MAJOR ARTICLE

Superiority of Transcriptional Profiling Over Procalcitonin for Distinguishing Bacterial From Viral Lower Respiratory Tract Infections in Hospitalized Adults

Nicolas M. Suarez,[1,2] Eleonora Bunsow,[1,2] Ann R. Falsey,[3,4] Edward E. Walsh,[3,4] Asuncion Mejias,[1,2] and Octavio Ramilo[1,2]

[1]Center for Vaccines and Immunity, and [2]Division of Pediatric Infectious Diseases, The Research Institute at Nationwide Children's Hospital, and The Ohio State University College of Medicine, Columbus; [3]Department of Medicine, University of Rochester, and [4]Rochester General Hospital, New York

(See the editorial commentary by Tsalik, McClain, and Zaas on pages 173–5.)

Background. Distinguishing between bacterial and viral lower respiratory tract infection (LRTI) remains challenging. Transcriptional profiling is a promising tool for improving diagnosis in LRTI.

Methods. We performed whole blood transcriptional analysis in 118 patients (median age [interquartile range], 61 [50–76] years) hospitalized with LRTI and 40 age-matched healthy controls (median age, 60 [46–70] years). We applied class comparisons, modular analysis, and class prediction algorithms to identify and validate diagnostic biosignatures for bacterial and viral LRTI.

Results. Patients were classified as having bacterial (n = 22), viral (n = 71), or bacterial-viral LRTI (n = 25) based on comprehensive microbiologic testing. Compared with healthy controls, statistical group comparisons ($P < .01$; multiple-test corrections) identified 3376 differentially expressed genes in patients with bacterial LRTI, 2391 in viral LRTI, and 2628 in bacterial-viral LRTI. Patients with bacterial LRTI showed significant overexpression of inflammation and neutrophil genes (bacterial > bacterial-viral > viral), and those with viral LRTI displayed significantly greater overexpression of interferon genes (viral > bacterial-viral > bacterial). The K–nearest neighbors algorithm identified 10 classifier genes that discriminated between bacterial and viral LRTI with a 95% sensitivity (95% confidence interval, 77%–100%) and 92% specificity (77%–98%), compared with a sensitivity of 38% (18%–62%) and a specificity of 91% (76%–98%) for procalcitonin.

Conclusions. Transcriptional profiling is a helpful tool for diagnosis of LRTI.

Keywords. microarrays; lower respiratory tract infection; procalcitonin; viral infections; bacterial infections.

Lower respiratory tract infections (LRTIs) are among the most frequent reasons for hospitalization in adults. However, the spectrum of the disease is broad, and adult patients with LRTI can present with a minor self-limiting illness to potentially life-threatening disease [1, 2]. Until recently, bacterial pathogens were considered the leading cause of LRTI/pneumonia in adults. However, there is increasing evidence suggesting that viral respiratory infections also play an important role [3, 4]. Antibiotic resistance, the increased adverse drug reactions and the costs of antimicrobial therapy should be considered in order to implement a rational use of antimicrobials. Hence, elucidating the causes of LRTIs has become a pivotal factor for improving patient management.

To tackle this issue, in recent years there has been increased interest in the development of rapid and accurate diagnostic tests for detection of respiratory pathogens (eg, polymerase chain reaction [PCR]) [5] and biomarkers with potential ability to distinguish between viral and bacterial infections, such as procalcitonin

Received 7 August 2014; accepted 26 November 2014; electronically published 29 January 2015.

Presented in part: IDWeek Annual Meeting, San Francisco, California, 2–6 October 2013. Poster 1317.

Correspondence: Octavio Ramilo, MD, Division of Pediatric Infectious Diseases and Center for Vaccines and Immunity, The Research Institute at Nationwide Children's Hospital, 700 Children's Dr, WA 4021, Columbus, OH 43205 (octavio.ramilo@nationwidechildrens.org).

The Journal of Infectious Diseases® 2015;212:213–22

DOI: 10.1093/infdis/jiv047

(PCT) [6] and C-reactive protein [7]. However, these tools have limitations, especially for the diagnosis of bacterial pathogens. Thus, there is an urgent need for the development of new methods to accurately differentiate viral from bacterial LRTIs to allow targeted use of antimicrobial therapy.

Transcriptional analysis is a promising tool that allows a comprehensive and efficient analysis of a large number of genes and helps elucidate the host response to the disease [8]. This approach has been applied to patients with different infectious diseases and has successfully demonstrated the potential to differentiate bacterial and viral infections, mainly in the pediatric population [9–11]. Moreover, recent studies have also demonstrated the value of this approach as a prognostic tool [10, 12]. Fewer studies have evaluated the role of microarrays for the diagnosis of bacterial and viral LRTI in adults [12–14].

The purpose of this study was to explore the value of transcriptional profiling as a potential diagnostic tool in adults hospitalized with LRTIs. Specifically, our focus was to evaluate the application of transcriptional profiles in the differential diagnosis of bacterial and viral LRTIs in adults.

PATIENTS, MATERIALS, AND METHODS

Study Design and Patients

We conducted a prospective study in adult patients (aged ≥21 years) hospitalized with LRTI over 3 respiratory seasons (2008–2011). All patients were enrolled at Rochester General Hospital, New York, and screened for bacteria and respiratory viruses with a comprehensive microbiologic diagnostic algorithm, as described elsewhere [15]. Exclusion criteria included antibiotic treatment before admission, immunosuppression, cavitary lung disease, and witnessed aspiration. Informed consent was obtained from subjects before sample collection. At enrollment, demographic, clinical, and laboratory information was collected, and a pulmonary specialist assigned the admitting diagnosis after examination of each patient and review of laboratory and radiographic data [15].

Blood samples were obtained for microarray analyses and white blood cell counts with differential, and PCT concentrations were measured in all study patients using a time-resolved amplified cryptate emission technology assay (Kryptor PCT; Brahms). The functional sensitivity of the assay is 0.06 ng/mL [16]. Blood samples were also obtained from 40 healthy volunteers matched for age, race, and sex, and screened for respiratory illnesses, as reference controls (Supplementary Figure 2). The study was approved by the institutional review boards at Rochester General Hospital, the University of Rochester, and Nationwide Children's Hospital.

Specimen Collection and Microbiologic Methods

Nasopharyngeal, sputum, urine, and blood samples were obtained at admission for bacterial and viral detection and were processed at Rochester General Hospital clinical laboratory, as described elsewhere [15]. Briefly, respiratory viruses were identified by a combination of PCR, culture, and serology, and bacterial pathogens were identified by culture, serology, and antigen detection. To identify respiratory viruses, we used PCR for influenza A and B viruses, respiratory syncytial virus [RSV], human metapneumovirus [HMPV], parainfluenza virus [PIV] types 1–3, human rhinovirus [HRV], and coronavirus 229E and OC43; viral culture for influenza, RSV, PIV, HRV, adenovirus, and enterovirus; and acute and convalescent serum immunoglobulin G titers for influenza, RSV, HMPV, PIV, HRV, and coronavirus 229E and OC43. Bacterial pathogens were identified with Gram staining and semiquantitative bacterial cultures from sputum and blood samples as well as nasopharyngeal and throat PCR for *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae*. In addition, urinary antigen testing was also performed for *Streptococcus pneumoniae* and *Legionella pneumoniae*. Finally, pneumococcal surface protein A antigens covering families 1 and 2 were used for pneumococcal serology.

Microarray Data and Statistical Analysis

Blood samples (1–3 mL) for microarray analyses were collected in Tempus tubes (Applied Biosystems) and stored at −20°C. Whole blood RNA was processed and hybridized into Illumina Human HT-12 v4 BeadChip kit (47 323 probes) and scanned on the Illumina BeadStation 500 [12, 17]. Illumina GenomeStudio software was used to subtract background and scale average samples' signal intensity, and GeneSpring GX 7.3 (Agilent Technologies) software to perform further normalization and analyses [9, 12, 18]. Briefly, transcripts were first selected if they were present in ≥10% of all samples and had a minimum of 2-fold expression change compared with the median intensity across all samples. Using this approach, we obtained a total of 16 710 quality control transcripts (Supplementary Figure 1*A*; 158 samples [22 bacterial LRTI, 71 viral LRTI, 25 bacterial-viral coinfections, and 40 healthy controls]).

We then continued with the following strategy. First, we performed *supervised analysis* (comparative analyses between predefined sample groups) using the Mann–Whitney test ($P < .01$), followed by Benjamini–Hochberg multiple-test corrections and a ≥1.25 fold change in expression level relative to the control group [10, 12]. Next, we applied *unsupervised clustering* (unbiased grouping of samples based on their transcriptional profile without prior knowledge of sample classification) to the validation set. We then applied *class prediction,* using the K–nearest neighbors (K-NN) algorithm, with 12 neighbors and a *P* value ratio cutoff of .5, to identify the top-ranked genes that best discriminated between bacterial and viral infections [9]. Finally, we performed *functional gene analyses* using a modular analysis, as described elsewhere [17, 19, 20] (module transcript content and annotations are available online at http://www.biir.net/public_wikis/module_annotation/V2_Trial_8_Modules). The data are

deposited in the National Center for Biotechnology Information Gene Expression Omnibus (accession No. GSE6024).

Patient demographic and clinical characteristics were compared using χ^2 or Fisher exact tests, whenever appropriate. Normally distributed continuous variables were compared using *t* tests or 1-way analysis of variance, and results were expressed as means and standard deviations. Nonnormally distributed continuous variables were compared using Mann–Whitney or Kruskal–Wallis tests (for 2 or >2 groups, respectively), and results were expressed as medians and interquartile ranges. Differences were considered significant at $P < .05$ for all statistical analyses. The IBM SPSS software package, version 19.0 (IBM), and GraphPad Prism version 6.03 for Windows (GraphPad Software), were used to perform statistical analyses.

RESULTS

Patient Demographic Characteristics and Etiologic Diagnosis

During the study period, 118 patients and 40 healthy controls (matched for age, sex, and race) were enrolled. Patients' median age was 61 years (interquartile range, 50–76 years), 69 (58.4%) were female, and the majority were white (76.3%). The most common clinical presentation was chronic obstructive pulmonary disease exacerbation (34 episodes; 28.8%), followed by community-acquired pneumonia (32 episodes; 27.1%). The most common presenting symptoms were cough (97.4%) and dyspnea (94%). None of the patients enrolled died during the study period. The remaining clinical characteristics of the patients with LRTI and the control group are summarized in Table 1 and Supplementary Table 1.

Of the 118 patients hospitalized with LRTI, a respiratory virus infection was diagnosed in 71 (60.2%) patients, a bacterial pathogen in 22 (18.6%), and a bacterial-viral coinfection in 25 (21.2%). Of the 71 viral infections, 32 (45%) were caused by influenza A, 9 (12.7%) by influenza B, 17 (23.9%) by RSV, and 7 (9.9%) by HMPV, and 6 (8.4%) were viral-viral coinfections. Among the bacterial infections, we identified 13 *Streptococcus pneumoniae*, 2 *Staphylococcus aureus*, 4 *Moraxella catarrhalis,* and 3 bacterial-bacterial coinfections.

Robust Transcriptional Biosignature in Adults Hospitalized With LRTI

We obtained blood samples from the 118 patients (including bacterial, viral, and bacterial-viral coinfections) and 40 healthy controls to define the whole blood biosignature of LRTI in adults (Supplementary Table 2). Samples were randomly divided into 2 independent cohorts ("training" and "test" sets). We used the training set to identify the transcriptional signature of LRTI and then validated it in the test set. Statistical group comparisons between the training set of 59 patients with LRTI and 20 healthy controls, matched for age, sex, and race, yielded 3986 differentially expressed transcripts (Figure 1*A*). This signature was validated in the independent test set of 59 patients and 20 healthy controls, also matched for age, sex and race. Hierarchical clustering of the test set samples confirmed the consistency of the gene expression patterns and correctly grouped 53 (90%) of 59 patients with LRTI (Figure 1*B*).

To better understand the host response in adults with LRTI and the immune pathways activated or suppressed, we used an analytical framework of 62 transcriptional modules that group together genes with shared expression patterns and similar biologic functions [19]. Module maps were derived independently for the training (Figure 1*C*) and test sets (Figure 1*D*), using their respective healthy control groups as references. Overall, patients with LRTI demonstrated significant overexpression of modules linked to the innate immune response (inflammation [M4.6, M5.1, and M5.7], interferon [IFN; M1.2, M3.4 and M5.12], and neutrophils [M5.15]) and underexpression of modules linked to the adaptive immune response (T cells [M4.1 and M4.15], B cells [M4.10], and lymphoid lineage [M6.19]). These findings were validated in the test set, as demonstrated by a significant correlation between training and test sets (Figure 1*F*; $P < .001$; Spearman $r = 0.98$), confirming the robustness of these observations.

Distinct Transcriptional Profiles in Patients With Bacterial, Viral, and Bacterial-Viral LRTIs

Next, to define the specific transcriptional profiles induced by viral or bacterial pathogens, we analyzed separately the gene expression profiles from 22 patients with bacterial infections, 71 with viral infections, and 25 with bacterial-viral coinfections, using 18 age-, sex-, and race-matched healthy controls as reference. Statistical group comparisons between the bacterial LRTI group and healthy controls identified 3376 differentially expressed transcripts. A similar approach revealed 2391 transcripts differentially expressed between viral LRTI and controls and 2628 between patients with bacterial-viral coinfections and controls. A hierarchical clustering algorithm was applied to the 3 patient cohorts to visualize the transcriptional pattern (Figure 2*A*–*C*). There were 1222 transcripts shared among the 3 groups. They represented 36% of transcripts for the bacterial LRTI signature, 51% of the viral LRTI signature, and 47% of the bacterial-viral coinfection signature. Likewise, a significant number of transcripts were specific for each of the groups (Figure 2*D*).

To better characterize the biologic significance of the differences in gene expression profiles identified in the 3 groups, we performed modular analysis [19]. Patients with bacterial LRTI demonstrated significant overexpression of modules related to the innate immune response, including inflammation and neutrophils modules, and underexpression of genes regulating the adaptive immunity, such as B- and T-cell modules, as did the viral infection group. However, the latter presented a marked overexpression of the IFN-related modules, which was absent in the bacterial infection group (Figure 3). Furthermore, the

Table 1. Demographic, Clinical, Radiologic, and Laboratory Data for Enrolled Patients With LRTI[a]

Patient Characteristics	Total (n = 118)	Viral LRTI (n = 71)	Bacterial LRTI (n = 22)	Bacterial-Viral LRTI (n = 25)	*P* Value	Missing Values, No. of Patients
Age, median (IQR), y	62 (50–76)	61 (50–77)	67 (46–76)	60 (51–72)	.95	0
Male sex	49 (41.5)	27 (38.0)	10 (45.5)	12 (48.0)	.63	0
Race						
White	90 (76.3)	59 (83.1)	14 (63.6)	17 (68.0)	.19	0
Black	26 (22.0)	11 (15.5)	8 (36.4)	7 (28.0)		0
Asian	2 (1.7)	1 (1.4)	0 (0.0)	1 (4.0)		0
Residence						3
Home	107 (93.0)	63 (92.6)	21 (95.5)	23	.54	. . .
Assisted living	4 (3.5)	3 (4.4)	1 (4.5)	0 (0.0)		. . .
Nursing home	4 (3.5)	2 (2.9)	0 (0.0)	2 (8.0)		. . .
Underlying conditions						
Chronic sinus disease	13 (11.4)	7 (10.4)	4 (18.2)	2 (8.0)	.51	4
Diabetes mellitus	43 (36.4)	27 (38)	8 (36.4)	8 (32)	.86	0
CHF	20 (16.9)	13 (18.3)	3 (13.6)	4 (16.0)	.87	0
COPD	47 (39.8)	25 (35.2)	11 (50.0)	11 (44.0)	.41	0
CRF	1 (0.8)	0 (0.0)	0 (0.0)	1 (4.0)	.16	1
BMI, median (IQR)	27.8 (22.8–36.5)	28.1 (22.8–37.4)	27.8 (22.5–37.5)	27.3 (23.3–34.3)	.89	27
Underweight	4 (4.4)	1 (1.9)	2 (10.5)	1 (4.8)	.84	NA
Healthy weight	31 (34.1)	18 (35.3)	5 (26.3)	7 (38.1)		NA
Overweight	18 (19.8)	10 (19.6)	4 (21.8)	4 (19.0)		NA
Obese	38 (41.7)	22 (43.1)	8 (42.1)	8 (38.1)		NA
Risk factors						
Smoking (active)	45 (38.1)	23 (32.4)	9 (40.9)	13 (52.0)	.21	0
Oral steroids	14 (11.9)	9 (12.9)	2 (9.1)	3 (12.0)	.89	1
Inhaled steroids	50 (42.4)	21 (29.6)	15 (68.2)	14 (56.0)	.002[b]	0
NSAIDs	22 (18.6)	14 (19.7)	4 (18.2)	4 (16.0)	.92	0
Home oxygen	24 (20.3)	14 (19.7)	8 (36.4)	2 (8.0)	.054	0
Influenza vaccine	72 (63.7)	43 (64.2)	17 (81.0)	12 (48)	.07	5
Pneumococcal vaccine	57 (58.2)	33 (60.0)	14 (70.0)	10 (43.5)	.20	3
Clinical diagnosis						
Pneumonia	32 (27.1)	12 (16.9)	11 (50.0)	9 (36.0)	.07	0
COPD exacerbation	34 (28.8)	22 (31.0)	6 (27.3)	6 (24.0)		0
Asthma	20 (16.9)	13 (18.3)	2 (9.1)	5 (20.0)		0
CHF	8 (6.8)	6 (8.5)	2 (9.1)	0 (0.0)		0
Bronchitis	15 (12.7)	13 (18.3)	0 (0.0)	2 (8.0)		0
Other[c]	9 (7.6)	5 (7.0)	1 (4.5)	3 (12.0)		0
Clinical manifestations						
Congestion	68 (57.6)	47 (66.2)	8 (36.4)	13 (52.0)	.03[b]	0
Cough	115 (97.5)	71 (100.0)	20 (90.9)	24 (96.0)	.05	0
Purulent sputum	75 (66.4)	42 (63.6)	14 (63.6)	19 (76.0)	.51	5
Dyspnea	111 (94.1)	66 (93.0)	22 (100.0)	23 (92.0)	.42	0
Wheezing	75 (63.5)	49 (69.0)	12 (54.5)	14 (56.0)	.32	0
Confusion	14 (12.3)	8 (11.9)	3 (13.6)	3 (12.0)	.98	4
Chest radiographic findings						
Normal	50 (42.7)	32 (45.7)	6 (27.3)	12 (48.0)	.23	1
Atelectasis	33 (28.2)	22 (31.4)	6 (27.3)	5 (20.0)	.55	1
Infiltrate, no consolidation	49 (41.8)	27 (38.6)	10 (45.5)	12 (48.0)	.66	1
Consolidation	6 (5.1)	2 (2.9)	3 (13.6)	1 (4.0)	.13	1
Edema	18 (15.3)	10 (14.3)	5 (22.7)	3 (12.0)	.55	1
Pleural effusion	7 (6.0)	2 (2.9)	4 (18.2)	1 (4.0)	.03[b]	1

Table 1 continued.

Patient Characteristics	Total (n = 118)	Viral LRTI (n = 71)	Bacterial LRTI (n = 22)	Bacterial-Viral LRTI (n = 25)	*P* Value	Missing Values, No. of Patients
CT scan pattern						
Normal	5 (27.7)	5 (50.0)	0 (0.0)	0 (0.0)	.06	100
Atelectasis	5 (27.7)	3 (30.0)	2 (33.3)	0 (0.0)	.64	100
Infiltrate, no consolidation	8 (44.4)	2 (20.0)	5 (83.3)	1 (50.0)	.047[b]	100
Consolidation	2 (11.1)	1 (10.0)	1 (16.7)	0 (0.0)	.80	100
Laboratory parameters						
WBC count, median (IQR), $\times 10^3$/mL	10.0 (7.15–14.7)	8.6 (6.6–11.4)	12.9 (9.2–17.5)	13.1 (8.8–17.5)	<.001[b]	5
Leukocytosis[d]	40 (60.6)	13 (19.7)	12 (54.5)	15 (60.0)	<.001[b]	5
Leukopenia[e]	2 (3.0)	1 (1.5)	1 (4.5)	0 (0.0)	.48	5
Neutrophils median (IQR), %	73.0 (66.0–83.0)	72.5 (65.2–81.2)	80.0 (68.2–86.2)	71.0 (65.0–82.0)	.07	15
Band neutrophils, median (IQR), %	1.0 (0.0–6.0)	0.0 (0.0–5.0)	0.0 (0.0–3.7)	6.0 (0.0–14.0)	.02[b]	23
Platelet count, median (IQR), $\times 10^3$/mL	228.0 (180.5–287.5)	217.0 (164.0–253.0)	268.0 (190.5–326.2)	231.0 (206.5–302.0)	.08	5
Serum urea nitrogen, median (IQR), mg/dL	17.5 (11.0–24.0)	18.0 (11.7–24.2)	16.0 (11.0–22.5)	17.5 (10.2–24.0)	.98	6
PCT, median (IQR), ng/mL[f]	0.13 (0.08–0.34)	0.11 (0.08–0.19)	0.13 (0.05–1.17)	0.67 (0.16–2.89)	<.001[b]	5

Abbreviations: BMI, body mass index; CHF, chronic heart failure; COPD, chronic obstructive pulmonary disease; CRF, chronic renal failure; CT, computed tomographic; IQR, interquartile range; LRTI, lower respiratory tract infection; NA, not available; NSAIDS, nonsteroidal anti-inflammatory drugs; PCT, procalcitonin; WBC, white blood cell.

[a] All data represent No. (%) of patients, unless otherwise specified.

[b] Significant associations ($P \leq .05$).

[c] Other clinical diagnoses include influenza, acidosis, and viral syndrome, among others.

[d] Leukocytosis was defined as a WBC count >12 000.

[e] Leukopenia was defined as a WBC count <4000.

[f] The highest PCT value between day 1 and day 2 measurements.

plasma cells module was only overexpressed in the viral infection group. For those modules that presented the same trend in the level of expression, we also found differences in magnitude, as was the case with neutrophils (more overexpressed in the bacterial infection group; $P < .001$) and natural killer cells (more underexpressed in the bacterial infection group; $P < .001$) modules (Figure 3 and Supplementary Figure 1*B*). The module based analysis also showed a distinct pattern for the bacterial-viral coinfection group, characterized by relative underexpression of the platelets and monocytes modules (Figure 3) and intermediate expression of inflammation and IFN-related modules compared with the bacterial and viral infection groups (Supplementary Figure 1*B*).

Discrimination Between Viral and Bacterial LRTIs With Classifier Genes

To explore the value of transcriptional profiles to discriminate between viral and bacterial LRTI, we applied the K-NN class prediction algorithm. To perform this analysis, we divided patients with bacterial and viral LRTI into 2 independent cohorts (K-NN training set and test set), each comprising 11 patients with bacterial and 12 with viral LRTI. The K-NN algorithm identified 10 classifier genes that best discriminated bacterial from viral LRTI (Table 2). Eight of those 10 classifiers were IFN-related genes (*IFI44, IFIT3, IFI27, RSAD2, OAS2, OASL, IFIT2,* and *PARP9*). Using the 10 classifier genes, leave-one-out cross-validation of the training set correctly classified 21 of 23 samples (91% accuracy; Figure 4*A*). In the validation analysis (K-NN in the test set), the classifier genes correctly categorized 22 of 23 new patient samples (96% accuracy; Figure 4*B*). To further confirm these findings, using the 10 classifier genes, we applied an unsupervised hierarchical clustering algorithm to a third cohort of patients comprising the 11 patients with bacterial LRTI used in the K-NN training set and a new independent set of 12 patients with viral LRTI. In this analysis, 22 of 23 patients (96%) were correctly classified (Figure 4*C*).

Sensitivity and Specificity of Transcriptional Profiles Compared With Serum PCT for Distinguishing Bacterial From Viral LRTI

Next, we assessed the ability of the 10 classifier genes signature and serum PCT to discriminate between bacterial and viral infections by comparing their sensitivity and specificity. The

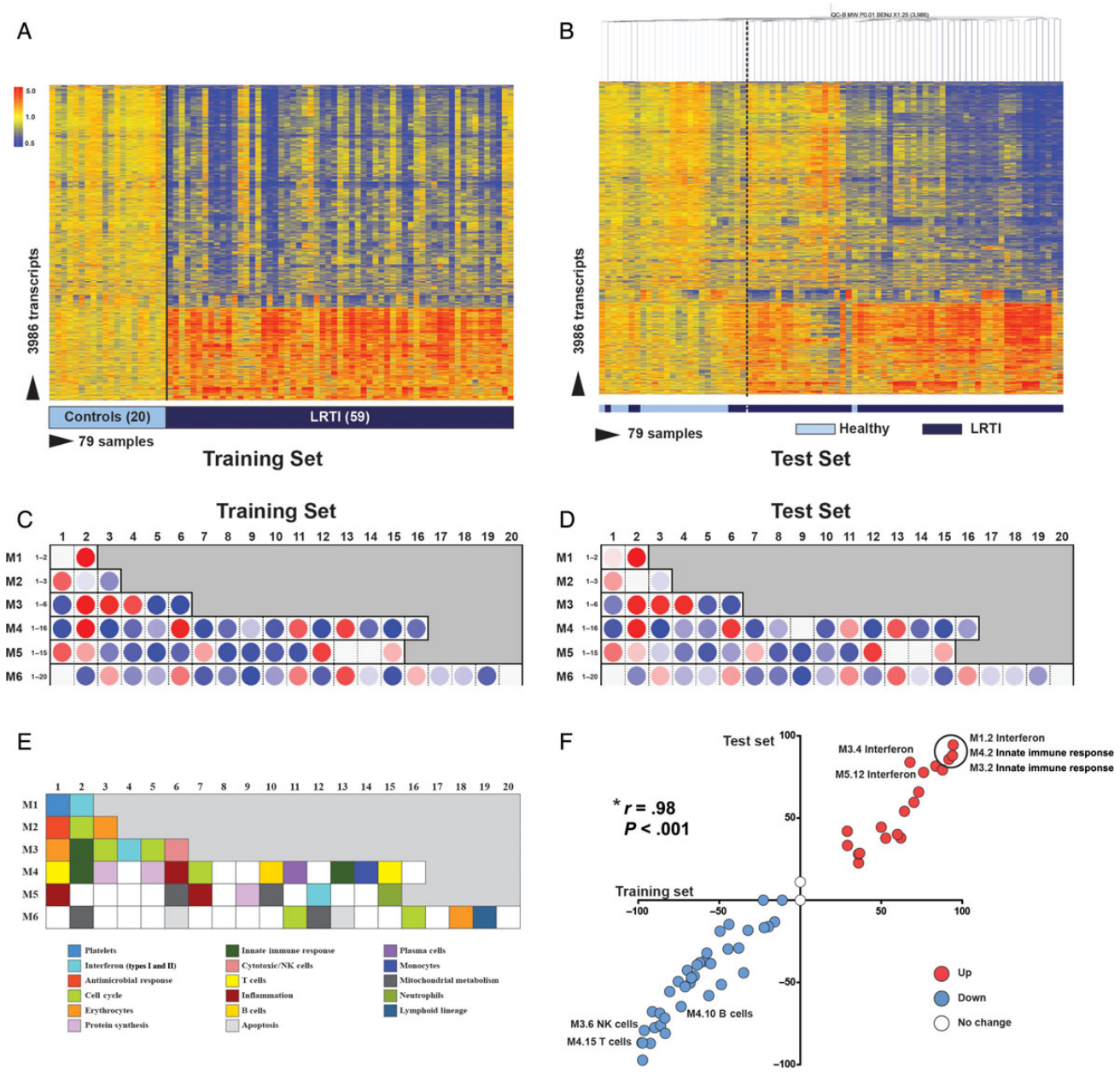

Figure 1. Lower respiratory tract infection (LRTI) whole blood transcriptional signature. *A*, Heat map representing the transcriptional profile of 20 healthy controls and 59 patients with LRTI based on 3986 transcripts obtained from a nonparametric test ($P < .01$), 1.25-fold change, and Benjamini–Hochberg multiple-test correction. Transcripts were organized by hierarchical clustering (standard correlation) according to similarities in expression profiles. Transcripts are represented in rows, and individual subjects in columns. Normalized log ratio levels are indicated in red (overexpressed) or blue (underexpressed), as compared with the median expression of the healthy controls. *B*, Unsupervised hierarchical clustering (distance method) of the transcriptional profiles from the same 3986 transcripts in an independent test cohort comprising 20 healthy controls and 59 patients with LRTI. *C*, Average modular transcriptional profile for patients with LRTI compared with healthy controls in the training set. *D*, Average modular transcriptional profile for patients with LRTI as compared with healthy controls in the test set. *E*, Module functional annotations legend. *F*, Scatterplot representing the module expression correlation (Spearman) between the training (x-axis) and test (y-axis) sets. Abbreviation: NK, natural killer.

combined sensitivity of the 10 classifier genes derived from the 3 cohorts of patients analyzed was 95% (95% confidence interval [CI], 77%–100%) with a specificity of 92% (77%–98%) (Table 3). Following Schuetz et al [21], a serum PCT concentration of ≥0.25 ng/mL, measured within the first 48 hours of hospitalization, is considered indicative of possible bacterial infection. This PCT cutoff value was applied to the same patient cohorts that were included in the K-NN analysis and demonstrated a sensitivity of 38% (95% CI, 18%–62%) and a specificity of 91% (76%–98%) (Table 3). Even after we included all the

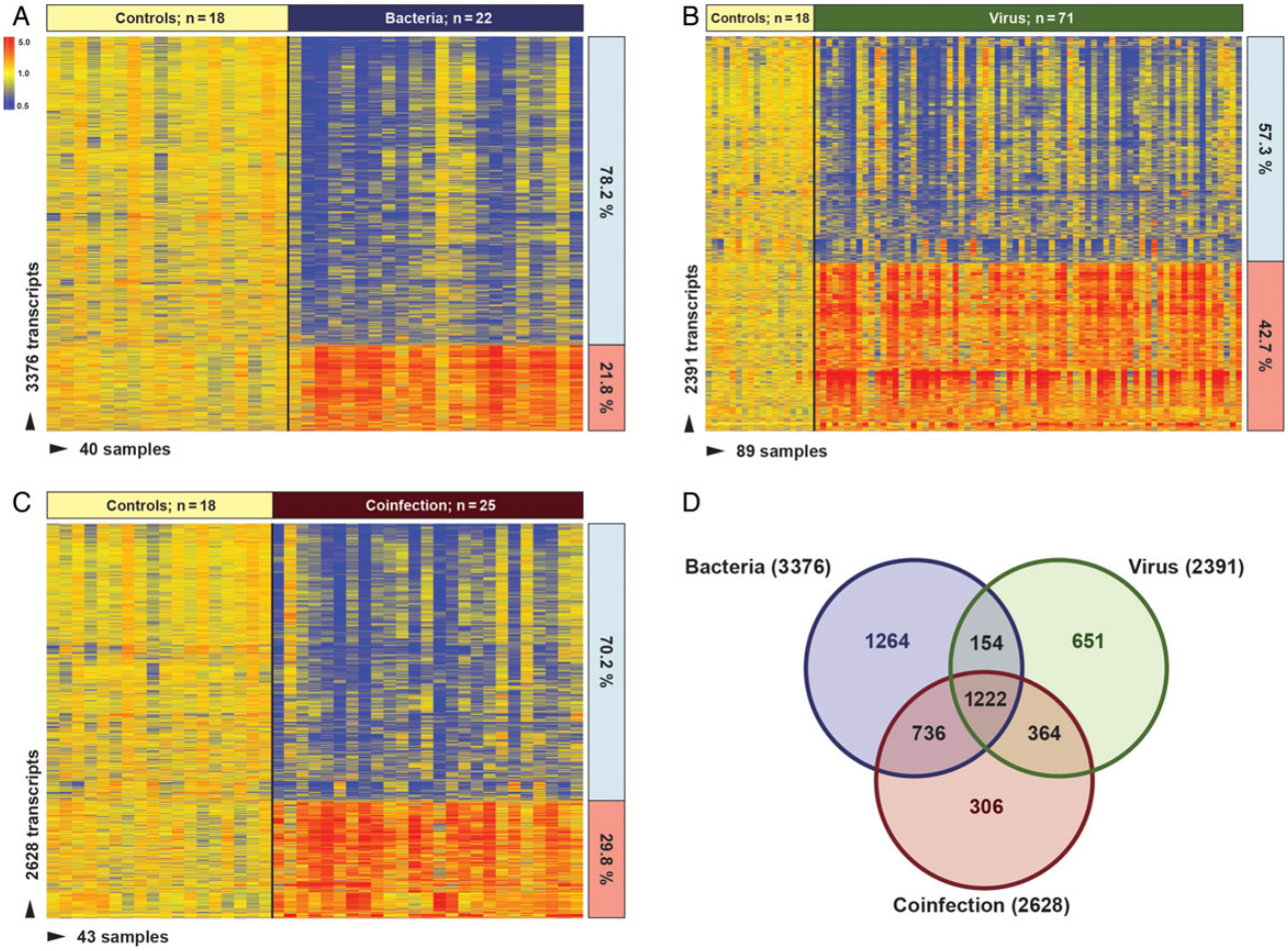

Figure 2. Transcriptional profiles in patients with bacterial, viral, and bacterial-viral (coinfection) lower respiratory tract infection (LRTI). Heat maps represent the transcriptional profiles of 18 healthy controls and 22 patients with a bacterial LRTI based on 3376 transcripts (*A*); 18 healthy controls and 71 patients with a viral LRTI based on 2391 transcripts (*B*), and 18 healthy controls and 25 patients with a bacterial-viral LRTI based on 2628 transcripts (*C*). All transcripts were identified after applying a nonparametric test (Mann–Whitney) ($P < .01$), 1.25-fold change, and Benjamini–Hochberg multiple-test correction. *D*, Venn diagram displaying the overlap among the significant transcripts identified in the 3 LRTI groups.

patients with bacterial and viral infection who had PCT concentrations measured (n = 88), the sensitivity and specificity did not improve (sensitivity, 36% [95% CI, 17%–59%]; specificity, 80% [69%–89%]).

DISCUSSION

This study was designed to evaluate the applicability of transcriptional profiling as a diagnostic tool in adults hospitalized with LRTI. Using an objective analytical approach, we found that the systemic response in patients with LRTI displays a robust and reproducible biosignature, with specific characteristics depending on the causative pathogen. Using these biosignatures and applying strict statistical analyses, we demonstrated that virus and bacterial infections can be differentiated with high sensitivity and specificity at the host response level.

LRTIs are a frequent cause of hospitalization and a major reason for treatment with antibiotics worldwide. The role of respiratory viruses in LRTIs in older adults is increasingly recognized, suggesting that accurate diagnostic tests to identify the etiologic mechanism of LRTIs have the potential to reduce overall antibiotic use. Our results showed that a distinct systemic host response is elicited in each infection group (ie, bacterial vs viral) and can be detected in the blood and at an early stage of the disease, as reported elsewhere for children [9, 10, 17, 22]. In agreement with other studies using gene expression analyses in patients with respiratory viral infections [10, 23], we observed that the major difference between the groups was found in the IFN response, with the viral infection group showing the highest expression level. Indeed, the 10 transcripts identified with the K-NN algorithm as the best classifiers included 8 IFN-related genes, allowing us to categorize patients with

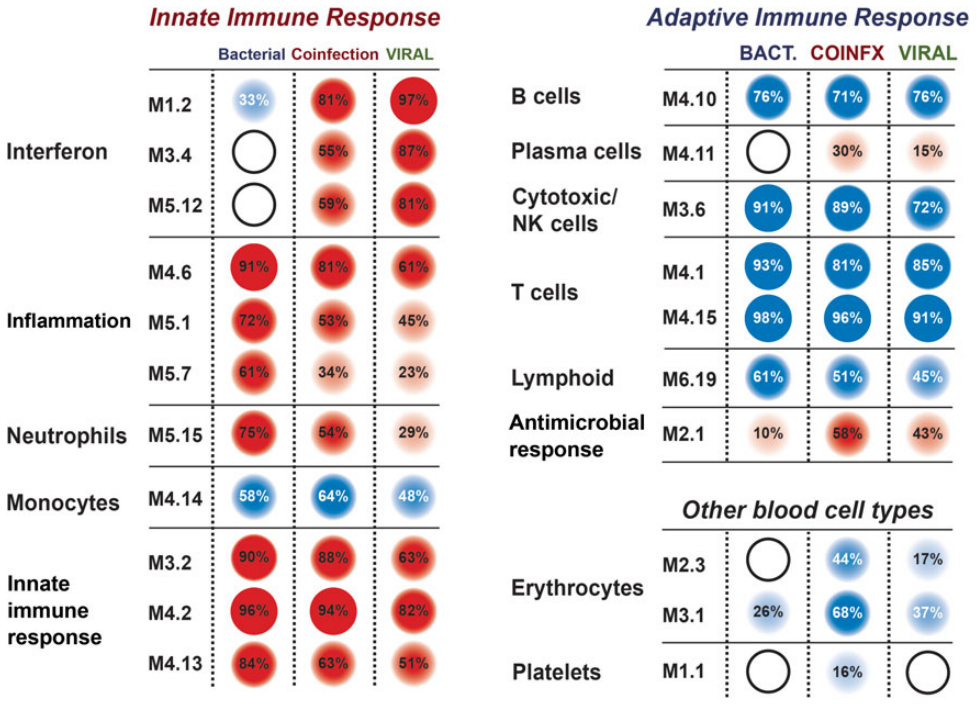

Figure 3. Modular transcriptional fingerprint comparison among the 3 lower respiratory tract infection groups. Mean modular transcriptional fingerprint for bacterial (22 patients and 18 matched controls), viral (25 patients and 18 matched controls), and bacterial-viral coinfection (25 patients and 18 matched controls). Modules are organized based on its relation to the innate and adaptive immune response. Abbreviation: NK, natural killer.

bacterial versus viral LRTI with high sensitivity (95%) and specificity (92%).

Different biomarkers to help differentiate bacterial from viral infections have been evaluated in adults with LRTI, with conflicting results [24]. C-reactive protein was considered a useful tool to distinguish bacterial from viral infections [7], but studies have shown [25] that this biomarker lacks sufficient sensitivity and specificity. In addition, although there are promising results supporting the applicability of PCT as a prognostic biomarker, its value as a diagnostic tool is still under debate. Its diagnostic accuracy in LRTI is variable and depends on the cutoff values used and the conditions under study, such as community-acquired pneumonia, acute exacerbations of chronic obstructive pulmonary disease, or bronchitis [26]. Our results indicate that the 10 classifier genes have superior sensitivity and similar specificity compared with serum PCT, demonstrating the potential value of transcriptional profiling as a diagnostic tool to differentiate viral from bacterial LRTI in the adult population. Importantly, these findings are in agreement with the Infectious Diseases Society of America/American Thoracic Society recommendations for improvement in the diagnosis of LRTI, because the symptoms of bacterial community-acquired pneumonia overlap with viral causes of pneumonia as well as exacerbations of asthma or chronic obstructive pulmonary disease [27].

Our study has limitations. The biosignatures were obtained from hospitalized patients who represent the most severe forms of LRTI. Nevertheless, it serves as the initial basis for future studies including patients with less severe infections who are managed as outpatients. Profiles were derived from peripheral whole blood and not from the respiratory mucosa, which may not necessarily reflect what is occurring in the primary

Table 2. Classifier Genes That Best Discriminate Bacterial From Viral LRTI[a]

Gene	Bacterial LRTI	Viral LRTI
BTN3A3	0.40	1.21
IFI27	2.16	57.49
RSAD2	0.73	14.24
KIAA1618	0.86	2.73
OAS2	0.85	3.51
IFIT3	0.64	4.35
IFI44	0.83	9.17
OASL	1.11	5.22
IFIT2	0.86	2.56
PARP9	1.35	2.62

Abbreviation: LRTI, lower respiratory tract infection.

[a] Numeric values represent median expression values per transcript per study group. Interferon-related transcripts are shown in bold.

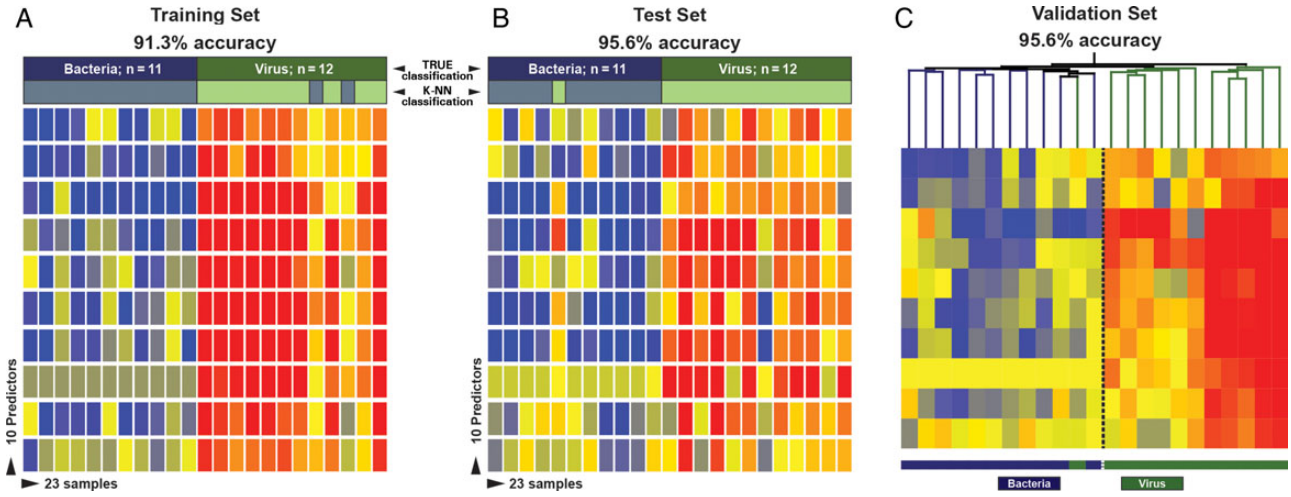

Figure 4. Transcriptional profile discrimination between bacterial and viral lower respiratory tract infection (LRTI). The 10 top-ranked genes that best differentiated bacterial from viral LRTI (Table 2) were identified, after use of a supervised learning K–nearest neighbors (K-NN) algorithm with 12 neighbors and a *P* value ratio cutoff of .5. *A*, Use of those genes in a training set correctly classified 21 of 23 individual subjects (91.3%). *B*, Cross-validation in a test set correctly classified 95.6% of patients. *C*, Validation in a third cohort, applying an unsupervised hierarchical clustering (distance method), correctly classified 95.6% of patients.

site of infection. However, studies have demonstrated the value of whole blood transcriptional profiles in other patient cohorts with respiratory infections [10, 12–14]. In addition, blood samples are convenient samples for diagnostic and prognostic analysis, because they are routinely obtained from patients with LRTI and do not require additional, more invasive interventions. Our results were obtained from a single cohort of 118 patients with LRTI hospitalized in a single center. It would be necessary to validate our findings in independent cohorts. Although we attempted to identify a biosignature in those patients with documented bacterial-viral coinfections, the small number of patients and the heterogeneity of the viruses and bacteria identified did not allow us yet to identify transcripts that can separate them from the other 2 groups. This important subgroup will require further studies using a larger sample.

In conclusion, this study provides evidence of the applicability of transcriptional profiling obtained from a single blood sample for diagnosis of viral versus bacterial LRTIs in adults within 24 hours of hospitalization. Indeed, there are ongoing studies applying PCR-based platforms to advance the implementation of host transcriptional profiles in the clinical setting [13]. In addition, our study demonstrates that with appropriate bioinformatics resources, vast amount of expression data can be simplified to obtain practical biomarkers as currently demanded by clinicians.

Table 3. Comparative Sensitivity and Specificity of PCT and Classifier Genes to Discriminate Between Viral and Bacterial LRTI

Method	Correct, No. (%)	Incorrect, No. (%)	Sensitivity (95% CI), %	Specificity (95% CI), %
PCT (n = 55)				
Bacterial	8 (38.1)	13 (61.9)	38 (18–62)	91 (76–98)
Viral	31 (91.2)	3 (8.8)		
Total	39 (70.9)	16 (29.1%)		
Classifier genes (n = 58)				
Bacterial	21 (95.5)	1 (4.5)	95 (77–100)	92 (77–98)
Viral	33 (91.7)	3 (8.3)		
Total	54 (93.1)	4 (6.9)		

Abbreviations: CI, confidence interval; LRTI, lower respiratory tract infection; PCT, procalcitonin.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank the staff of the Rochester General Hospital clinical microbiology, hematology, and chemistry laboratory for assistance with specimen retrieval, members of the respiratory therapy department for help with sputum induction, and Kenneth Becker and Eric Nylen for PCT measurements. We also thank Phuong Nguyen, Esperanza Anguiano, and Nicole Baldwin at Baylor Institute for Immunology Research for microarray sample processing and data curation.

Financial support. This work was supported by the National Institutes of Health (grants AI089987 and AI079446) and Nationwide Children's Hospital (intramural funds).

Potential conflicts of interest. All authors: No reported conflicts.

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. Macfarlane J, Holmes W, Gard P, et al. Prospective study of the incidence, aetiology and outcome of adult lower respiratory tract illness in the community. Thorax **2001**; 56:109–14.
2. Mizgerd JP. Lung infection—a public health priority. PLoS Med **2006**; 3:e76.
3. Creer DD, Dilworth JP, Gillespie SH, et al. Aetiological role of viral and bacterial infections in acute adult lower respiratory tract infection (LRTI) in primary care. Thorax **2006**; 61:75–9.
4. Hall CB, Weinberg GA, Iwane MK, et al. The burden of respiratory syncytial virus infection in young children. N Engl J Med **2009**; 360: 588–98.
5. Caliendo AM. Multiplex PCR and emerging technologies for the detection of respiratory pathogens. Clin Infect Dis **2011**; 52(suppl 4): S326–30.
6. Becker KL, Snider R, Nylen ES. Procalcitonin assay in systemic inflammation, infection, and sepsis: clinical utility and limitations. Crit Care Med **2008**; 36:941–52.
7. Husebekk A, Hansson LO. C-reactive protein in clinical practice. Oslo: Nycomed Pharma, Diagnostics Division, **1999**.
8. Chaussabel D, Pascual V, Banchereau J. Assessing the human immune system through blood transcriptomics. BMC Biol **2010**; 8:84.
9. Ramilo O, Allman W, Chung W, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood **2007**; 109:2066–77.
10. 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.
11. Hu XR, Yu JS, Crosby SD, Storch GA. Gene expression profiles in febrile children with defined viral and bacterial infection. Proc Natl Acad Sci U S A **2013**; 110:12792–7.
12. Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature **2010**; 466:973–7.
13. Zaas AK, Burke T, Chen MH, et al. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. Sci Transl Med **2013**; 5:203ra126.
14. Kaforou M, Wright VJ, Oni T, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. PLoS Med **2013**; 10:e1001538.
15. Falsey AR, Becker KL, Swinburne AJ, et al. Bacterial complications of respiratory tract viral illness: a comprehensive evaluation. J Infect Dis **2013**; 208:432–41.
16. Becker KL, Snider R, Nylen ES. Procalcitonin in sepsis and systemic inflammation: a harmful biomarker and a therapeutic target. Br J Pharmacol **2010**; 159:253–64.
17. Banchereau R, Jordan-Villegas A, Ardura M, et al. Host immune transcriptional profiles reflect the variability in clinical disease manifestations in patients with *Staphylococcus aureus* infections. PLoS One **2012**; 7:e34390.
18. Allantaz F, Chaussabel D, Stichweh D, et al. Blood leukocyte microarrays to diagnose systemic onset juvenile idiopathic arthritis and follow the response to IL-1 blockade. J Exp Med **2007**; 204:2131–44.
19. Chaussabel D, Quinn C, Shen J, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity **2008**; 29:150–64.
20. Chaussabel D, Baldwin N. Democratizing systems immunology with modular transcriptional repertoire analyses. Nat Rev Immunol **2014**; 14:271–80.
21. Schuetz P, Christ-Crain M, Thomann R, et al. Effect of procalcitonin-based guidelines vs standard guidelines on antibiotic use in lower respiratory tract infections the ProHOSP randomized controlled trial. JAMA **2009**; 302:1059–66.
22. Ardura MI, Banchereau R, Mejias A, et al. Enhanced monocyte response and decreased central memory T cells in children with invasive *Staphylococcus aureus* infections. PLoS One **2009**; 4:e5446.
23. Smith SB, Dampier W, Tozeren A, Brown JR, Magid-Slav M. Identification of common biological pathways and drug targets across multiple respiratory viruses based on human host gene expression analysis. PLoS One **2012**; 7:e33174.
24. Blasi F, Stolz D, Piffer F. Biomarkers in lower respiratory tract infections. Pulm Pharmacol Ther **2010**; 23:501–7.
25. van der Meer V, Neven AK, van den Broek PJ, Assendelft WJ. Diagnostic value of C reactive protein in infections of the lower respiratory tract: systematic review. BMJ **2005**; 331:26.
26. Seligman R, Ramos-Lima LF, Oliveira Vdo A, Sanvicente C, Pacheco EF, Dalla Rosa K. Biomarkers in community-acquired pneumonia: a state-of-the-art review. Clinics (Sao Paulo) **2012**; 67:1321–5.
27. Caliendo AM, Gilbert DN, Ginocchio CC, et al. Better tests, better care: improved diagnostics for infectious diseases. Clin Infect Dis **2013**; 57(suppl 3):S139–70.