

Rapid, Sample-to-Answer Host Gene Expression Test to Diagnose Viral Infection

Ephraim L. Tsalik,^{1,2,a} AyeAye Khine,³ Abdossamad Talebpour,³ Alaleh Samiei,³ Vilcy Parmar,³ Thomas W. Burke,¹ Micah T. McClain,^{1,4} Geoffrey S. Ginsburg,¹ Christopher W. Woods,^{1,4} Ricardo Henao,^{1,5} and Tino Alavie^{3,a}

¹Duke Center for Applied Genomics and Precision Medicine, Department of Medicine, Duke University School of Medicine, Durham, North Carolina, USA; ²Emergency Medicine Service, Durham Veterans Affairs Health Care System, Durham, North Carolina, USA; ³Qvella Corporation, Richmond Hill, Ontario, Canada; ⁴Medicine Service, Durham Veterans Affairs Health Care System, Durham, North Carolina, USA; ⁵Pratt School of Engineering, Duke University, Durham, North Carolina, USA

Objective. Distinguishing bacterial, viral, or other etiologies of acute illness is diagnostically challenging with significant implications for appropriate antimicrobial use. Host gene expression offers a promising approach, although no clinically useful test has been developed yet to accomplish this. Here, Qvella's FAST HR (Richmond Hill, Ontario, Canada) process was developed to quantify previously identified host gene expression signatures in whole blood in <45 minutes.

Method. Whole blood was collected from 128 human subjects (mean age 47, range 18–88) with clinically adjudicated, microbiologically confirmed viral infection, bacterial infection, noninfectious illness, or healthy controls. Stabilized mRNA was released from cleaned and stabilized RNA-surfactant complexes using e-lysis, an electrical process providing a quantitative real-time reverse transcription polymerase chain reaction-ready sample. Threshold cycle values (C_T) for 10 host response targets were normalized to hypoxanthine phosphoribosyltransferase 1 expression, a control mRNA. The transcripts in the signature were specifically chosen to discriminate viral from nonviral infection (bacterial, noninfectious illness, or healthy). Classification accuracy was determined using cross-validated sparse logistic regression.

Results. Reproducibility of mRNA quantification was within 1 cycle as compared to the difference seen between subjects with viral versus nonviral infection (up to 5.0 normalized C_T difference). Classification of 128 subjects into viral or nonviral etiologies demonstrated 90.6% overall accuracy compared to 82.0% for procalcitonin ($P = .06$). FAST HR achieved rapid and accurate measurement of the host response to viral infection in less than 45 minutes.

Conclusions. These results demonstrate the ability to translate host gene expression signatures to clinical platforms for use in patients with suspected infection.

Clinical Trials Registration. NCT00258869.

Key words: gene expression profiling; infection; molecular diagnostic techniques; real-time polymerase chain reaction; virus diseases.

INTRODUCTION

Respiratory tract infections are among the most common reasons for healthcare visits. Among the challenges in managing these presentations is the ambiguity of whether symptoms are caused by a viral pathogen, bacterial pathogen, or noninfectious etiologies. Routinely available clinical information inadequately discriminates these etiologies, resulting in excessively high rates of inappropriate antibiotic prescriptions. Despite a viral etiology in the majority of cases, 73% of ambulatory care patients in the United States with acute respiratory infection (ARI) are

prescribed an antibiotic, accounting for 41% of all antibiotics prescribed in this setting [1, 2]. Diagnostics that discriminate between viral and nonviral etiologies offer a means to dispel this uncertainty and improve antibiotic management for such patients.

Currently available diagnostic strategies largely focus on pathogen identification. This includes rapid antigen tests, culture, and nucleic acid amplification tests. These diagnostic tools are indispensable but, depending on the approach, come with limitations such as poor sensitivity, long time to result, inability to discriminate infection from colonization, or a priori suspicion of the etiologic agent. A complementary strategy focuses on the host's response to illness. Infection manifests when pathogens induce a maladaptive host response associated with clinical illness. That host response is pathogen class specific and, in some cases, has been shown to be pathogen specific [3–5]. Host response markers historically have been proteins, such as C-reactive protein or, more recently, procalcitonin. Recent developments, however, have shown that peripheral blood gene

Received 4 October 2019; editorial decision 16 October 2019; accepted 23 October 2019.

aE. L. T. and T. A. contributed equally to this work.

Correspondence: Ephraim L. Tsalik, MD, MHS, PhD, DUMC 102359, Durham, NC 27710. E-mail: e.t@duke.edu

Open Forum Infectious Diseases®

Published by Oxford University Press on behalf of Infectious Diseases Society of America 2019. This work is written by (a) US Government employee(s) and is in the public domain in the US. DOI: 10.1093/ofid/ofz466

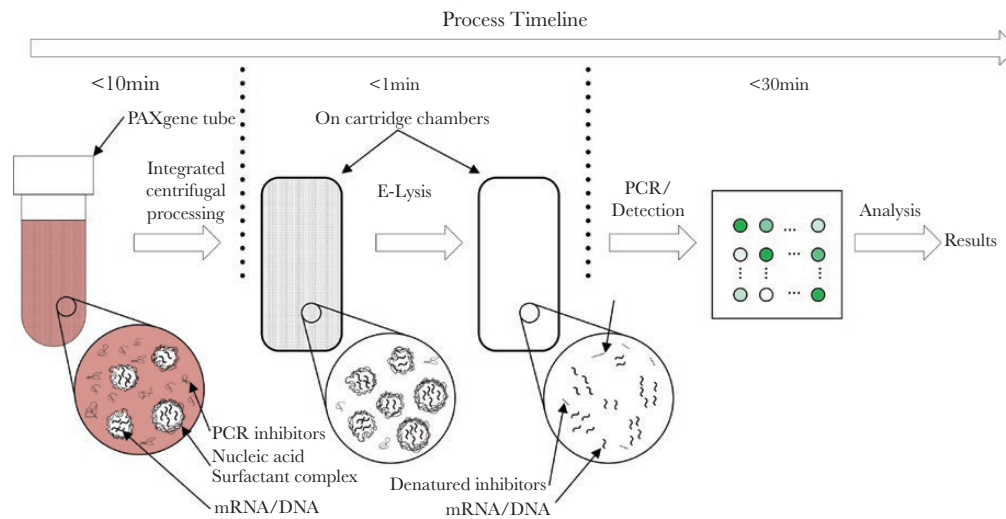


Figure 1. The FAST HR process separates stabilized mRNA from debris in the PAXgene Blood RNA tubes and releases them in a nuclease-free and inhibitor-free liquid medium suitable for performing quantitative reverse transcription polymerase chain reaction (PCR) assays.

expression offers a more robust ability to discriminate bacterial, viral, or noninfectious etiologies [3, 4, 6–10].

These advances, however, have been limited to the research setting in part because of technical limitations. A platform suitable for clinical use would have to offer sample-to-answer capability, provide results quickly (eg, <1 hour), be adequately quantitative so as to measure differential gene expression, and offer sufficient multiplexing so all transcripts in the signature can be detected. To address this need, we developed Qvella's FAST HR (Richmond Hill, Ontario, Canada) sample processing and detection platform to develop a rapid (45 minute), fully-integrated, sample-to-answer system that measures the host response to viral infection allowing for the discrimination of viral and nonviral etiologies of acute illness [Figure 1](#).

METHODS

Study Design

Studies were approved by relevant Institutional Review Boards and in accordance with the Declaration of Helsinki. All subjects or their legally authorized representatives provided written informed consent.

Patients with community-onset, suspected infection were enrolled in the Emergency Departments (ED) of Duke University Medical Center (Durham, NC) or the Durham VA Health Care System (Durham, NC) as part of the Community Acquired Pneumonia and Sepsis Outcome Diagnostic (CAPSOD) study (ClinicalTrials.gov NCT00258869) [11–13]. The CAPSOD study was a prospective sample collection study targeting community-acquired pneumonia or suspected sepsis at the time of ED presentation. Additional patients were enrolled through the University of North Carolina Health Care Emergency Department (Chapel Hill, NC) as part of the Community

Acquired Pneumonia and Sepsis Study. For both studies, enrollment was through convenience sampling. Patients were eligible if they had a known or suspected infection and if they exhibited 2 or more systemic inflammatory response syndrome (SIRS) criteria [14]. Upon adjudication and subject selection (described in the online Supplementary Data), banked samples were accessed and analyzed.

Sample Collection and Processing

At initial clinical presentation, subjects were enrolled and whole blood was collected directly into Qiagen PAXgene Blood RNA tubes (Germantown, MD) and stored at -80°C . Samples were later thawed to generate aliquots, which were refrozen and maintained at -80°C until use. For testing, 100 μL (~ 27 uL whole blood volume) was used to measure the relative abundance of target mRNA through real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR). All sample processing steps as described in the online Supplementary Data were performed on the integrated instrument followed by qRT-PCR (Illumina Eco Real-Time PCR System, San Diego, CA).

Procalcitonin Measurements

Concentrations were measured at different times during the study and, as a result, different platforms were utilized. Serum measurements were made on a Roche Elecsys 2010 analyzer (Roche Diagnostics, Basel, Switzerland) or the miniVIDAS immunoassay (bioMerieux, Marcy-l'Étoile, France). Plasma measurements were made using the B-R-A-H-M-S PCT sensitive KRYPTOR (Thermo Fisher Scientific, Waltham, MA). All procalcitonin measurements were treated equivalently, regardless of testing platform with results reported as ng/ml. Neither serum nor plasma was available from the healthy controls used in this study and, therefore, procalcitonin was not measured for that group.

Signature Translation and Statistical Analysis

The host response transcripts used for test development derived from previously published research, which identified signatures characteristic of viral versus nonviral infection [9, 15]. The transcripts in these signatures were prioritized based on their importance to the model that predicted disease status. Assays were developed for 20 top-performing transcripts for use on the FAST HR test. Normalization to hypoxanthine phosphoribosyltransferase 1 (HPRT1) was performed by subtracting its C_T value from that of each target in the log domain. These normalized counts then were used to build an elastic logistic regression model to discriminate viral from nonviral infections [16]. Targets with variability comparable to that of the reference (HPRT1) were unlikely to contribute to classification. To avoid setting an arbitrary cutoff for the variance (and by extension, the number of targets included in the model), the cutoff was selected using cross-validation (CV). The variance cutoff, regularization parameter of the model, and performance metrics were estimated using nested CV, where the internal 10-fold CV was used for the variance cutoff and the regularization parameter. The outer leave-one-out CV (LOOCV) was used to define performance estimates: area under the receiver operating characteristic curve (AUC), positive percent agreement (PPA), and negative percent agreement (NPA) [17]. For each LOOCV fold, the subset of N-1 observations was first used to sort targets by variance to then build 10-fold CV models for 1 through all ($d = 20$) target inputs selected by variance. This CV procedure also accounted for different regularization parameters using the regularization path approach [16]. The CV model with the best AUC, corresponding to 1 specific subset of target inputs and a regularization parameter value, was used to make a prediction on the left-out observation. This procedure was repeated N times as part of LOOCV. Where appropriate, 95% confidence intervals are reported. The threshold used to differentiate viral from nonviral infection was selected based on the empirical prevalence of 27% observed in this study. The χ^2 test was used to compare procalcitonin to gene expression.

RESULTS

Clinical Cohort

Eighty-nine subjects were selected based on clinically adjudicated, acute illness with confirmatory microbiology as well as

39 healthy controls. Mean age was 47 years (range 18–88 years), 42% were male, and 65% were hospitalized (excluding healthy controls). The cohort included 34 subjects with confirmed acute viral etiologies, 30 with confirmed acute bacterial etiologies, 25 with acute noninfectious illness, and 39 healthy controls (Table 1). Among subjects with bacterial infection, 13 had lower respiratory tract infections, 5 had urinary tract infections, and the remainder represented a variety of infection sites (Supplementary Table 1). Of the subjects with viral infection, all were respiratory in nature and included 7 different viral etiologies; influenza ($n = 14$) and human metapneumovirus ($n = 7$) were the most common. The noninfectious conditions were varied, although all had respiratory symptoms and clinical suspicion of infection at initial presentation. Additional details regarding each ill subject are presented in Supplementary Table 1.

Normalization

Host gene expression-based tests must be quantitative, because the amount of each transcript in the signature is what conveys biological meaning. However, there are a number of parameters that influence transcript abundance. The most relevant is the underlying biological state of the patient. Other sources of variability include white blood cell count, leukocyte differential, and variability in the sample volume loaded onto the test. In order to control for these variables, a reference transcript was included in the signature that does not vary as a function of biological state. We first evaluated 6 housekeeping genes commonly used as controls for qRT-PCR. Primers were designed for Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), Beta actin (ACTB), hypoxanthine phosphoribosyltransferase 1 (HPRT1), Heat shock 90kD protein 1, beta (HSPCB), Succinate dehydrogenase (SDHA), and TATA-box-binding protein (TBP). Two criteria were used to select the normalization assay for the host response viral test: (1) minimal or no gDNA detection and (2) C_T value that was close to the average C_T value for transcripts in the viral host response signature. Based on these criteria, HPRT1 was selected as the normalization control. Specifically, gDNA was not amplified (Supplementary Figure 1) and the C_T value for HPRT1 in control subjects was 0.12 cycles higher than the average C_T values observed for the viral host response transcripts (Supplementary Table 2).

Table 1. Cohort Characteristics

	Number of Subjects	Gender (Male/Female)	Mean Age, Years (Range)	Ethnicity (Black/White/Other)	Admitted
Total Cohort	128	54/74	47 (18–88)	41/65/22	
Viral	34	10/24	47 (18–88)	19/12/3	35%
Bacterial	30	15/15	54 (24–85)	12/18/0	93%
Noninfectious illness	25	9/16	56 (20–83)	10/14/1	72%
Healthy	39	20/19	36 (20–59)	0/21/18	0%

The transcripts in the viral host response signature were selected from our prior research [9, 15] in which we identified differentially expressed genes that discriminated viral from nonviral illness, largely representing the interferon signaling and response pathway. Having defined the signature for translation, we then verified the reproducibility of the FAST HR process for quantifying transcripts using PAXgene Blood RNA samples from 3 healthy individuals. Each sample was processed 5 times and qRT-PCR was performed to measure the mRNA levels of 2 viral host response transcripts, 2'-5'-Oligoadenylate Synthetase (OAS)1 and OAS2, and the HPRT1 normalization transcript. Transcript quantification was based on C_T values. A summary of the results on the 5 replicates is presented in [Supplementary Table 2](#). Precision was high based on a standard deviation across replicates of <1 cycle for OAS1, OAS2, and HPRT1.

The relative expression level (ΔC_T) of target transcripts was specified by the difference between the threshold C_T values for that target relative to HPRT1 (ie, target C_T - HPRT1 C_T). A negative ΔC_T indicates that the transcript had a higher expression level relative to HPRT1 whereas a positive ΔC_T indicates a lower transcript abundance. In order to account for background noise, a ΔC_T value of 10 was assigned for scenarios where no detection occurred. The average ΔC_T for all 20 viral host response targets stratified by clinical subgroup are presented in [Table 2](#). The signature measured here includes transcripts that are upregulated in response to viral infection. Consistent with this, we observed marked differences in the ΔC_T for transcripts in the viral cohort as compared to the nonviral groups. For example, the average normalized ΔC_T of IFI27 was 5.0 cycles lower in subjects with viral infection compared to nonviral subjects, corresponding to a 32-fold difference in IFI27 abundance. Similarly, a greater than 3-cycle difference in average ΔC_T (ie, 8-fold difference in

abundance) was observed for OAS3, RSAD2, IFIT1, MX1, and ISG15. However, some transcripts demonstrated low variance similar to the HPRT1 normalization transcript, making them unhelpful for classification. Using CV as described in Methods, we identified an optimal signature comprising 10 transcripts: Serpin Family G Member 1 (SERPING1), OAS3, Radical S-Adenosyl Methionine Domain-Containing Protein 2 (RSAD2), Interferon Induced Protein With Tetratricopeptide Repeats 1 (IFIT1), Septin 4 (SEPT4), MX Dynamin Like GTPase 1 (MX1), ISG15 Ubiquitin Like Modifier (ISG15), Interferon Induced Protein With Tetratricopeptide Repeats 3 (IFIT3), Interferon Alpha Inducible Protein 27 (IFI27), and OAS1 ([Figure 2](#)).

Classification

The host gene expression signature utilized for this test was developed for the purpose of discriminating viral from nonviral infection [9, 15]. It is not expected to distinguish nonviral etiologies from each other. Specifically, it would not discriminate bacterial from noninfectious etiologies or from healthy controls as all are considered nonviral. Gene expression data were normalized and a CV sparse logistic regression model was used to assign subjects a probability of viral infection as described in Methods. Using a probability threshold of 27% (empirical prevalence of viral infection in this cohort), 32 of 34 viral infection cases were correctly identified corresponding to a PPA of 94% (95% confidence interval [CI], 80.3%–99.3%) compared to clinical adjudication as the reference standard. Moreover, 84 of 94 nonviral cases agreed with the adjudicated diagnoses corresponding to a NPA of 89.4% (95% CI, 81.3%–94.8%) ([Figure 3](#)). The overall accuracy was therefore 90.6% (95% CI, 84.2%–95.1%), AUC was 0.95, positive likelihood ratio was 8.9 (95% CI, 4.9–16), and negative likelihood ratio was 0.07 (95%

Table 2. Changes in Transcript Abundance Stratified by Clinical Categories. Average Change in C_T Value Relative to HPRT1 are Reported for Each Target in the Viral Host Response Test. Results are Stratified by Clinical Group, Followed by Averages for Control Groups, Either With or Without the Inclusion of Healthy Subjects. The Last Two Rows Provide the Differences in Normalized C_T Values Between the Viral Infection Group and Control Groups, Either With or Without the Inclusion of Healthy Subjects

Clinical Categories	Average ΔC_T																			
	OAS2	LAMP3	SERPING1	OAS1	SIGLEC1	OAS3	RSAD2	IFIT1	SEPT4	GBP1	ADAR	DDX58	MX1	ISG15	TNFAIP6	IFIT3	IFI44	IFI6	RTP4	IFI27
Healthy (H)	-1.52	-1.96	-4.71	-0.33	-2.79	1.05	0.39	-1.02	2.58	-1.44	-2.65	-0.64	-0.96	2.99	0.10	-2.40	-0.98	-3.91	1.01	4.20
Bacterial (B)	-1.22	0.36	-2.50	-0.67	-0.92	1.40	0.84	0.17	1.98	-2.09	-3.02	-1.11	-0.77	3.66	-1.54	-1.30	-1.10	-3.14	1.27	3.97
Noninfectious (N)	-1.51	-0.01	-2.31	-0.45	-0.61	0.99	0.91	-0.95	3.18	-1.40	-3.04	-1.13	-1.06	4.10	-0.92	-2.36	-0.89	-3.32	1.85	4.32
Viral (V)	-4.07	-0.38	-3.27	-3.45	-2.28	-2.19	-2.98	-4.98	0.39	-3.28	-3.61	-2.85	-4.14	0.21	-2.07	-4.29	-3.49	-4.43	-1.02	-0.80
Nonviral samples (average of H + B + N)	-1.42	-0.54	-3.17	-0.48	-1.44	1.15	0.71	-0.60	2.58	-1.64	-2.91	-0.96	-0.93	3.58	-0.79	-2.02	-0.99	-3.46	1.38	4.16
Nonviral illnesses (average of B + N)	-1.37	0.18	-2.40	-0.56	-0.76	1.20	0.87	-0.39	2.58	-1.74	-3.03	-1.12	-0.92	3.88	-1.23	-1.83	-1.00	-3.23	1.56	4.14
Viral – nonviral samples	-2.65	0.16	-0.10	-2.97	-0.84	-3.33	-3.69	-4.38	-2.18	-1.64	-0.70	-1.88	-3.21	-3.38	-1.28	-2.27	-2.50	-0.98	-2.40	-4.97
Viral – Nonviral illnesses	-2.70	-0.56	-0.87	-2.89	-1.52	-3.38	-3.85	-4.59	-2.18	-1.54	-0.58	-1.72	-3.23	-3.67	-0.84	-2.46	-2.49	-1.20	-2.58	-4.95

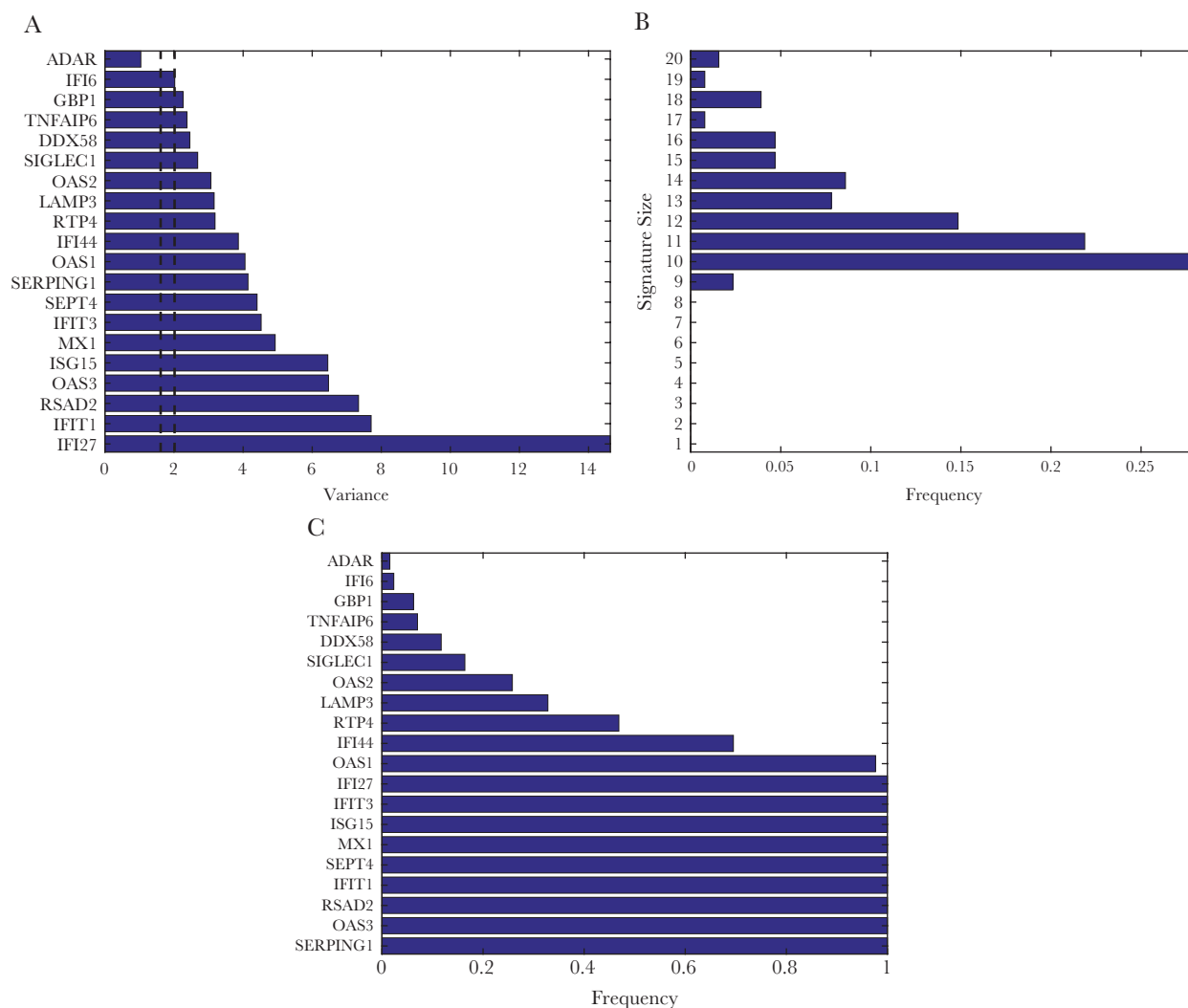


Figure 2. Target Selection by Variance. A, sorted variance of all targets in the assay; dashed lines represent 95% empirical quantiles of the variance of the normalizing target (HPRT1). B, distribution of signature size obtained by leave-one-out cross-validation (LOOCV); the most frequent signature size was 10 and signatures smaller than 9 targets were never selected. C, frequency of targets selected by LOOCV; 10 targets were selected in more than 90% of the cases and 9 targets were selected fewer than 50% of the cases.

CI, 0.02–0.25). Among the bacterial or noninfectious cases that were classified as viral, we did not detect a viral pathogen through multiplex viral pathogen detection testing. We compared test performance stratified by clinical severity as indicated by hospitalization (excluding healthy subjects) and observed no significant difference: 90.3% overall accuracy for the nonhospitalized group versus 84.5% for the hospitalized group ($P = .45$). We also observed no difference in the number of Gram-positive and Gram-negative infections that were correctly classified: 15 of 17 (88%) and 9 of 11 (82%), respectively.

We compared classification performance to procalcitonin, a peptide biomarker commonly used to differentiate bacterial from nonbacterial infections. Values <0.25 ng/ml indicated the absence of a bacterial infection whereas values ≥ 0.25 ng/ml indicated the presence of a bacterial infection [18]. Subjects with adjudicated viral infection had procalcitonin values <0.25 ng/

ml in 30 of 34 cases (88%). Among cases of adjudicated noninfectious illness, 20 of 25 subjects (80%) had low procalcitonin levels. Twenty-three of 30 subjects (77%) with adjudicated bacterial infection had elevated procalcitonin concentrations. Excluding healthy subjects for whom procalcitonin was not measured, the overall accuracy across all symptomatic subjects was 82.0% (95% CI, 73%–89%). This was lower than the 90.6% (95% CI, 84.2%–95.1%) accuracy for the gene expression test when measured in all subjects ($P = .06$) but similar to the 86.5% (95% CI, 79%–94%) accuracy observed when restricting to those subjects who had procalcitonin measured.

DISCUSSION

The host response to infection as measured through gene expression patterns in peripheral blood offers unique diagnostic

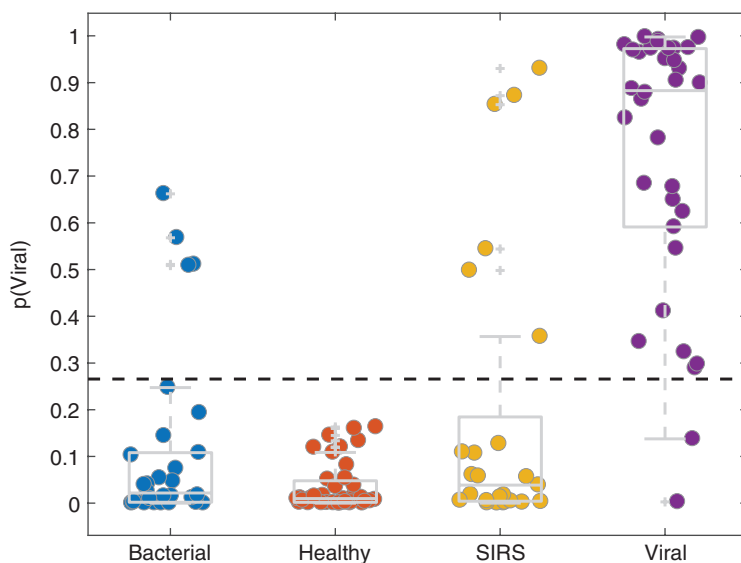


Figure 3. *Classification Performance.* Whole blood from 128 subjects was collected in PAXgene Blood RNA tubes and used to measure host response gene expression in response to viral infection. A logistic regression equation was used to generate a probability of viral infection based on the level of expression of mRNAs in the signature. This signature classifies subjects as viral or nonviral. Therefore, all cases of nonviral infection (bacterial, healthy, and noninfectious illness (NI)) are expected to appear similar. Each symbol represents 1 subject. The box denotes the interquartile range. Plus signs (+) represent statistical outliers. The dashed line at $p(\text{Viral}) = 0.27$ represents the prevalence of viral ARI in the study cohort.

and prognostic potential. This includes defining whether infection is present [3], identifying presymptomatic infection [7], defining the pathogen class [15], distinguishing sepsis from noninfectious etiologies of illness [19], and predicting sepsis outcomes [20], among others. Although these host response signatures have been well described in the literature, there are no currently available host gene expression tests for infectious diseases. Whereas some fields, such as oncology, may be able to tolerate delayed results or a complex assay performed in a central laboratory, the care of patients with infectious diseases requires rapid diagnosis in order to inform real-time clinical decision making. In this study, we have for the first time developed a system that accurately distinguishes viral from nonviral syndromes by measuring a host gene expression signature in as little as 45 minutes.

The host response as a diagnostic approach has several advantages over molecular pathogen detection tests. The host response is conserved to viral infections generally and is not limited to the pathogens included on a pathogen detection panel [6, 21]. Identifying a viral pathogen does not exclude bacterial/viral coinfection, whereas the host response can detect the presence of both [22]. In the subjects tested here, several cases of adjudicated bacterial or noninfectious illness were classified as viral. Although these could represent cases of coinfection, we did not detect a viral pathogen using a commercial respiratory pathogen panel. Finally, detecting a pathogen does not make it the cause of illness considering the high rates of asymptomatic carriage and shedding in both pediatric and adult populations [23–25].

Several studies have described host gene expression signatures that discriminate viral from bacterial infection. In most cases, these comprise a single signature that assigns subjects as belonging to 1 class or the other [3, 8–10, 21, 26, 27]. Because the host response to viral infection is dominated by very robust interferon-related pathways, any subject with a nonviral process, such as healthy, bacterial infection, or noninfectious illness, will look similar. It is therefore more accurate to consider these signatures as classifying viral versus nonviral syndromes. The results presented in this study validate a previously characterized viral versus nonviral signature [9, 15]. As a first step in the translation of host gene expression signatures to a clinically relevant platform, we developed a host response test for viral versus nonviral infection. In doing so, we have demonstrated the feasibility of Qvella's FAST HR process as a rapid (~45 minutes) and simple, low-complexity, sample-to-answer system for monitoring the host response to infection and discriminating viral from nonviral etiologies.

We began by evaluating 20 transcripts but determined that 10 represented the optimal signature size. This is similar to the signature size described in other studies, although some have demonstrated discrimination of viral from nonviral infection with as few as 1–2 transcripts [27–29]. The gene expression-based test described herein performed marginally better than the commonly used peptide biomarker, procalcitonin. This is in contrast to most other studies comparing gene expression to procalcitonin, which showed superior performance of gene expression [15, 26, 30]. This may be because of a relatively small sample size or better-than-expected performance

of procalcitonin observed in our study. For example, a recent meta-analysis evaluating procalcitonin for its ability to discriminate bacterial and viral etiologies of pneumonia demonstrated only 55% sensitivity and 76% specificity [31].

In order to develop a host gene expression test, there are a number of technical hurdles that must be overcome. This includes sample preparation techniques to isolate mRNA from blood into a matrix free of RNase enzymes and qRT-PCR inhibitors. This is not easily amenable to full automation on a single cartridge that also performs amplification and quantification. Qvella's FAST HR process circumvents this issue by centrifugally removing qRT-PCR inhibitors, separating them from the nucleic acid content. In addition, RNase enzymes are inactivated by electrically-induced flash heating. The result is a suspension of nucleic acid in a qRT-PCR-ready matrix, which can be readily amplified and quantified on the instrument. Carrying out all these steps on a fully integrated platform results in a low-complexity assay with no preprocessing and a time to result of approximately 45 minutes. Measuring a host gene expression signature using conventional laboratory processes (nucleic acid extraction and purification, reverse transcription and multiplexed amplification, and real-time detection) would require 6–8 hours in a high complexity laboratory environment [22]. The 45-minute time-to-result for the FAST HR system is comparable to the time necessary to measure procalcitonin, which is about 20 minutes without accounting for preprocessing time.

One limitation of this approach is the inability to differentiate between various causes of nonviral infection. Specifically, this signature does not discriminate bacterial from noninfectious etiologies. This may not be a problem if the test is used in cases of confirmed infection where the only question is whether the etiology is bacterial or viral. In that scenario, a low probability of viral infection by default means the illness is bacterial. This oversimplification fails to consider cases of coinfection nor the difficulty in determining whether an infection is present at all. Therefore, the most clinically useful test would measure the host response to viral infection, bacterial infection, coinfection, or no infection. In addition, we selected cases of microbiologically confirmed bacterial and viral infections. However, most cases of acute respiratory illness have no microbiological diagnosis despite rigorous testing [32]. When more extensive validation is performed that includes these more ambiguous cases, it is possible that test accuracy will be lower. As this study was primarily focused on the translation of a host gene expression signature to a clinically useful platform, validation in larger and more heterogeneous cohorts will be needed.

In future work, the test will require validation in larger cohorts, including those with chronic viral infections (eg, HIV, hepatitis C) and those with immunocompromise. Pediatric patients, particularly the very young, as well as pregnant women should be evaluated given the possibility that their immune responses may differ from the adult population represented

here. A larger validation cohort also would identify any potential impact of comorbidities, such as chronic kidney disease, chronic obstructive pulmonary disease, diabetes, and others on the test's performance. The test also should be expanded to include components of a bacterial host response signature for reasons already stated, which also will require testing in subjects with ambiguous phenotypes, such as those with bacterial/viral coinfection, those with suspected infection but no confirmatory microbiology, and those with atypical (eg, intracellular bacterial) pathogens. Moreover, Qvella's FAST HR technology can be used to measure other host gene expression signatures to offer better diagnostic and prognostic tools for other infectious disease-related questions.

Advances in transcriptomics and machine learning have identified an ever-growing number of diagnostic and prognostic host gene expression signatures. None of the signatures currently used in clinical care are available rapidly enough to be helpful in the context of acute infectious disease. This study is the first to demonstrate the translation of a host response signature for viral versus nonviral infection to a clinically-suitable platform. This rapid, simple, sample-to-answer system demonstrates the feasibility of measuring host response signatures, which opens up a new pathway for diagnostic and prognostic test development.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We are grateful for the contributions made by our many colleagues in the execution of the CAPSOD, CAPSS, and ARLG/RADICAL studies, which enabled us to perform the studies reported here.

Author Contribution. E.L.T., A.K., A.T., T.W.B., M.T.M., G.S.G., C.W.W., R.H., and T.A. conceived and designed the analysis. E.L.T., A.K., A.T., A.S., V.P., and T.W.B. collected data. E.L.T., A.T., and R.H. performed analysis. E.L.T., A.K., A.T., T.W.B., M.T.M., G.S.G., C.W.W., R.H., and T.A. performed data interpretation. E.L.T. and A.T. wrote the paper.

Financial support. This work was supported in part by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (UM1AI104681). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. It also was supported in part by the U.S. DARPA (N66001-09-C-2082). Qvella supported the work performed by its employees A.K., S.T., A.S., V.P., and T.A..

Potential conflicts of interest. E.L.T., T.B., G.S.G., and C.W.W. have equity in Predigen, Inc. S.T. and A.A.K. have pending patent applications related to sample preparation methods for gene expression analysis. E.L.T., T.W.B., M.T.M., G.S.G., C.W.W., and R.H. have submitted, pending, or granted patents related to the gene expression classifier. A.K., A.T., A.S., V.P., and T.A. are employees of Qvella Corp. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Shapiro DJ, Hicks LA, Pavia AT, Hersh AL. Antibiotic prescribing for adults in ambulatory care in the USA, 2007-09. *J Antimicrob Chemother* **2014**; 69:234–40.
2. Lee GC, Reveles KR, Attridge RT, et al. Outpatient antibiotic prescribing in the United States: 2000 to 2010. *BMC Med* **2014**; 12:96.
3. Ramilo O, Allman W, Chung W, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **2007**; 109:2066–77.
4. Mejias A, Dimo B, Suarez NM, et al. Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. *PLOS Med* **2013**; 10:e1001549.
5. Ahn SH, Tsalik EL, Cyr DD, et al. Gene expression-based classifiers identify *Staphylococcus aureus* infection in mice and humans. *PLOS ONE* **2013**; 8:e48979.
6. Zaas AK, Chen M, Varkey J, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe* **2009**; 6:207–17.
7. Woods CW, McClain MT, Chen M, et al. A host transcriptional signature for presymptomatic detection of infection in humans exposed to influenza H1N1 or H3N2. *PLOS ONE* **2013**; 8:e52198.
8. Parnell GP, McLean AS, Booth DR, et al. A distinct influenza infection signature in the blood transcriptome of patients with severe community-acquired pneumonia. *Crit Care* **2012**; 16:R157.
9. Zaas AK, Burke T, Chen M, et al. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. *Sci Transl Med* **2013**; 5:203ra126.
10. Hu X, Yu J, Crosby SD, Storch GA. Gene expression profiles in febrile children with defined viral and bacterial infection. *Proc Natl Acad Sci* **2013**; 110:12792–7.
11. Langley RJ, Tsalik EL, van Velkinburgh JC, et al. An integrated clinico-metabolomic model improves prediction of death in sepsis. *Sci Transl Med* **2013**; 5:195ra95.
12. Tsalik EL, Jagers LB, Glickman SW, et al. Discriminative value of inflammatory biomarkers for suspected sepsis. *J Emerg Med* **2012**; 43:97–106.
13. Glickman SW, Cairns CB, Otero RM, et al. Disease progression in hemodynamically stable patients presenting to the emergency department with sepsis. *Acad Emerg Med* **2010**; 17:383–90.
14. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest* **1992**; 101:1644–55.
15. Tsalik EL, Henao R, Nichols M, et al. Host gene expression classifiers diagnose acute respiratory illness etiology. *Sci Transl Med* **2016**; 8:322ra11.
16. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer; **2001**.
17. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* **2006**; 27:861–74.
18. Schuetz P, Wirz Y, Sager R, et al. Procalcitonin to initiate or discontinue antibiotics in acute respiratory tract infections. *Cochrane Database Syst Rev* **2017**; 10:CD007498.
19. McHugh L, Seldon TA, Brandon RA, et al. A molecular host response assay to discriminate between sepsis and infection-negative systemic inflammation in critically ill patients: discovery and validation in independent cohorts. *PLOS Med* **2015**; 12:e1001916.
20. Sweeney TE, Perumal TM, Henao R, et al. A community approach to mortality prediction in sepsis via gene expression analysis. *Nat Commun* **2018**; 9:694.
21. Sampson DL, Fox BA, Yager TD, et al. A four-biomarker blood signature discriminates systemic inflammation due to viral infection versus other etiologies. *Sci Rep* **2017**; 7:2914.
22. Lydon EC, Henao R, Burke TW, et al. Validation of a host response test to distinguish bacterial and viral respiratory infection. *EBioMedicine*. **2019**; 48:453–61.
23. Byington CL, Ampofo K, Stockmann C, et al. Community surveillance of respiratory viruses among families in the Utah better identification of germs-longitudinal viral epidemiology (BIG-LoVE) study. *Clin Infect Dis* **2015**; 61:1217–24.
24. Self WH, Williams DJ, Zhu Y, et al. Respiratory viral detection in children and adults: comparing asymptomatic controls and patients with community-acquired pneumonia. *J Infect Dis* **2016**; 213:584–91.
25. Shaman J, Morita H, Birger R, et al. Asymptomatic summertime shedding of respiratory viruses. *J Infect Dis* **2018**; 217:1074–7.
26. Suarez NM, Bunsow E, Falsey AR, Walsh EE, Mejias A, Ramilo O. Superiority of transcriptional profiling over procalcitonin for distinguishing bacterial from viral lower respiratory tract infections in hospitalized adults. *J Infect Dis* **2015**; 212:213–22.
27. Herberg JA, Kaforou M, Wright VJ, et al; IRIS Consortium. Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children. *JAMA* **2016**; 316:835–45.
28. Tang BM, Shojaei M, Parnell GP, et al. A novel immune biomarker IFI27 discriminates between influenza and bacteria in patients with suspected respiratory infection. *Eur Respir J* **2017**; 49:1602098. doi:10.1183/13993003.02098-2016.
29. Kaforou M, Herberg JA, Wright VJ, Coin LJM, Levin M. Diagnosis of bacterial infection using a 2-transcript host RNA signature in febrile infants 60 days or younger. *JAMA* **2017**; 317:1577–8.
30. Scicluna BP, Klein Klouwenberg PMC, van Vught LA, et al. A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. *Am J Respir Crit Care Med*. **2015**; 192:826–35.
31. Kamat IS, Ramachandran V, Eswaran H, Guffey D, Musher DM. Procalcitonin to distinguish viral from bacterial pneumonia: a systematic review and meta-analysis. *Clin Infect Dis*. **2019**. doi:10.1093/cid/ciz545. [Epub ahead of print]
32. Jain S, Self WH, Wunderink RG, et al; CDC EPIC Study Team. Community-acquired pneumonia requiring hospitalization among U.S. adults. *N Engl J Med* **2015**; 373:415–27.