nature communications

Article

https://doi.org/10.1038/s41467-025-65361-3

A four-gene signature from blood to exclude bacterial etiology of lower respiratory tract infection in adults

Received: 20 February 2025

Accepted: 13 October 2025

Published online: 24 November 2025

Check for updates

Ann R. Falsey[1] ✉, Derick R. Peterson[2], Edward E. Walsh[1], Soumyaroop Bhattacharya[3], Andrea M. Baran[2], Chinyi Chu[3], Angela R. Branche[1], Daniel P. Croft[1], Michael Peasley[1], Anthony M. Corbett[4], John Ashton[5] & Thomas J. Mariani[3]

Unnecessary antibiotic use is a major driver of antimicrobial resistance, an urgent public health threat. Acute respiratory infection (ARI) is a leading cause of inappropriate antibiotic use, creating an unmet need for improved diagnostics to identify bacterial etiology in ARI. In this work we show a 4-gene signature defining the absence of bacterial ARI which may be useful for managing antibiotics in ARI. Hospitalized adults with ARI underwent comprehensive microbiologic testing and those with definitive viral (n = 280), bacterial (n = 129), or mixed viral-bacterial infection (n = 95) had whole blood RNA sequencing. A hard-thresholded, mostly relaxed, LASSO-constrained logistic regression model is used to select a parsimonious gene set (*ITGB4, ITGA7, IFI27, FAM20A*) highly capable of discriminating any bacterial from nonbacterial infection (cross-validated AUC = 0.90). The 4-gene signature is validated in five independent adult RNAseq cohorts (AUC = 0.89–0.98), two adult microarray cohorts (AUC = 0.73–0.90), and one pediatric pneumonia RNAseq cohort (AUC 0.74). Thresholding the 4-gene risk score to yield 90% sensitivity to detect bacterial infection results in 71% specificity and 91% negative predictive value.

Acute respiratory infections (ARI) account for substantial morbidity and mortality in adults and are also a leading cause of antibiotic overuse. Although the actual microbial etiology of ARI is frequently viral, the specific microbiologic diagnosis is frequently unknown to the treating clinician[1–6]. Notably, unnecessary antibiotic use is a major driver of increasing antimicrobial resistance, one of the most urgent threats to global public health, and as such, more accurate microbiologic diagnostics for ARI are critically needed[6–8]. Although polymerase chain reaction (PCR) testing allows rapid diagnosis of respiratory viruses, the impact on antibiotic prescription has been modest primarily due to concern about bacterial co-infection[9–11]. Transcriptomics represents a powerful approach for analysis of the host response during infection[12]. Earlier studies indicate that viral and bacterial infections trigger specific host transcriptional patterns in blood, yielding unique "bio-signatures" that may discriminate viral from bacterial causes of infection[13–20]. Importantly, mixed viral-bacterial infection must be categorized with bacterial infection since antibiotic therapy is warranted, and additionally, predictive genes should ideally be limited in number in order to be adaptable to the development of rapid commercial tests. Substantial progress has been

[1]Department of Medicine, University of Rochester, Rochester, NY, USA. [2]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA. [3]Department of Pediatrics and Center for Children's Health Research, University of Rochester, Rochester, NY, USA. [4]Clinical and Translational Science Institute, University of Rochester, Rochester, NY, USA. [5]Genomics Research Center, University of Rochester, Rochester, NY, USA. ✉e-mail: ann_falsey@urmc.rochester.edu

made in defining the host response to pathogens of global significance as well as in certain clinical syndromes, including sepsis and chronic infection[21–25]. However, translation of this knowledge into improved diagnostic tools to support clinician decision-making in the management of respiratory infections remains limited[23,26,27].

Therefore, in this work, we show that a 4-gene set defined by RNA sequencing accurately discriminates bacterial and nonbacterial illness in adults hospitalized with ARI. This 4-gene signature may offer clinicians treating ARI a tool to supplement clinical judgment and help curb unnecessary antibiotic use.

Results

Cohort description

Between March 2019 and April 2023, 4346 potential participants were screened for eligibility. The most common reasons for exclusion were immunosuppression (15%) and low likelihood to make a microbiologic diagnosis (16%) (Fig. 1). In addition, 14% refused participation and 13% could not provide consent, leaving 1111 enrolled of which 1103 were evaluable. Of the 1021 cases of ARI, 504 were adjudicated to have a definitive microbiologic diagnosis and underwent RNA sequencing and were included in the primary analysis. In addition, 82 cases were enrolled as non-infected (NI) control subjects of whom 64 were adjudicated as having sufficient microbiologic assessment to be classified as non-infected. The clinical characteristics of those included in the primary analysis of ARI compared to those without a definite microbiologic diagnosis who were not analyzed are shown in Supplementary Table 1. The primary analysis group was slightly younger and had significantly fewer chronic medical conditions, including a history of smoking, COPD, home oxygen use and heart disease, than the unanalyzed group. The analysis group required intensive care use more often but had lower rates of radiographic pneumonia. Finally, discharge diagnoses also differed, with higher rates of bronchitis and viral syndrome in the analyzed compared to the unanalyzed subjects, the latter who had higher rates of acute exacerbations of chronic obstructive pulmonary disease (AECOPD).

The analysis group was composed of 280 viral alone (V), 129 bacterial alone (B) and 95 mixed viral bacterial (VB) illnesses (Details of each case are included in the supplementary information). The primary analysis compared cases with any bacterial ARI illness (B and VB, $N = 224$) to cases with nonbacterial ARI illnesses comprised of viral alone (V) illnesses ($N = 280$). The clinical features of each group are shown in Table 1. The any bacterial group was older, with a higher percentage of smokers, more often had sputum production, exhibited confusion and more abnormalities of vital signs. Laboratory values such as total white blood cell count and serum procalcitonin were significantly higher, and infiltrates on chest radiographs were more common in any bacterial group. In addition, any bacterial group had longer hospital stays and higher rates of intensive care and ventilatory support. Nonbacterial subjects were younger and more often had underlying asthma and upper respiratory infection symptoms. The discharge diagnoses were also significantly different between the two groups and aligned with the microbiologic category, with pneumonia and sepsis more frequent in the any bacterial group and asthma exacerbation and viral syndrome in the nonbacterial group. Of note, when comparing the VB subgroup to the B and V subgroups, the VB group most closely resembled the B group in terms of clinical presentation, laboratory parameters and outcomes, although there was a higher frequency of underlying asthma and URI symptoms at presentation (Supplementary Table 2).

The 63 NI controls included a variety of illnesses (28 congestive heart failure [CHF], 18 pulmonary emboli [PE], 7 atrial fibrillation with rapid ventricular response, 4 asthma exacerbations, 3 AECOPD, and 3 other) deemed non-infectious by the provider team and confirmed as such by the clinical adjudication committee. The reasons for exclusion of 18 NI are shown in Fig. 1. Additional details of the NI cohort compared to the primary analysis group are shown in Supplementary Table 3.

The most frequent viral detections in the primary analysis group are shown in Supplementary Table 4 and were rhinovirus (23%), influenza A (23%) and respiratory syncytial virus (10%) with few viral coinfections (4.1%). The most common bacterial pathogens detected were *Streptococcus pneumoniae* (12%), *Hemophilus influenzae* (7%), and Legionella (5%), and 2% had multiple bacterial detections. Of the viral-bacterial coinfections, the most common were influenza A and RSV, with *Streptococcus pneumoniae* (9% each) and rhinovirus with *Hemophilus influenzae* (11%).

Differential gene expression analysis of blood samples

Gene expression was compared between subjects with nonbacterial etiology ($n = 280$) and subjects with any bacterial infection ($n = 224$). On average 40 ± 11 million reads were generated from each of the cDNA libraries, with a mapping rate of 84.2 ± 1.1% and transcriptome

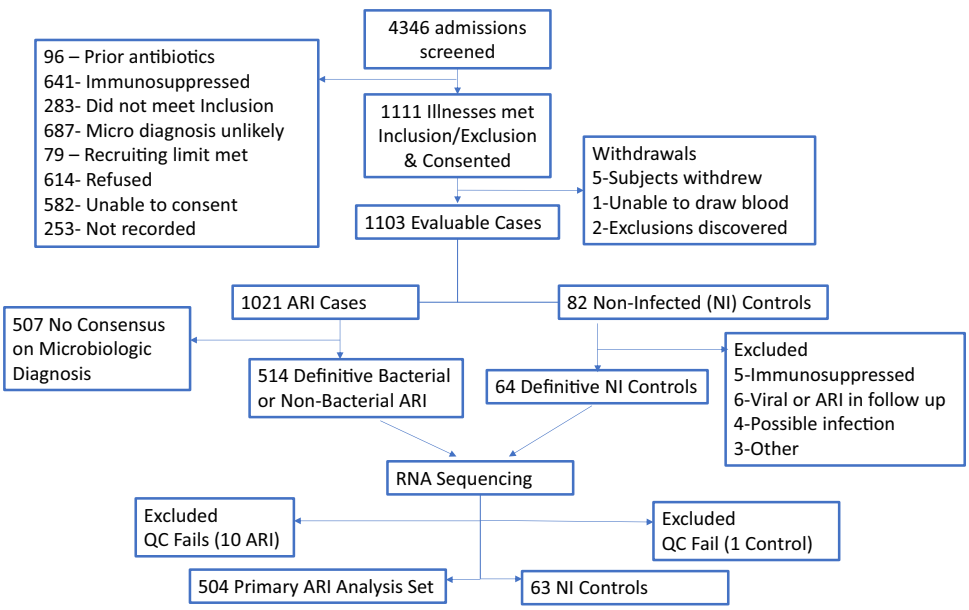


Fig. 1 | Consort diagram. Description of study population and analysis groups. Data for the figure is provided in the Source Data files.

https://doi.org/10.1038/s41467-025-65361-3

Table 1 | Study Populations and Illness Characteristics of Primary Analysis Cases

Characteristic	Bacterial[a] N = 224	Nonbacterial[b] N = 280	P-value
Mean Age (SD)	63.1 ± 15.9	59.8 ± 18.4	0.03
Female, No (%)	105 (47)	176 (63)	0.0004
Race, No (%)	–	–	0.051
White	166 (74)	181 (65)	
Black/AA	55 (25)	90 (32)	
Other	3 (1.3)	9 (3)	
Hispanic	17 (8)	36 (13)	0.059
Medical Conditions	–	–	–
Mean BMI (SD)	31.5 ± 10.1	32.3 ± 9.8	0.369
Asthma	61 (27)	128 (46)	< 0.0001
COPD	82 (37)	83 (30)	0.105
Any Smoking	177 (79)	196 (70)	0.025
Home Oxygen	26 (10)	30 (11)	0.777
CAD	34 (15)	53 (19)	0.288
CHF	32 (14)	33 (12)	0.425
Diabetes Mellitus	68 (30)	84 (30)	1.0
Chronic Kidney Disease	27 (12)	29 (10)	0.571
Any medical condition	218 (97)	276 (99)	0.351
Mean No. Medical Conditions (SD)	3.8 ± 1.8	3.8 ± 1.8	1.0
Symptoms & Signs			
Nasal Congestion	109 (49)	195 (70)	< 0.0001
Sore throat	66 (29)	138 (49)	< 0.0001
Cough	209 (93)	271 (97)	0.091
Dyspnea	208 (93)	253 (90)	0.340
Sputum production	171 (76)	174 (62)	0.0007
Feverish	153 (68)	170 (61)	0.093
Confusion	22 (10)	12 (4)	0.019
Highest pulse	108 ± 23.5	105 ± 18.3	0.083
Lowest Systolic BP	116 ± 20.5	125 ± 26.3	< 0.0001
Respiratory rate	26 ± 8.1	24 ± 7.1	0.003
Oxygen Saturation	90 ± 7.1	91.2 ± 6.8	0.054
Temperature	37.9 ± 1.0	37.5 ± 0.90	< 0.0001
Laboratory			
WBC	14.7 ± 7.7	8.8 ± 3.8	< 0.0001
BUN	12.0 ± 4.4	17.0 ± 11	< 0.0001
PCT	6.9 ± 18.6	0.13 ± 0.15	< 0.0001
CXR – any infiltrate	120 (54)	46 (16)	< 0.0001
Hospital Course			
ICU	53 (24)	22 (8)	< 0.0001
Non-invasive Ventilation	27 (12)	22 (8)	0.131
Mechanical Ventilation	12 (5)	1 (0.4)	0.0004
In hospital death	5 (2)	2 (0.7)	0.250
Primary Discharge Diagnoses			
Asthma	9 (4)	40 (14)	0.0001
Bronchitis	26 (12)	19 (7)	0.083
AECOPD	28 (13)	54 (19)	0.052
Pneumonia	84 (38)	16 (6)	0.0001
Respiratory Failure	12 (5)	15 (5)	1.0
Viral Syndrome	13 (6)	76 (27)	< 0.0001
Sepsis	14 (6)	0	< 0.0001
Other	38 (17)	60 (21)	0.215

[a]Bacterial includes 129 bacterial alone and 95 mixed viral bacterial cases.
[b]Non-bacterial are viral alone cases.
Two sided *t* test and Chi-Square or Fishers exact as appropriate, unadjusted for multiple comparisons.

coverage of 66.7 ± 10.8%. Differential expression analysis comparing any bacterial to nonbacterial groups identified 5401 genes as significantly differentially expressed at FDR < 0.05. We leveraged the existence of multiple sequencing batches to explore consistency in differentially expressed genes by leaving out each of the 7 batches, one at a time. Of the 5401 genes identified using all batches, 4584 (85%) were consistently identified as differentially expressed when leaving out any one batch (Fig. 2).

Diagnostic feature selection

We investigated the ability of gene expression patterns to discriminate any bacterial from nonbacterial infections. The cross-validation (CV) procedure selected relaxed LASSO parameters that resulted in a 4-gene diagnostic signature for any bacterial infection, Fig. 3 inset. The final signature included weighted expression of ITGA7, IFI27, FAM20A and ITGB4. Nested CV was used to estimate performance of the procedure via receiver operating characteristic (ROC) analysis (CV-area under the curve [AUC] = 0.90) (Fig. 3A). Density plots for the selected genes displayed separate, but overlapping expression patterns with *ITGA7, ITGB4* and *FAM20A* exhibiting higher expression in subjects with bacterial infection (Fig. 4) and the interferon-related gene *IFI27* exhibiting higher expression in subjects lacking a bacterial infection. Interestingly, *ITGA7* and *ITGB4* expression levels were not distinctly different between bacterial subjects with (VB) or without viral co-infection (B), but *IFI27* and *FAM20A* showed clearly distinct patterns in the 3 underlying diagnosis groups (Supplementary Fig. 1). The 4 gene-signature performed equally well when adding the 63 ill noninfected (NI) controls to the nonbacterial group (B + VB vs. V + NI) with an AUC of 0.90 (Fig. 3B).

Compared with traditional biomarkers of bacterial infection, such as white blood cell count (WBC) and procalcitonin (PCT) (Supplementary Fig. 2A, B), our 4-gene signature performed very well. However, caution should be exercised when interpreting the AUC of WBC and PCT since clinical adjudicators considered these data when adjudicating illnesses as bacterial or nonbacterial, resulting in incorporation bias.

Validation of the 4-gene diagnostic signature

We assessed the performance of our 4-gene signature using six external, independent RNAseq data sets that were appropriate for validation and included our study outcome (included bacterial and viral diagnoses). Early pandemic COVID-19 cases were excluded since they were not included in our primary analysis and model development. Validation was performed with and without controls as available. (Table 2) First, we examined our own previously published cohort (phs001248.v1.p1)[17], which consisted of 68 subjects (22 bacterial, 37 viral and 9 mixed bacterial-viral infections) recruited from the same Rochester, NY area population at an earlier epoch (January through June 2013). Our 4-gene signature discriminated between any bacterial and nonbacterial subjects with an AUC of 0.94 in this cohort. Next, we utilized data from a recent publication by Ko et al.[21] (GSE211567) of 224 subjects (101 bacterial, 123 viral, co-infection status not defined)

https://doi.org/10.1038/s41467-025-65361-3

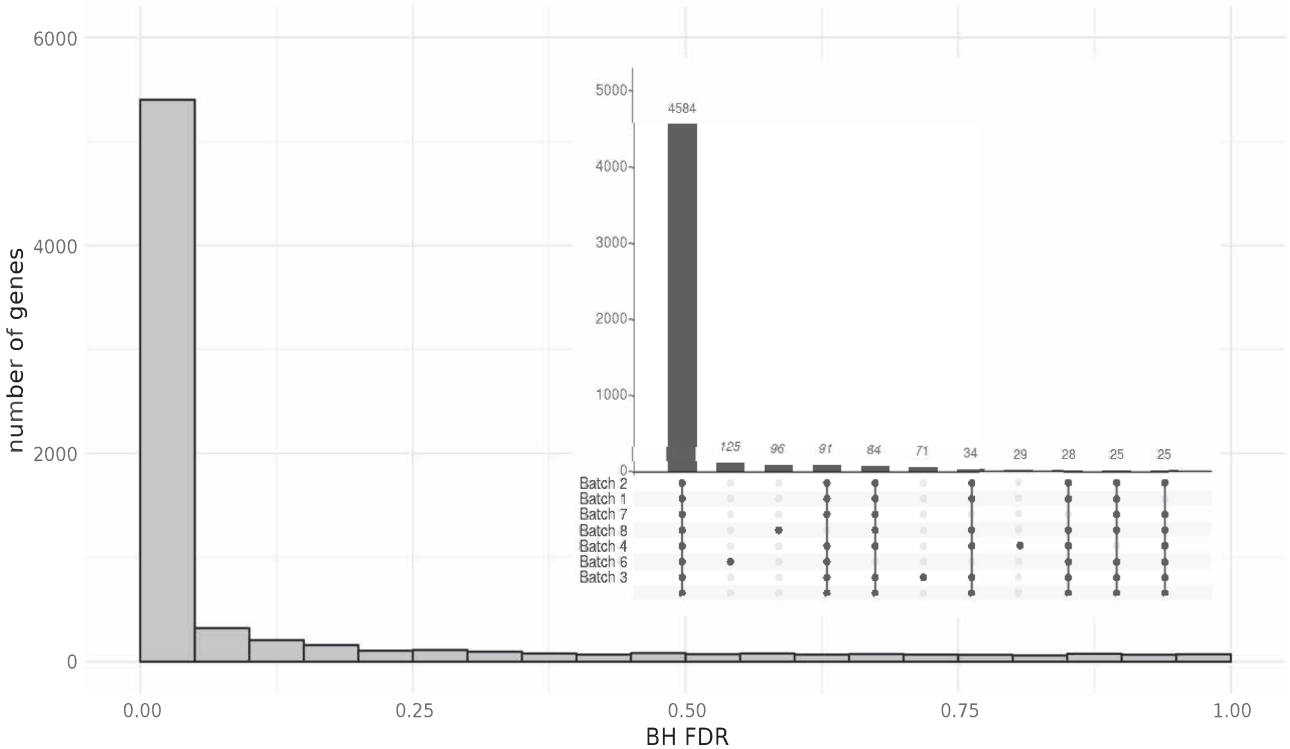


Fig. 2 | Differential gene expression. We identified 5401 differentially expressed genes (DEG) at a Benjamini-Hochberg False Discovery Rate (BH FDR) < 0.05 when comparing all subjects with a bacterial infection (B + BV, $n = 224$) to those subjects with only a viral infection (V; $n = 280$) using DESeq2. (Inset) We leveraged the existence of multiple sequencing runs (batches) in our cohort/data set to explore consistency in DEG identification, using a leave-one-batch-out (LOBO) approach (see "Methods"). Of the 5401 DEG found using all batches, 4584 genes (85%) were consistently found to be differentially expressed in all LOBO analyses. Data for the figure is provided in the Source Data files.

recruited from the US and Sri Lanka with a variety of infectious syndromes, and our 4-gene model provided an AUC of 0.90 and an AUC including noninfected ill controls of 0.86 in this second cohort. The gene signature performed well in 2 additional adult data sets (GSE 163151: 6 bacterial and 20 viral)[28] with an AUC of 0.89 and (GSE161731: 24 bacterial and 78 viral)[29] with an AUC of 0.98. If healthy controls were included, the AUCs dropped to 0.82 and 0.97 for each data set, respectively. In addition, we performed validation in a recently available RNAseq data set of adult ARI (GSE282464)[30] without COVID (early pandemic cases), which is more similar to our data, as well as with COVID; our 4-gene signature performed well with an AUC of 0.89 and 0.84, respectively. Finally, we evaluated our 4-gene set in a pediatric pneumonia study (GSE261482)[31] that included 98 definite bacterial and 12 definite viral infections, resulting in an AUC of 0.74[31]. If the healthy controls were included, the AUC rose to 0.92.

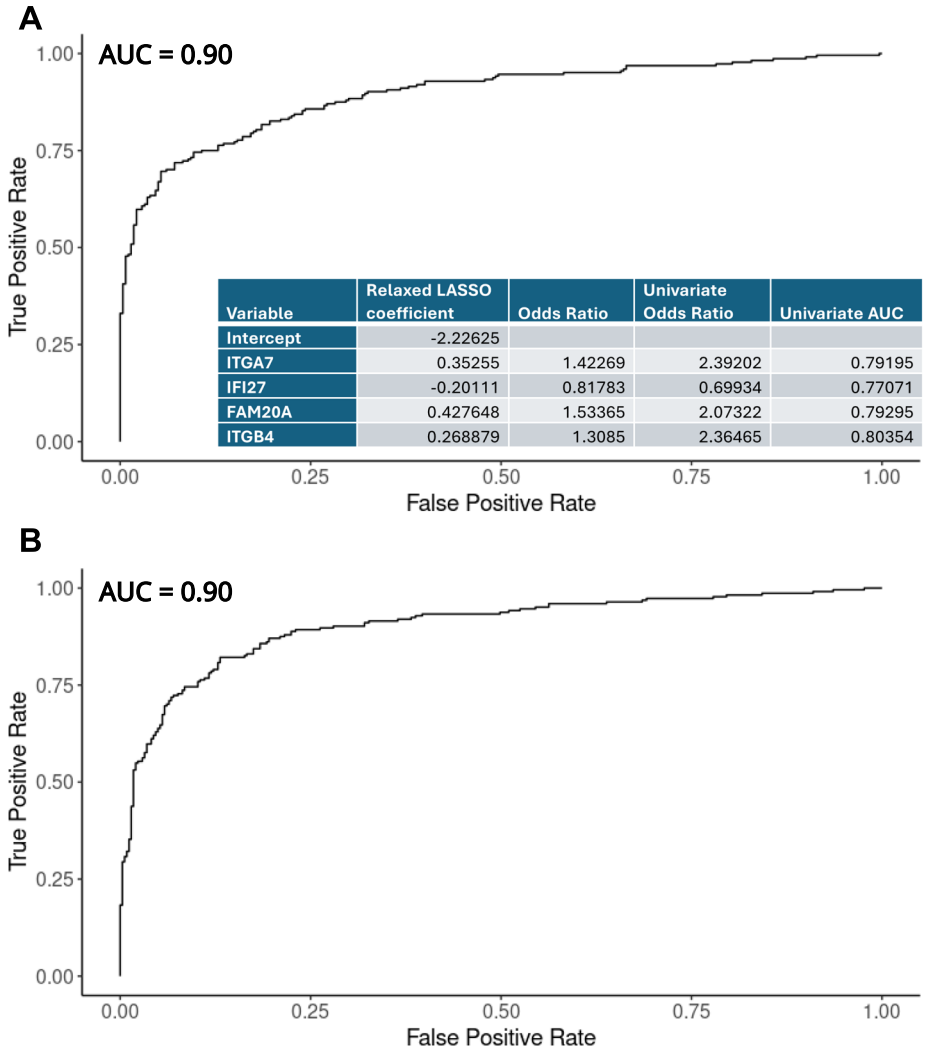
Despite the difference in sequencing platforms, we also performed validations using two microarray data sets. One included 154 adults with viral and bacterial respiratory infection and healthy controls (GSE60244)[16] resulting in an AUC of 0.73; another (GSE63990) included adults with ARI and controls that lacked data for one of our signature genes (FAM20A), so we imputed its expression, resulting in an AUC of 0.90[15].

Defining a threshold for classification

We sought to define a threshold for the gene signature that can be used for the classification of any bacterial infection. Based upon prior feedback from clinicians regarding the acceptability of a diagnostic test to exclude bacterial cause of LRTI[32], we aimed to achieve at least 90% sensitivity while also providing > 90% negative predictive value (NPV) for the determination of bacterial infection etiology in LRTI across a range of likely prevalences. Our 4-gene signature is a linear combination of the expression levels of 4 selected genes and represents a risk score or estimated log (odds) of bacterial infection for a patient with the given expression levels. The higher the risk score, the more likely the patient is to have a bacterial infection. Classifying patients as bacterial if the risk score exceeds − 0.886 (i.e., if the estimated probability of bacterial infection > 29%) and nonbacterial if the risk score < − 0.886 (i.e., if the estimated probability of bacterial infection < 29%) resulted in 90% sensitivity and 71% specificity to detect any bacterial infection in our training data. NPV ranged from > 91% at 40% prevalence of bacterial infection to > 95% at < 25% prevalence (Fig. 5A).

Choosing this threshold to achieve 90% sensitivity resulted in 23 cases adjudicated as bacterial being classified as nonbacterial. Examining the 23 missed any bacterial infections using the 4-gene threshold score, we found that 74% were VB and 74% were judged to be nonpneumonic ARI (Supplementary Table 5). Of the 6 cases adjudicated as bacterial pneumonia that were scored as nonbacterial by our threshold, none were bacteremic or had consolidation on chest radiograph. Additionally, using this threshold, only 1 of 10 mycoplasma pneumonia cases was misclassified as nonbacterial. The sensitivity of our 4-gene predictor was higher for more serious bacterial infections, such as pneumonia and bacteremia, at 93% and 100%, respectively. The majority of nonbacterial cases classified as bacterial using the 4-gene signature were either due to rhinovirus or influenza A and were nonpneumonic (86%) without distinguishing clinical features. The 4-gene signature also provided good separation of the noninfected controls from those with any bacterial infection (Fig. 5B). Forty-seven (75%) of

https://doi.org/10.1038/s41467-025-65361-3



Variable	Relaxed LASSO coefficient	Odds Ratio	Univariate Odds Ratio	Univariate AUC
Intercept	-2.22625			
ITGA7	0.35255	1.42269	2.39202	0.79195
IFI27	-0.20111	0.81783	0.69934	0.77071
FAM20A	0.427648	1.53365	2.07322	0.79295
ITGB4	0.268879	1.3085	2.36465	0.80354

Fig. 3 | Diagnostic gene signature. A A nested leave-one-batch-out cross-validation (CV) procedure tuning a hard-thresholded, mostly relaxed, LASSO-constrained logistic regression model was used to construct a parsimonious gene signature that distinguishes LRTI subjects with any bacterial infection from those with a viral-only etiology (CV-AUC = 0.90) (Inset). The table lists the names, adjusted contribution (relaxed LASSO coefficient/odds ratio), and univariate performance characteristics (univariate odds ratio and AUC) of the selected genes. **B** Performance of the 4-gene signature using LRTI subjects and including non-infected controls distinguishes LRTI subjects with any bacterial infection (B + BV) from those with a nonbacterial (V + NI) etiology (CV-AUC = 0.90). Data for the figure is provided in the Source Data files.

the non-infected controls were classified as nonbacterial using this threshold. Interestingly, a higher proportion of those misclassified as bacterial had a diagnosis of pulmonary embolism compared to those who were correctly classified as nonbacterial, 50% vs. 21%, $p = 0.051$, respectively.

Discussion

Bacterial antibiotic resistance has been identified as a major threat to global public health and is estimated to cause approximately 2 million illnesses and 23,000 deaths occur yearly in the United States (US)[8,33]. Notably, ARI is one of the most common reasons for emergency room visits and hospitalizations in the US and the majority are treated with broad-spectrum antibiotics, largely driven by concern for possible bacterial infection[11,34,35]. Due to the inability to readily sample the lower airways and the difficulty of clinically distinguishing any bacterial from nonbacterial causes, and identifying bacterial-viral coinfections, the management of ARI is severely hampered by diagnostic uncertainty and thus, ARI represents a clinical entity that may benefit most from a host response based diagnostic test. Serum biomarkers such as procalcitonin (PCT) and C-reactive protein (CRP) have shown limited success to supplement clinical judgment in assessing patients with ARI, although currently biomarker testing is not routinely recommended for management of respiratory infections[36,37].

In our study, a number of traditional clinical features, such as lack of URI symptoms, higher rates of sputum production, confusion and higher WBC and PCT levels were significantly different in participants adjudicated to have bacterial infection compared to viral alone. And yet, studies indicate that such clinical parameters do not provide sufficient precision to reliably distinguish viral from bacterial infection[38]. Our findings are consistent with the literature in that only half of the cases enrolled in the study could be definitively clinically adjudicated as having a clear microbiologic classification using traditional methods, leaving the other half with diagnostic uncertainty.

In our study, we sought to define a predictive gene expression signature capable of classifying nonbacterial ARI from ARI with any bacterial involvement, which could support a decision to withhold antibiotics. We identified a 4-gene set (*ITGB4, ITGA7, IFI27, FAM20A*) using nested cross-validation (CV) that was capable of discriminating any bacterial from nonbacterial infection with a CV-AUC of 0.90, specifically in adults hospitalized with cardiopulmonary illnesses. Its

https://doi.org/10.1038/s41467-025-65361-3

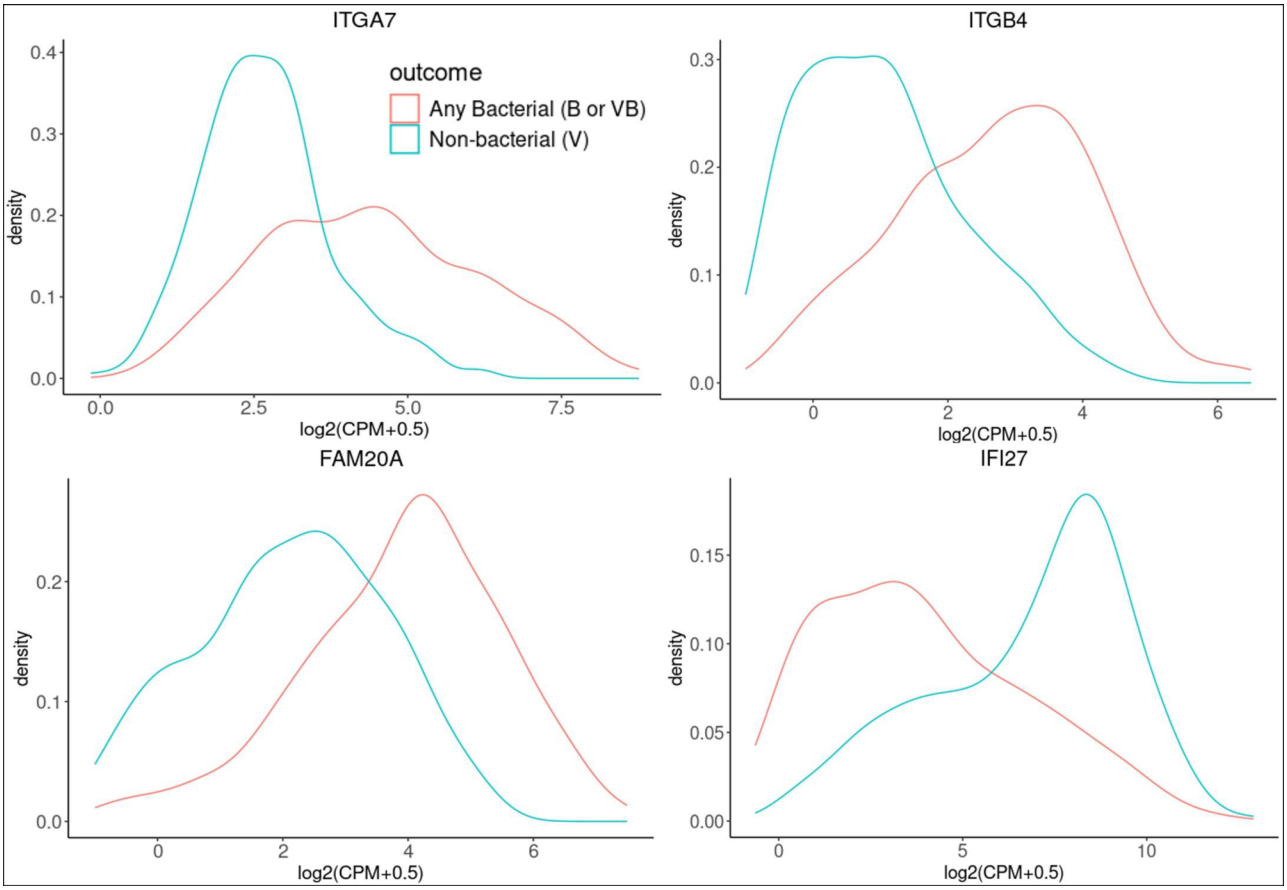


Fig. 4 | Expression profiles for the 4 genes in the signature. Distribution of counts per million (CPM) normalized gene expression for each of the 4 genes comprising the diagnostic gene signature by outcome. Any bacterial (B or VB) shown in red, and nonbacterial (V) shown in blue. Data for the figure is provided in the Source Data files.

parsimony makes the 4-gene signature set an optimal candidate for translation to a clinically useful test to appropriately target antibiotic use for respiratory infections. Importantly, this 4-gene signature was validated in five independent adult cohorts using RNAseq in subjects with ARI, from which data had been processed differently and independently, with an AUC range of 0.89 to 0.98. Notably, one of the validation cohorts was a global study that included a variety of pathogens not included in the current study (dengue fever, leptospirosis and rickettsial diseases). The alignment was not as high in the validation using RNAseq data from a pediatric pneumonia study (AUC of 0.74), suggesting optimal gene sets for children and adults may differ[21,31]. Many of the published microarray data sets evaluating gene expression signatures of bacterial/viral discrimination are pediatric. However, we were able to perform two validations using adult microarray data sets, which yielded AUCs of 0.73–0.90, suggesting RNAseq results may not always translate as well to microarray results. RNAseq has multiple improved performance characteristics as compared to microarrays (e.g., dynamic range, sensitivity, etc.), and some studies have reported improved clinical endpoint prediction for RNAseq[39].

Of the four genes used in our classifier, only one (*IFI27*) is a canonical viral response, interferon related gene. *IFI27* encodes interferon alpha inducible protein 27, an important regulator of Interferon Stimulated Genes (ISGs) shown to counteract the innate immune response to avoid harmful overstimulation[40]. *ISI27* has been included in other gene signatures for both viral and bacterial classification[41,42]. Two of the other four classifier genes encode integrins (ITGB4 and ITGA7), membrane proteins that mediate a wide spectrum of cell to cell and cell to matrix interactions. *ITGB4* encodes instructions for the synthesis of the β4 subunit of airway epithelial cell integrins and has been found to be important in in vivo viral infections such as RSV, and in local inflammatory and immune responses[43]. *ITGA7* encodes integrin subunit alpha 7, and has been studied as part of asthma related disease[44,45]. Specifically, overexpression of ITGA7 has been associated with a decreased inflammatory response of the airways in mice but a phenotype of contractile airway smooth muscle cells in adults[45]. The final classifier gene *FAM20A* codes for a golgi localized type 2 transmembrane protein, which exhibits increased expression in the lung and liver, in sepsis and in blood neutrophils of acute respiratory distress syndrome and was found to be among the 5 genes predicting bacterial infection in a pediatric pneumonia study[31,46–48]. The mechanisms by which these 4 differentially expressed genes can accurately discriminate any bacterial from nonbacterial respiratory infections will require further study. However, their role, or lack thereof, in infection response pathophysiology does not preclude their diagnostic potential.

Thresholding the 4-gene risk score to yield 90% sensitivity resulted in 71% specificity to detect bacterial infection with a 91% negative predictive value at 40% prevalence of bacterial infection and > 95% at prevalence < 25%). In our study, the goal was to support decisions to withhold or withdraw antibiotics in hospitalized adults by developing an expression profile to exclude bacterial etiology, thereby prioritizing sensitivity over specificity to yield high NPV. In prior surveys we conducted amongst practitioners, only 8% would accept a bacterial misdiagnosis rate of ≥ 20%[32]. Since the default position for most physicians has been to prescribe antibiotics "just to be safe", especially for patients ill enough to warrant hospitalization, the specificity of 71% should still result in a substantial overall decrease in antibiotic use. In our prior study of ARI in hospitalized adults, we found that 90% of patients judged to have viral infection alone received antibiotics during admission[11]. Importantly, the cases of bacterial infection misclassified in the current study were primarily diagnoses of mixed viral-bacterial bronchitis,

https://doi.org/10.1038/s41467-025-65361-3

Table 2 | Validation Cohorts and Results

Validation	Dataset	Platform	Age	Study Population	AUC[a]	Sensitivity[b]	Specificity
1	PhsO01248.v1.p1[17]	RNA-Seq	Adult	ARI (viral, bacterial, coinfection)	0.94 ($p < 0.0001$)	90%	84%
2	GSE211567[21]	RNA-Seq	Adult	Febrile Illness (viral, bacterial)	0.90 ($p < 0.0001$)	90%	77%
				Febrile Illness (viral, bacterial) + ill noninfected controls	0.86 ($p < 0.0001$)	90%	65%
3	GSE161731[29]	RNA-Seq	Adult	ARI (viral, bacterial)	0.98 ($p < 0.0001$)	92%	100%
				ARI (viral, bacterial) + healthy controls	0.97 ($p < 0.0001$)	92%	99%
4	GSE163151[28]	RNA-Seq	Adult	Infection (bacterial sepsis, influenza)	0.89 ($p = 0.001$)	100%	65%
				Infection (bacterial sepsis, influenza) + healthy controls	0.82 ($p = 0.005$)	100%	45%
5	GSE261482[31]	RNA-Seq	Pediatric	Pneumonia (viral, bacterial)	0.74 ($p = 0.04$)	90%	33%
				Pneumonia (viral, bacterial) + healthy controls	0.92 ($p < 0.0001$)	90%	80%
6	GSE282464[30]	RNA seq	Adult	Infection (bacterial, sepsis, shock, viral, coinfection)	0.89 ($p < 0.0001$)	90%	72%
7	GSE60244[16]	Microarray	Adult	ARI (viral, bacterial, coinfection)	0.73 ($p < 0.0001$)	92%	23%
				ARI (viral, bacterial, coinfection) + healthy controls	0.67 ($p = 0.0003$)	92%	19%
8	GSE63990[15]	Microarray missing FAM20A	Adult	Bacterial ($N = 73$) vs Viral ($N = 117$)	0.90 ($p < 0.0001$)	90%	67%
				Bacterial ($N = 73$) vs Viral ($N = 117$) and non-infectious illness ($N = 90$)	0.79 ($p < 0.0001$)	90%	44%

[a]AUC-(1-sided Wilcoxon p-value), [b]Sensitivity target was at least 90%.

suggesting minimal harm from using the gene expression score to guide antibiotic use. It is also possible that despite rigorous adjudication that some cases were not categorized correctly. Illnesses with both viral and bacterial detections may reflect bacterial colonization rather than true invasive infection. Nevertheless, even accepting that no test is 100% accurate, this finding also raises the possibility that gene expression could differ based on parenchymal vs mucosal infection. Thus, it may be worthwhile in the future to examine gene expression stratified by clinical syndrome, such as pneumonia versus non-pneumonic ARI.

Numerous studies over the past decade have demonstrated whole blood gene expression analysis can discriminate viral from bacterial causes of infection with greater accuracy than PCT[12,16,26,49,50]. In a recent systematic comparison of gene signatures for bacterial/viral discrimination, 28 gene signatures were validated in 51 publicly available data sets compromised of 4589 patients[41]. In a number of studies, investigators used public data sets to develop predictors, followed by validation with relatively small numbers of prospectively enrolled new patients. Discovery phenotypes were varied (viral, healthy, bacterial, SIRS, COVID) as were populations (neonates, children, adults) and diseases studied. The number of genes included in predictor gene sets ranged from 1 to 398 and weighted mean AUC for bacterial vs. non-bacterial discrimination ranged from 0.547 to 0.863. Interestingly, there has been little overlap in the predictive genes identified amongst the various studies[41]. The difference in the predictive genes identified are likely explained by diverse populations, types of infection and control groups studied, along with alternate analytic tools, with most using micro-array technology[22,23]. Only a few studies to date have specifically focused on respiratory illness in adults[16,17,27–29]. In the largest study of respiratory infections to date, Tsalik et al. used a microarray to evaluate gene expression to discriminate bacterial from viral infection or non-infectious illness in 273 subjects with community onset ARI[15]. Investigators defined 130 predictor genes in a model with an accuracy of 87% to discriminate clinically adjudicated bacterial, viral, and non-infectious illness. Recently, this group custom designed a PCR test to rapidly measure 45 host messenger RNA transcripts that could be performed on the automated BioFire Film Array platform[51] and are developing a portable molecular platform, Biomeme HR-B/V, demonstrating gene expression tools have viable paths forward for commercialization[52].

Although the field is advancing quickly, we feel our study is a significant advance. Importantly, we included a robust sample of mixed viral-bacterial infections as part of any bacterial infection group, 95 / 224 (42%). This very important subgroup has not been well evaluated in prior studies. As antecedent viral infection is common in bacterial lung infection, this is a critical subgroup to include in developing predictors that will be clinically useful[11]. Secondly, RNAseq technology may uncover new useful genes that were previously not identified, as evidenced by our study and the pediatric study by Viz-Lasheras et al. [31] Thirdly, a high NPV is very likely needed for provider acceptance, which we demonstrated over a range of bacterial infection prevalence. Additionally, mycoplasma has presented challenges for other host response diagnostic tests[31,53], and our gene set correctly classified mycoplasma infection as bacterial in 90% of cases. Finally, although our 4 gene signature is not the smallest discriminatory gene set described, it is parsimonious, which lends itself to further development.

Finally, two rapid host response-based tests are now commercially available. A rapid protein based assay, which combines measurements of TNF-related apoptosis-inducing ligand (TRAIL), IP-10 and CRP available as MeMed BV, has demonstrated improved diagnostic performance than any protein measurement alone[54]. Studies of MeMed BV in children and adults presenting to emergency rooms or urgent care with fever or ARI suggests physicians will respond appropriately to the information provided[55]. TriVerity®, a 29 gene-based predictor of bacterial sepsis (AUC 0.83), is FDA approved based on a large clinical study of suspected sepsis[24,25]. These tests were compared in 169 adults with respiratory infection by Dedeoglu et al. and both demonstrated similar overall

https://doi.org/10.1038/s41467-025-65361-3

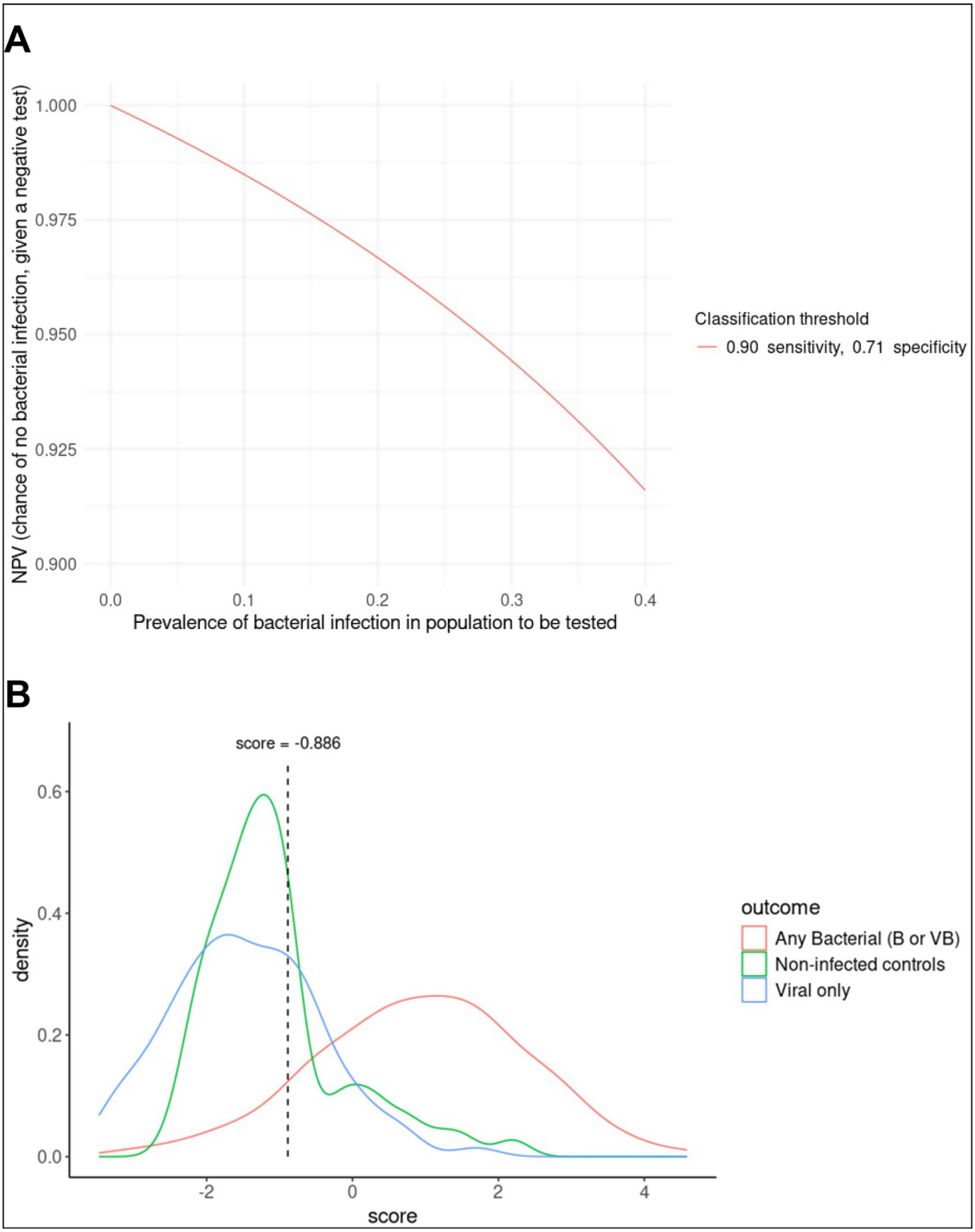


Fig. 5 | A gene signature-based diagnostic test. For practical implementation of a diagnostic test to exclude the involvement of bacterial etiology in LRTI, we aimed to find a risk score threshold that achieves both sensitivity and Negative Predictive Value (NPV) > 90%. **A** The risk score threshold of − 0.886 corresponds to 90% sensitivity and 71% specificity in the training data, and NPV ranges from 0.91 to 1.00 across a range of likely prevalences. **B** The distribution of the 4-gene signature by outcome group shows good separation between bacterial and nonbacterial subjects. The red line indicates any bacterial, the blue shows viral-only cases, and the green line indicates the non-infected controls. LRTI subjects with score ≥ − 0.886 are classified as bacterial and subjects with scores < − 0.886 are classified as non-bacterial. This threshold provides 75% specificity in the set of non-infected controls. Data for the figure is provided in the Source Data files.

accuracy[27]. However, more bacterial infections were missed by Triverity, whereas MeMed BV misclassified most viral infections as bacterial. Both tests provide a range of probability scores for the likelihood of bacterial or viral infection, with some subjects falling into an uncertainly range. The true utility of any host response test, including ours, will require prospective interventional trials adequately designed and powered to assess the impact on antibiotic use and clinical recovery.

Our study had a number of strengths but also some potential weaknesses. Strengths included the large sample size, inclusion of mixed viral-bacterial infections, comprehensive microbiologic evaluation and careful clinical adjudication by infectious diseases and pulmonary specialists. Potential weaknesses include dependence on clinical adjudication since a gold standard for bacterial respiratory infection does not exist, and interruption of the study by the COVID-19 pandemic,

which changed clinical testing practices and de-emphasized bacterial sputum cultures. There were also differences in the characteristics of those analyzed compared to those not included. Due to the need for correct microbiologic designation to develop the predictors, cases that were analyzed were skewed to extremes of the clinical spectrum (ex. asthmatics with URI symptoms and wheezy bronchitis and persons with consolidative pneumonias). Thus, there may be value to evaluate the gene score in the remainder of the participants with less clear-cut diagnoses. Finally, our data show associations between bacterial infection and clinical measurements traditionally considered surrogates of severe illness, which is unsurprising. Since this information was available to adjudicators and likely contributed categorizing some cases as either viral or bacterial, we excluded surrogates of disease severity from our model to avoid optimistically biasing performance. However, we acknowledge that our gene signature might relate not only to bacterial infection but also to disease severity and thus might be a potential confounder of the associations between genes and bacterial infection.

In conclusion, we developed a parsimonious gene set highly capable of discriminating any bacterial from nonbacterial respiratory infection. Our 4-gene signature yielded 90% sensitivity to detect bacterial infection with 91% NPV and was validated in six independent RNAseq cohorts. This 4-gene signature may offer clinicians treating ARI a tool to supplement clinical judgment regarding antibiotic management of these common infections. Further study will be required to test if physicians will accept such a tool and if antibiotic use can be safely reduced, and produce improved outcomes for patients hospitalized with ARI.

Methods

Details of the Methods are provided in the supplementary information.

Study period and sites

The study was conducted at two hospitals in Rochester, N.Y, University of Rochester Medical Center (URMC) and Rochester General Hospital (RGH) between March 2019 and April 2023. The study was approved by the University of Rochester and the Rochester Regional Health institutional review boards, and all participants or their legally authorized representatives signed written informed consent prior to study procedures. Enrollment was paused from March to October 2020 due to the COVID-19 pandemic.

Recruitment

Potential participants with symptoms of an acute cardiopulmonary illness or diagnoses compatible with acute respiratory infection (ARI, i.e., pneumonia, acute exacerbation of chronic obstructive pulmonary disease, bronchitis, asthma, upper respiratory infection, influenza, viral syndrome) were screened by reviewing hospital admission logs. A limited number of patients with noninfectious cardiopulmonary diagnoses were included as control cases. Participants were enrolled within 24 h of admission if hospitalized.

Acute illness evaluation

At enrollment demographic, clinical and laboratory information were collected from the medical record and direct patient and family interviews. Medical history and medications, date of illness onset and signs and symptoms were collected. Results of standard of care (SOC) testing were recorded.

Clinical and microbiological adjudication

Cases were adjudicated by a panel of four physicians (three infectious diseases and a pulmonary medicine specialist) and classified into discrete microbiologic categories of viral infection alone, bacterial infection alone or bacterial-viral coinfection. Confidence in the microbiological classification was rated as definitive, probable, or indeterminate, and required unanimous agreement by adjudicators. Only cases judged as definitive were included in the primary analysis, and blood samples sent for RNA sequencing. Noninfected (NI) control cases required a clear noninfectious event with a negative nasal swab PCR to exclude asymptomatic viral infection. Controls were contacted 7–14 days later to ensure their non infected status had not changed.

Laboratory methods

Whole blood was collected in Tempus™ Blood RNA Tubes and RNA isolated using the Tempus Spin RNA Isolation Kit (Applied BioSystems). Total RNA was processed for globin reduction using the GLOBINclear Human Kit, and cDNA library construction was performed using the TruSeq Stranded mRNA library kit (Illumina, San Diego, CA) as described previously[56]. Libraries were sequenced on the Illumina NovaSeq6000 (Illumina, San Diego, CA). Reads were mapped to the Human GRCh38/genecode38 reference using STAR, counts were summarized with HTSeq[57] and counts per million (CPM) normalized[58] on a linux cluster of processor nodes located at the Center for Integrated Research Computing (CIRC) of the University of Rochester. Gene-specific mean CPM values were used to filter genes with insufficient expression (mean CPM < 2) for downstream analyses. Principle component analysis (PCA) was used to identify and remove 11 outlier samples (10 ARI and 1 NI). Sequencing run (batch)-specific variance modeling was used to remove additional genes. After subject and gene filtering, we retained 504 ARI samples and 7352 genes for downstream analyses. Downstream differential expression analysis was performed using the library DESeq2[58] as implemented in R (v4.5.0).

Statistical methods

A leave-one-batch-out nested cross-validation procedure was used to tune a hard-thresholded, mostly relaxed, LASSO-constrained logistic regression model to construct a gene signature to discriminate any bacterial (B and VB) from nonbacterial (V) infections. Independent validation of the gene signature was performed in several datasets with comparable populations and clinical outcomes. Six RNAseq data sets were selected for validation, which represented all RNAseq data sets that included both viral and bacterial infections. Of the 50 Microarray data sets available, only 20 included both viral and bacterial diagnoses (14 pediatric and 6 adult). Because it was felt that validation with a different platform was suboptimal, we chose to focus on the 6 adult studies. Four studies were rejected because of sample size ≤ 10, no publication associated with the data, full data set unavailable, or longitudinal sampling, leaving two studies of adult ARI for validation. For each independent validation dataset, we applied the coefficients from the final gene signature to construct a risk score for each validation subject. An ROC curve with associated AUC was used to assess the performance of the risk scores in each independent validation set.

To derive a classifier from the gene signature, we identified the threshold that provided ≥ 90% sensitivity, which also yielded 91% NPV across a range of likely prevalences (≤ 40%) of bacterial infection. Any subject with a gene signature ≥ threshold is classified as bacterial, and any subject with a gene signature < threshold is classified as nonbacterial.

Clinical variables were compared between subjects with any bacterial infection vs. nonbacterial infection using t tests for continuous variables or Fisher's exact tests for categorical variables. P-values are two-sided, with $p \le 0.05$ indicating nominal statistical significance. Differences in expression between any bacterial and nonbacterial ARI subjects for each gene were assessed using DESeq2 with FDR < 0.05. Statistical analysis was performed using analytical packages and codes as implemented in R v4.5.0 and Matlab 2022a. The codes to reproduce the analysis are available at https://github.com/ambaran3/LRTI_CV.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The individual phenotypic and bulk RNA sequencing data generated in this study have been deposited in the dbGaP database under accession code phs003931.v1.p1. The individual phenotypic and bulk RNA sequencing data are available under restricted access for subject privacy and confidentiality; access can be obtained by submitting a Data Access Request via dbGaP's Authorized Access Portal [https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login]. The processed QC'd/filtered count matrices for both the infected and the NI controls are on GitHub at https://github.com/ambaran3/LRTI_CV. Source data for tables and figures in the main text and supplementary information are provided as a Source Data file. All code used to perform analysis in this manuscript is available at https://github.com/ambaran3/LRTI_CV. A compilation of the 504 cases with limited clinical information and microbiologic testing is available on Figshare at https://doi.org/10.6084/m9.figshare.28454330. The validation data sets used in this study are available under accession codes (Phs001248.v1.p1, GSE211567 GEO Accession viewer, GSE161731 GEO Accession viewer, GSE163151 GEO Accession viewer, GSE261482 GEO Accession viewer, GSE282464 GEO Accession viewer, GSE60244, GEO Accession viewer GSE63990 GEO Accession viewer). Source data are provided in this paper.

References

1. Havers, F. P. et al. Outpatient antibiotic prescribing for acute respiratory infections during influenza seasons. *JAMA Netw. Open* **1**, e180243 (2018).
2. Harris, A. M. et al. Appropriate antibiotic use for acute respiratory tract infection in adults: advice for high-value Care from the American college of physicians and the centers for disease control and prevention. *Ann. Intern. Med.* **164**, 425–434 (2016).
3. Smith, S. M., Fahey, T., Smucny, J. & Becker, L. A. Antibiotics for acute bronchitis. *Cochrane Database Syst. Rev.* **6**, CD000245 (2017).
4. Ong, D. S., Kuyvenhoven, M. M., van Dijk, L. & Verheij, T. J. Antibiotics for respiratory, ear and urinary tract disorders and consistency among GPs. *J. Antimicrob. Chemother.* **62**, 587–592 (2008).
5. Lee, G. C. et al. Outpatient antibiotic prescribing in the United States: 2000 to 2010. *BMC Med* **12**, 96 (2014).
6. Zumla, A. et al. Rapid point of care diagnostic tests for viral and bacterial respiratory tract infections-needs, advances, and future prospects. *Lancet Infect. Dis.* **14**, 1123–1135 (2014).
7. Caliendo, A. M. et al. Better tests, better care: improved diagnostics for infectious diseases. *Clin. Infect. Dis.* **57**, S139–S170 (2013). **Suppl 3**.
8. CDC. Antimicrobial Resistance. (2024).
9. Brendish, N. J. et al. Routine molecular point-of-care testing for respiratory viruses in adults presenting to hospital with acute respiratory illness (ResPOC): a pragmatic, open-label, randomised controlled trial. *Lancet Respir. Med.* **5**, 401–411 (2017).
10. Shiley, K. T., Lautenbach, E. & Lee, I. The use of antimicrobial agents after diagnosis of viral respiratory tract infections in hospitalized adults: antibiotics or anxiolytics?. *Infect. Control Hosp. Epidemiol.* **31**, 1177–1183 (2010).
11. Falsey, A. R. et al. Bacterial complications of respiratory tract viral illness: a comprehensive evaluation. *J. Infect. Dis.* **208**, 432–441 (2013).
12. Chaussabel, D., Pascual, V. & Banchereau, J. Assessing the human immune system through blood transcriptomics. *BMC Biol.* **8**, 84 (2010).
13. Ramilo, O. et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **109**, 2066–2077 (2007).
14. Hu, X., Yu, J., Crosby, S. D. & Storch, G. A. Gene expression profiles in febrile children with defined viral and bacterial infection. *Proc. Natl. Acad. Sci. USA* **110**, 12792–12797 (2013).
15. Tsalik, E. L. et al. Host gene expression classifiers diagnose acute respiratory illness etiology. *Sci. Transl. Med.* **8**, 322ra311 (2016).
16. Suarez, N. M. et al. Superiority of transcriptional profiling over procalcitonin for distinguishing bacterial from viral lower respiratory tract infections in hospitalized adults. *J. Infect. Dis.* **212**, 213–222 (2015).
17. Bhattacharya, S. et al. Transcriptomic biomarkers to discriminate bacterial from nonbacterial infection in adults hospitalized with respiratory illness. *Sci. Rep.* **7**, 6548 (2017).
18. Parnell, G. P. et al. A distinct influenza infection signature in the blood transcriptome of patients with severe community-acquired pneumonia. *Crit. Care* **16**, R157 (2012).
19. Zaas, A. K. et al. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. *Sci. Transl. Med.* **5**, 203ra126 (2013).
20. Ko, E. R. & Tsalik, E. L. A new era in host response biomarkers to guide precision medicine for infectious diseases. *J. Pediatr. Infect. Dis. Soc.* **11**, 477–479 (2022).
21. Ko, E. R. et al. Host-response transcriptional biomarkers accurately discriminate bacterial and viral infections of global relevance. *Sci. Rep.* **13**, 22554 (2023).
22. Rao, A. M. et al. A robust host-response-based signature distinguishes bacterial and viral infections across diverse global populations. *Cell Rep. Med.* **3**, 100842 (2022).
23. Xu, N. et al. A two-transcript biomarker of host classifier genes for discrimination of bacterial from viral infection in acute febrile illness: a multicentre discovery and validation study. *Lancet Digit. Health* **3**, e507–e516 (2021).
24. Bauer, W. et al. A novel 29-messenger RNA host-response assay from whole blood accurately identifies bacterial and viral infections in patients presenting to the emergency department with suspected infections: a prospective observational study. *Crit. Care Med.* **49**, 1664–1673 (2021).
25. Bauer, W. et al. Detection of viral infection and bacterial coinfection and superinfection in coronavirus disease 2019 patients presenting to the emergency department using the 29-mRNA host response classifier IMX-BVN-3: a multicenter study. *Open Forum Infect. Dis.* **9**, ofac437 (2022).
26. Ross, M. et al. A comparison of host response strategies to distinguish bacterial and viral infection. *PLoS ONE* **16**, e0261385 (2021).
27. Dedeoglu, B. E., Tanner, A. R., Brendish, N. J., Moyses, H. E. & Clark, T. W. Comparison of two rapid host-response tests for distinguishing bacterial and viral infection in adults with acute respiratory infection. *J. Infect.* **89**, 106360 (2024).
28. Ng, D. L. et al. A diagnostic host response biosignature for COVID-19 from RNA profiling of nasal swabs and blood. *Sci. Adv.* **7**, https://doi.org/10.1126/sciadv.abe5984 (2021).
29. McClain, M. T. et al. Dysregulated transcriptional responses to SARS-CoV-2 in the periphery. *Nat. Commun.* **12**, 1079 (2021).
30. Chew, T. et al. Molecular landscape of respiratory infection: A large-scale, multi-centre blood transcriptome dataset. *Sci. Data* **12**, 1175 (2025).
31. Viz-Lasheras, S. et al. A 5-transcript signature for discriminating viral and bacterial etiology in pediatric pneumonia. *iScience* **28**, 111747 (2025).
32. Falsey, A. R., Murata, Y. & Walsh, E. E. Impact of rapid diagnosis on management of adults hospitalized with influenza. *Arch. Intern. Med.* **167**, 354–360 (2007).
33. Cosgrove, S. E. The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. *Clin. Infect. Dis.* **42**, S82–S89 (2006).
34. Zaas, A. K. et al. The current epidemiology and clinical decisions surrounding acute respiratory infections. *Trends Mol. Med.* **20**, 579–588 (2014).

35. McDermott, K. W. & Roemer, M. Most frequent principle diagnoses for inpatient stays in U. S. Hospitals, 2018: Statistical Brief. (Agency for Healthcare Research and Quality, 2021).
36. Webster, K. E. et al. Diagnostic accuracy of point-of-care tests for acute respiratory infection: a systematic review of reviews. *Health Technol. Assess* 1–75 https://doi.org/10.3310/jlcp4570 (2024).
37. Metlay, J. P. et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and infectious diseases society of America. *Am. J. Respir. Crit. Care Med.* **200**, e45–e67 (2019).
38. Muller, B. et al. Diagnostic and prognostic accuracy of clinical and laboratory parameters in community-acquired pneumonia. *BMC Infect. Dis.* **7**, 10 (2007).
39. Zhang, W. et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol.* **16**, 133 (2015).
40. Villamayor, L. et al. The IFN-stimulated gene IFI27 counteracts innate immune responses after viral infections by interfering with RIG-I signaling. *Front. Microbio.* **14**, 1176177 (2023).
41. Bodkin, N. et al. Systematic comparison of published host gene expression signatures for bacterial/viral discrimination. *Genome Med.* **14**, 18 (2022).
42. Williams, D. J. et al. Transcriptomic biomarkers associated with microbiological etiology and disease severity in childhood pneumonia. *J. Infect. Dis.* **231**, e277–e289 (2024).
43. Du, X. et al. ITGB4 deficiency in airwayepithelium aggravates RSV infection and increases HDM sensitivity. *Front. Immunol.* **13**, 912095 (2022).
44. Ba, M. A. et al. Transgenic overexpression of alpha7 integrin in smooth muscle attenuates allergen-induced airway inflammation in a murine model of asthma. *FASEB Bioadv.* **4**, 724–740 (2022).
45. Teoh, C. M. et al. Integrin alpha7 expression is increased in asthmatic patients and its inhibition reduces Kras protein abundance in airway smooth muscle cells. *Sci. Rep.* **9**, 9892 (2019).
46. Lu, X. et al. Identification of key pathogenic genes of sepsis based on the Gene Expression Omnibus database. *Mol. Med. Rep.* **17**, 3042–3054 (2018).
47. Juss, J. K. et al. Acute respiratory distress syndrome neutrophils have a distinct phenotype and are resistant to phosphoinositide 3-Kinase inhibition. *Am. J. Respir. Crit. Care Med.* **194**, 961–973 (2016).
48. Nalbant, D. et al. FAM20: an evolutionarily conserved family of secreted proteins expressed in hematopoietic cells. *BMC Genomics* **6**, 11 (2005).
49. Huang, D. T., Yealy, D. M., Angus, D. C. & the Pro, A. C. T. I. Procalcitonin-guided antibiotic use. *N. Engl. J. Med.* **379**, 1973 (2018).
50. Chaussabel, D. et al. Analysis of significance patterns identifies ubiquitous and disease-specific gene-expression signatures in patient peripheral blood leukocytes. *Ann. N. Y Acad. Sci.* **1062**, 146–154 (2005).
51. Tsalik, E. L. et al. Discriminating bacterial and viral infection using a rapid host gene expression test. *Crit. Care Med.* **49**, 1651–1663 (2021).
52. Iglesias-Ussel, M. D. et al. A rapid host response blood test for bacterial/viral infection discrimination using a portable molecular diagnostic platform. *Open Forum Infect. Dis.* **12**, ofae729 (2025).
53. Papan, C. et al. A host signature based on TRAIL, IP-10, and CRP for reducing antibiotic overuse in children by differentiating bacterial from viral infections: a prospective, multicentre cohort study. *Clin. Microbiol. Infect.* **28**, 723–730 (2022).
54. Eden, E. et al. Diagnostic accuracy of a TRAIL, IP-10 and CRP combination for discriminating bacterial and viral etiologies at the Emergency Department. *J. Infect.* **73**, 177–180 (2016).
55. Kalmovich, B. et al. Impact on Patient Management of a Novel Host Response Test for Distinguishing Bacterial and Viral Infections: Real World Evience form the Urgent Care Setting. *Biomedicines* **11**, 1498 (2023).
56. Peterson, D. R. et al. Gene expression risk scores for COVID-19 illness severity. *J. Infect. Dis.* **227**, 322–331 (2023).
57. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

Acknowledgements

Supported by grant NIH/NIAID R01AI137364 (A.R.F. and T.J.M.) and the UR CTSI, supported in part by a Clinical and Translational Science Award (CTSA) (UL1 TR002001, KL2 TR001999, and TL1 TR002000) from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH) (AMC). We wish to acknowledge Mary Criddle and Sharon Moorehouse for assistance in participant recruitment and evaluation. We also acknowledge technical support from Jeffrey Malik, Cameron Baker, and Elizabeth Pritchett in the URMC Genomics Research Center.

Author contributions

A.R.F. -study design, data collection, clinical adjudication, data analysis, writing and review of the manuscript. DRP-statistical design, data analysis, writing and review of manuscript. E.E.W.-study design, data collection, clinical adjudication, data analysis, writing and review of the manuscript. S.B.-data analysis, writing and review of the manuscript. A.M.B.- data analysis, writing and review of the manuscript. C.C.- data analysis and review of the manuscript. A.R.B.- data collection, clinical adjudication, and review of the manuscript. D.C.- clinical adjudication, writing and review of the manuscript. M.P.- microbiologic evaluations, data analysis, review of the manuscript. A.M.C.- data management and analysis. J.A.- RNA sequencing, data analysis, review of the manuscript. T.J.M.- study design, data analysis, writing and review of the manuscript.

Competing interests

Edward Walsh declares grant support from Merck, Pfizer and received honoraria from Sanofi. Ann Falsey declares grant support from AstraZeneca, Merck, CyanVac, Moderna, Pfizer and consulting fees for ADMA Biologics, GSK, Sanofi, Merck and Shinogi. Angela Branche declares grant support from Cyanvac, Pfizer, Moderna, and Sanofi and serves as a consultant for GSK, Moderna, Merck, AstraZeneca and Sanofi. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-65361-3.

Correspondence and requests for materials should be addressed to Ann R. Falsey.

Peer review information *Nature Communications* thanks Gustavo Sganzerla, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.