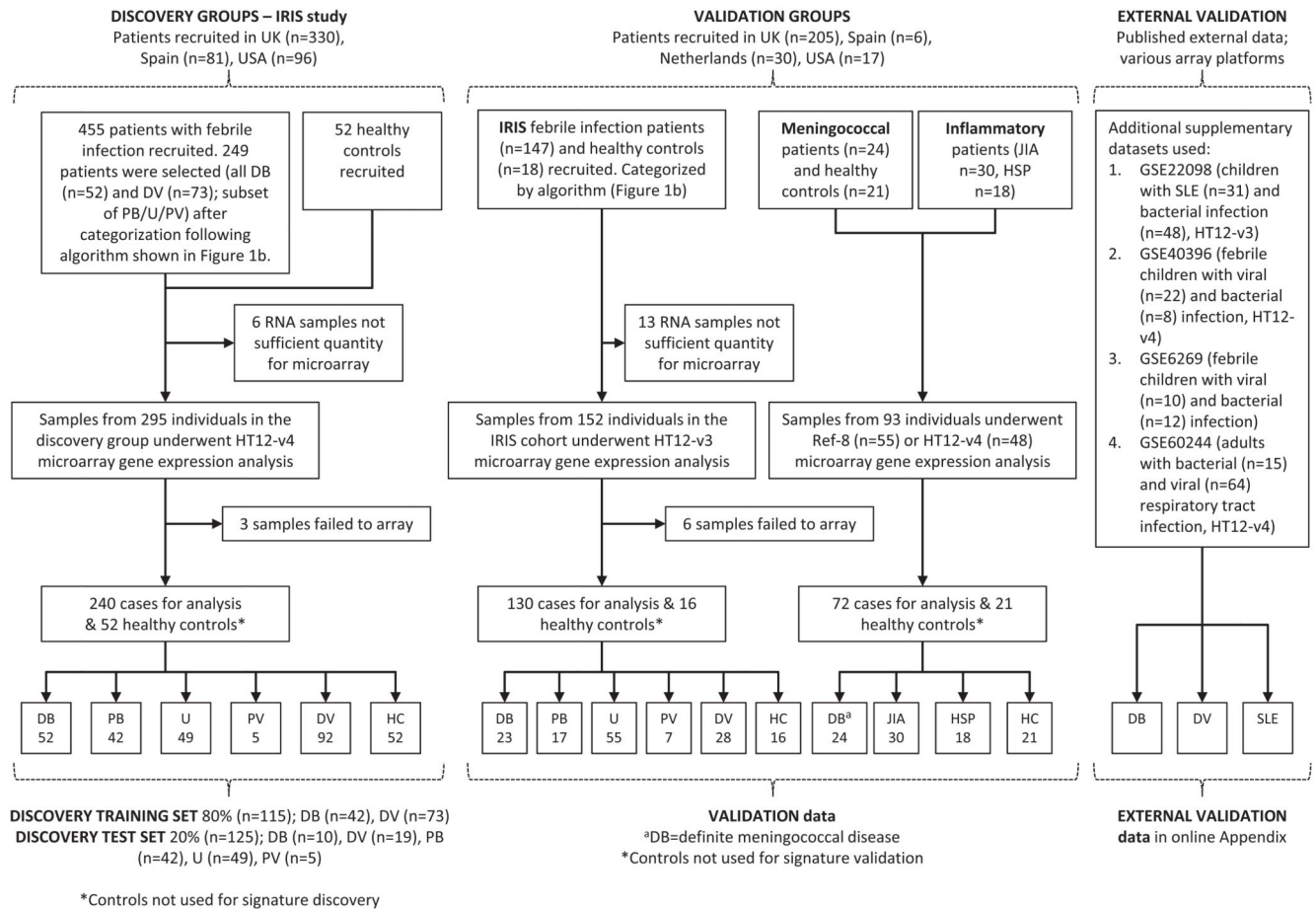
Can febrile children with bacterial infection be distinguished from those with viral infection and other common causes of fever using whole-blood gene expression profiling?

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% confidence interval [CI], 100 - 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.

**Figure 1. Study overview**

Overall flow of patients in the study showing patient recruitment and subsequent selection for microarray analysis.

HC Healthy Control; JIA juvenile idiopathic arthritis; HSP Henoch-Schönlein Purpura; SLE Systemic Lupus Erythematosus; DB Definite Bacterial; PB Probable Bacterial; U Unknown; PV Probable Viral; DV Definite Viral.

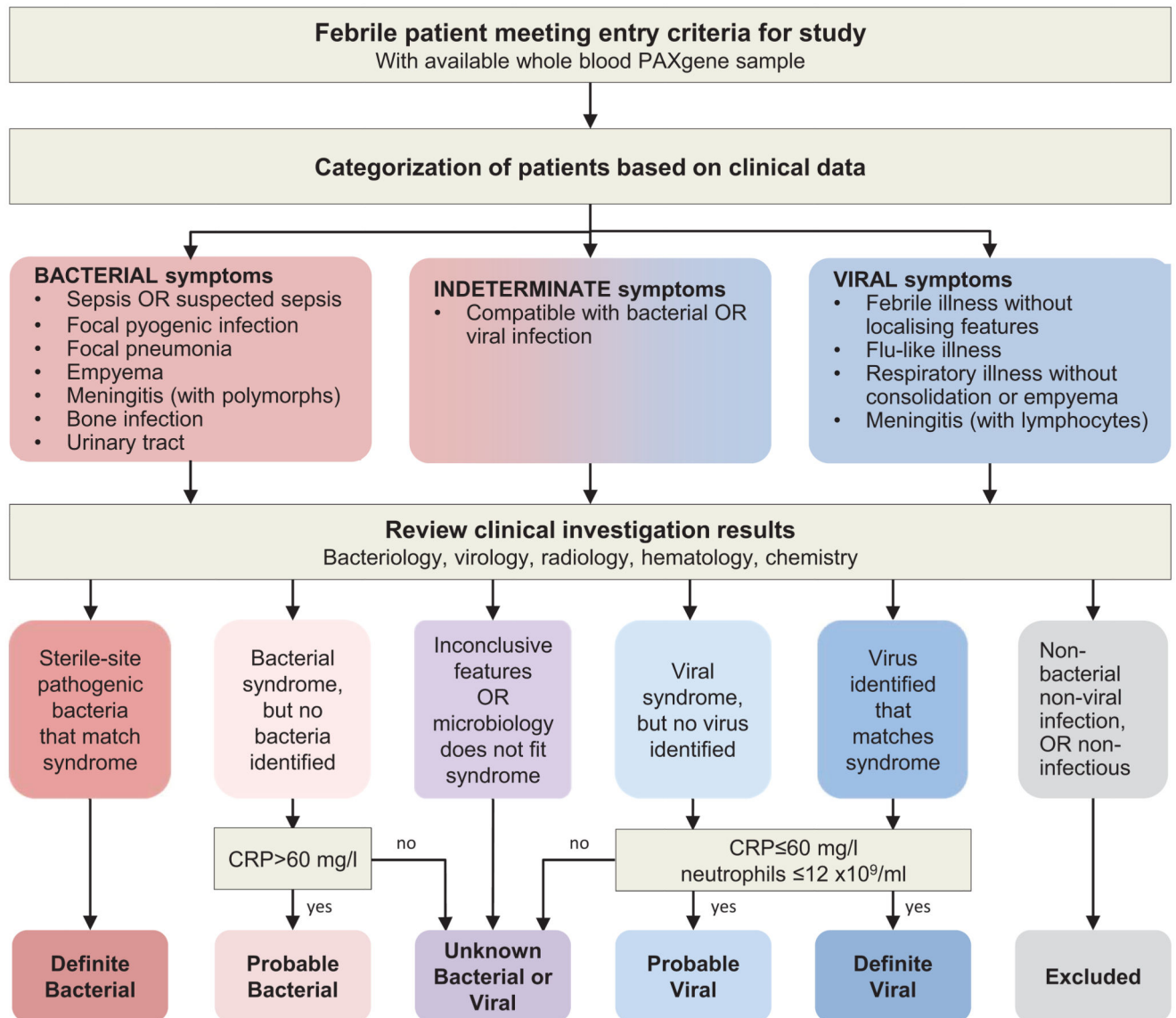


Figure 2. Classification of patients into diagnostic groups

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.

CRP: C-reactive protein

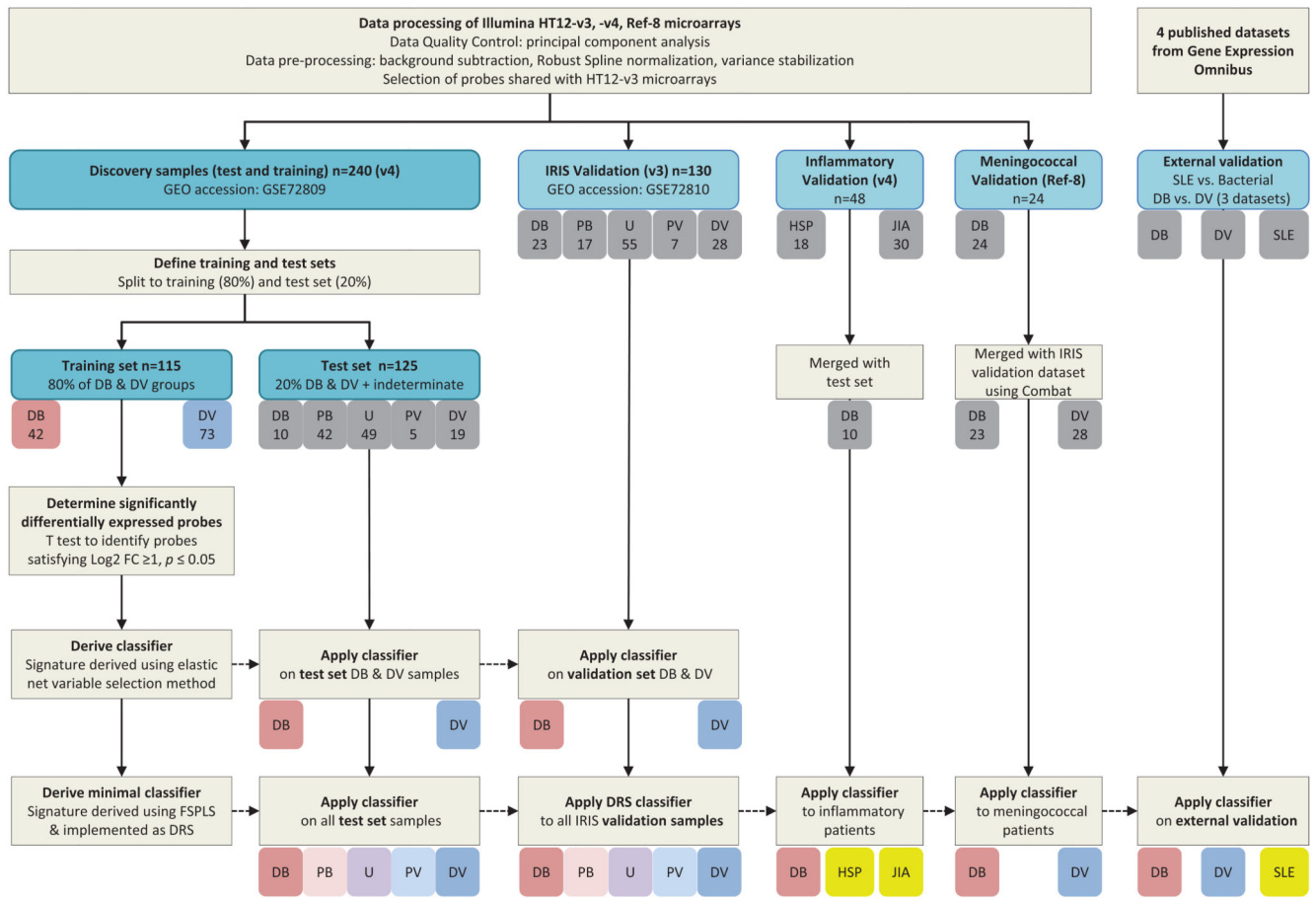


Figure 3. Analysis workflow

Overall study pipeline showing sample handling, derivation of test and training datasets, data processing, and analysis pipeline including application of 38-transcript elastic net classifier and 2-transcript DRS classifier, to the test set, the validation datasets and published (external) validation datasets.

DB Definite Bacterial; PB Probable Bacterial; U Unknown; PV Probable Viral; DV Definite Viral; HSP Henoch-Schönlein Purpura; JIA Juvenile Idiopathic Arthritis; SLE Systemic Lupus Erythematosus; HC Healthy Control; SDE Significantly Differentially Expressed; FC fold change; FS-PLS Forward Selection - Partial Least Squares; DRS Disease Risk Score

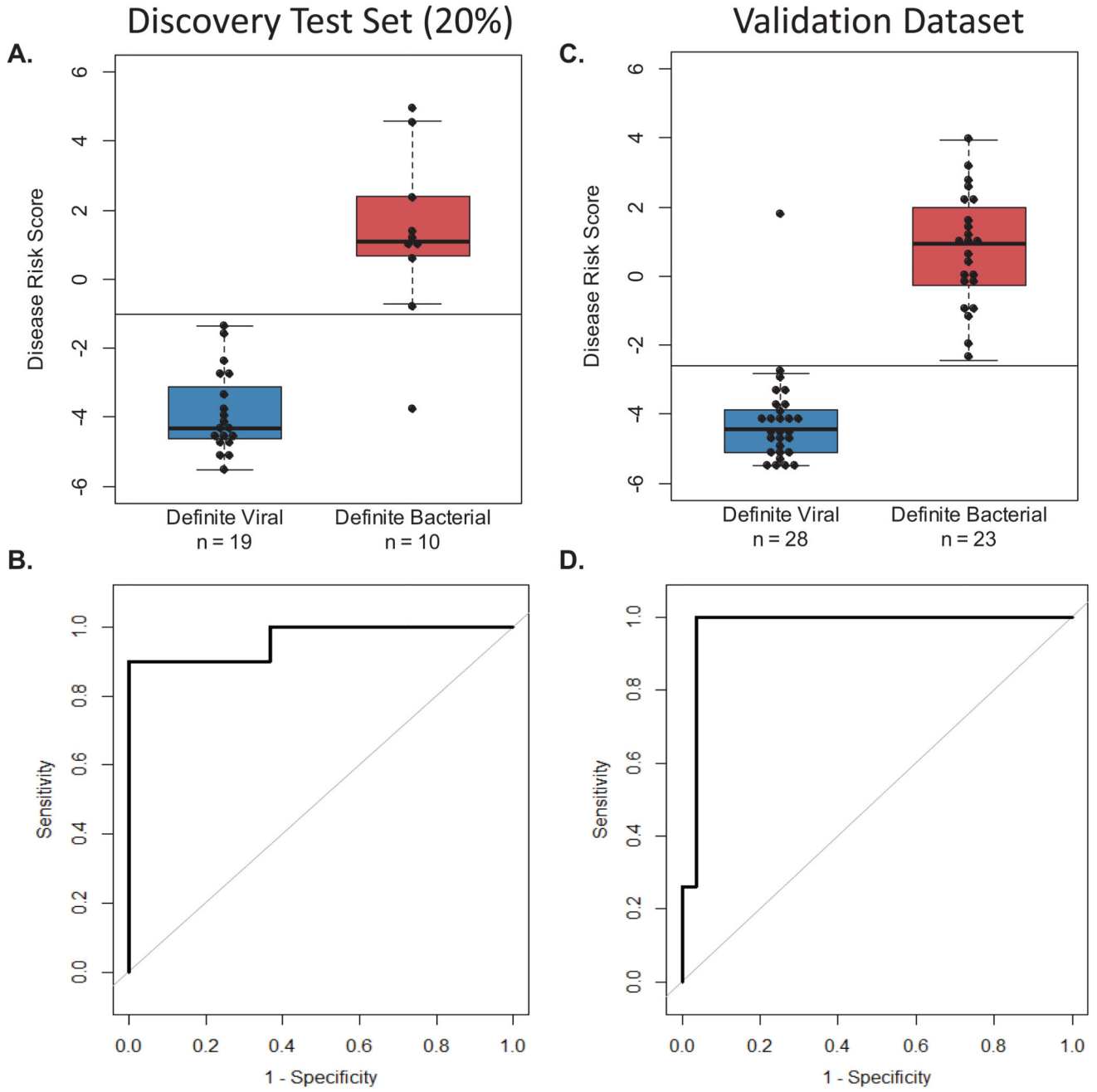


Figure 4. DRS and ROC curves based on the 2-transcript signature applied to Definite Bacterial and Viral groups

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 groups of the discovery test set (20% of the total discovery group) (A & B) and the IRIS validation dataset (C & D). Boxes show median with 25th and 75th quartiles; whiskers, plotted using ‘*boxplot*’ in R, extend 1 times the interquartile range. Sensitivity, specificity, and AUC are reported in eTable 4. The horizontal DRS threshold line 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.

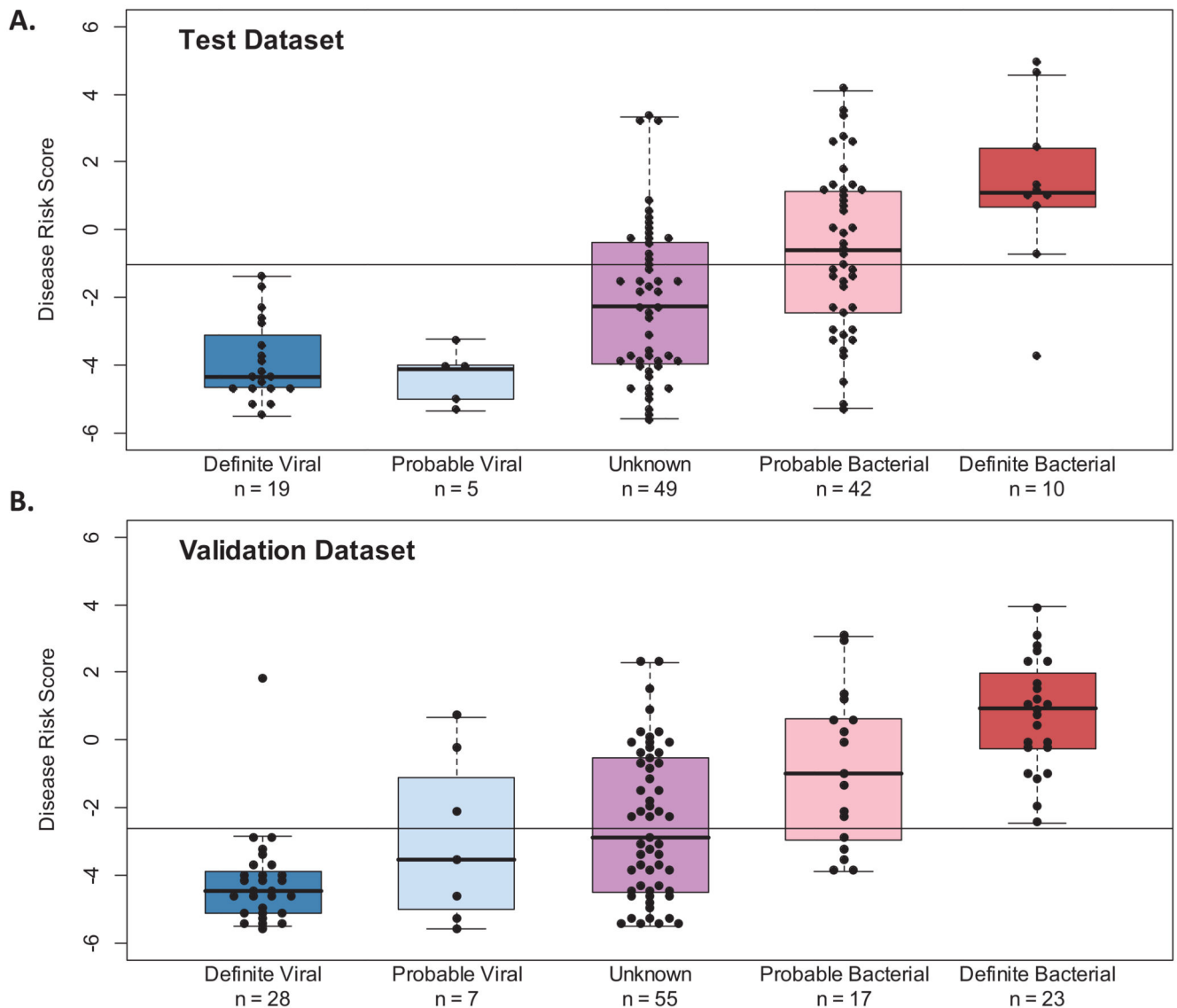


Figure 5. Performance of the 2-transcript DRS signature in indeterminate groups

Classification performance of the 2-transcript DRS (the combined *IFI44L* and *FAM89A* expression values) in the indeterminate groups of Probable Bacterial, Probable Viral, and Unknown of the discovery test (A) and IRIS validation (B) sets. Boxes show median with 25th and 75th quartiles; whiskers, plotted using ‘*boxplot*’ in R, extend 1 times the interquartile range. The horizontal DRS threshold line ($\text{threshold}_{\text{test_dataset}}=-1.03$; $\text{threshold}_{\text{validation_dataset}}=-2.63$) separates patients predicted as bacterial (above the line) or viral (below the line). It is determined by the point in the ROC curve that maximized sensitivity and specificity. For the test set, the training threshold was used.