communications medicine

A Nature Portfolio journal

Article

https://doi.org/10.1038/s43856-025-01370-z

Host gene expression analysis in the detection of bacterial and viral etiology in children hospitalized with a suspected severe infection

Ruut Piri[1] ✉, Milla Valta[2], Johanna Lempainen[1,2,3], Lauri Ivaska[1,4], Sini Ezer[5,6], Juha Kere[5,6,7], Shintaro Katayama[5,6,7,8] & Ville Peltola[1,4,8]

Abstract

Background Host gene expression profiling holds great potential in improving the differential diagnostics of bacterial and viral infections. We investigated its discriminative value in children with suspected serious infections.

Methods Peripheral blood gene expression profiles were analyzed by RNA sequencing in 268 children aged between 4 weeks and 16 years: 211 hospitalized due to a suspected severe infection, 15 with a confirmed viral respiratory tract infection managed as outpatients, and 42 healthy control children. We classified children according to the bacterial, viral, or other etiology of their final diagnosis, and determined expression profiles. We derived 2-transcript signatures discriminating bacterial infections and viral-bacterial co-infections from viral infections in a discovery group (n = 101) of children with respiratory tract infections and validated them in children (n = 109) with non-respiratory infections, including cases with probable and mixed etiologies in the analyses.

Results Here, we show that clustering of blood transcriptome cannot be unequivocally explained by the etiology of infection. A 2-transcript signature comprising *TSPO* and *SECISBP2* genes differentiates bacterial and viral-bacterial co-infections from viral infections with an area under the curve of 0.93 and 0.87 (95% confidence interval, 0.88–0.98 and 0.82–0.92) in the discovery and discovery plus validation groups, respectively. Both groups combined, the sensitivity is 77%, and specificity 87%.

Conclusions The identified 2-transcript signature demonstrates good accuracy in distinguishing between bacterial and viral infections in a complex population of children hospitalized for suspected severe infection. Heterogeneity of clinical manifestations needs to be considered in diagnostic gene expression studies.

Plain language summary

Gene expression analysis is a method used to study which genes are active in blood at a given time. It is a promising tool for distinguishing bacterial infections, which require antibiotic treatment, from self-resolving viral infections, as different genes are turned on, or "expressed", during these infections. We examined whether bacterial and viral infections can be differentiated through blood gene expression analysis in children hospitalized for suspected severe infection with various diagnoses. We identified a combination of two activated genes, *TSPO* and *SECISBP2*, that had good accuracy in distinguishing between bacterial and viral infections. The findings suggest that gene expression analysis could improve the diagnostics of infectious diseases in children and help reduce the overuse of antibiotics in the future.

Differential diagnosis of bacterial and viral infections in children is crucial in reducing the misuse of antibiotics and related harms, including increase in antimicrobial resistance and short- and long-term adverse effects. However, the tools available for this differentiation have deficiencies. Diagnostic approaches based on direct detection of microbes face limitations, including the potential inaccessibility of the infection site, time required for culture methods, and often questionable pathogenic role of a detected microbe[1–5]. Notably, young children are frequently respiratory virus-positive, on average every other week, regardless of symptoms[3]. Although rapid multiplex polymerase chain reaction (PCR) assays have become available, they do not distinguish between asymptomatic or co-incidental and true infections[1,2]. Despite widespread use, biomarkers for bacterial infection, such as C-reactive protein (CRP) and procalcitonin (PCT), lack the desired specificity to effectively guide antibiotic use[6,7]. While there are reports of promising novel biomarkers for viral infection, such as myxovirus resistance protein A (MxA)[8–11], none are so far implemented in routine clinical use.

A full list of affiliations appears at the end of the paper. ✉e-mail: ruut.piri@utu.fi

As an alternative approach, host gene expression profiling has shown great promise in improving the differential diagnostics of bacterial and viral infections, both in children and adults[12–24]. It has demonstrated excellent performance in discriminating between microbiologically confirmed severe bacterial and viral infections in children, with less pronounced results in groups with an indeterminate or probable etiology of infection[12,13]. An RNA signature comprising only two transcripts reliably categorized febrile children in a dichotomous manner[12], and a larger transcript combination was able to classify children into 18 categories of infectious or inflammatory diseases[22]. However, the application of gene expression profiling in complex and ambiguous cases, such as infections caused by a combination of microbes, with two or more foci of infection, or with atypical clinical presentation, is less explored in children. Ideally, markers of bacterial or viral infection should perform well in a patient population with a wide range of clinical illnesses and causative agents.

This study aims to investigate whether bacterial and viral infections can be differentiated through host gene expression analysis in a diverse population of children hospitalized for a suspected severe infection, with a wide age range and various clinical diagnoses. After initial analysis in the whole study population, we divide patients into groups with respiratory infections and other infections to manage heterogeneity in clinical manifestations. The performance of transcript signatures derived from a discovery group of children with respiratory infections is examined in a validation group of children with infections not affecting the respiratory tract. We identify a 2-transcript signature of *TSPO* and *SECISBP2* genes that discriminates between bacterial and viral infections with good accuracy.

Methods

Study design and conduct

This prospective two-center study was conducted at the pediatric emergency departments of Turku University Hospital and Seinäjoki Central Hospital, Finland, between December 2016 and April 2018. The inclusion criteria were: (1) age between 4 weeks and 16 years, (2) admission to hospital, and (3) blood bacterial culture drawn by the decision of the attending clinician. To ensure a balanced representation between bacterial and viral infections in the study population, we also included a convenience sample of children with a suspected or confirmed viral infection, either hospitalized or managed as outpatients. For children hospitalized due to a suspected serious viral infection between June 2017 and April 2018, the inclusion criteria were: (1) age between 4 weeks and 16 years, (2) admission to hospital for an acute infection, (3) no antibiotic treatment initiated on admission or used within one week before hospitalization, and (4) venous blood samples required for other reasons. Children with an RNA sample drawn were eligible for this study. We defined children hospitalized according to these criteria as having a suspected severe infection. Children with a confirmed rhinoviral upper respiratory tract infection managed as outpatients were part of an observational birth cohort study named Steps to the Healthy Development and Wellbeing of Children (STEPS study) conducted between November 2010 and March 2012 in Turku University Hospital. Asymptomatic virus-negative children participating in the STEPS study at their scheduled visits at the age of 2, 13, or 24 months served as healthy controls.

The study protocols were approved by the Ethics Committee of the Hospital District of Southwest Finland (approval numbers 48/1801/2016 and 16/180/2008). The parents of all children, and older children or adolescents themselves, provided their written informed consent at the enrollment.

Diagnostic measures

Of hospitalized children, blood samples were collected for blood bacterial culture, white blood cell (WBC) count, and determinations of plasma CRP and PCT levels, as well as for blood MxA level. WBC, CRP and PCT determinations were performed by routine methods in the hospital central laboratories. Whole blood samples for MxA measurement were diluted in 1:20 hypotonic buffer and stored at −70 °C until the enzyme immunoassay analysis was performed, as described earlier[25]. Nasopharyngeal swab samples were suspended into phosphate-buffered saline, and nucleic acids were extracted using NucliSENS easyMAG (bioMérieux). Multiplex RT-PCR Allplex respiratory panels 1–3 (Seegene) were used in Turku University Hospital, while FilmArray (BioFire Diagnostics) was employed at Seinäjoki Central Hospital for detecting respiratory viruses. Both methods detected adenovirus; influenza A and B viruses; parainfluenza viruses type 1, 2, 3, and 4; respiratory syncytial virus; human metapneumovirus; coronaviruses 229E, NL63, and OC43; rhinovirus; and enteroviruses. Additionally, Allplex detected human bocavirus, and FilmArray identified coronavirus HKU1. FilmArray results for rhinovirus/enterovirus were further analyzed with Allplex to specifically document rhinovirus or enterovirus. In the STEPS study, nucleic acids were extracted from nasal swabs using NucliSENS easyMAG (bioMérieux) or MagNA Pure 96 (Roche). Rhinovirus, enteroviruses, respiratory syncytial virus, and influenza A and B viruses were analyzed using PCR as described earlier[26].

Other microbiological samples were collected, and radiographic imaging performed in hospitalized children when deemed necessary by the attending physician.

Classification of children according to etiology

Initially, children were classified into eight etiologic groups (hereafter referred to as detailed etiologic classification; see Supplementary Methods for details): (1) definite bacterial infection, a clinical and microbiological diagnosis of bacterial infection in the absence of viral infection; (2) definite viral infection, a clinical and microbiological diagnosis of viral infection in the absence of bacterial infection; (3) probable bacterial infection, and; (4) probable viral infection, same criteria as for definite infections but without microbiological confirmation; (5) viral-bacterial co-infection, an infection with viral and bacterial etiology, or simultaneous viral and bacterial infections at distinct foci; (6) bacterial infection with an asymptomatic virus finding, a definite or probable bacterial infection and a respiratory virus finding without respiratory symptoms or an incidental finding of another virus not causing symptoms; (7) infection of undetermined etiology; and (8) noninfectious disease.

For subsequent analyses, the etiologic classification was simplified into two groups (hereafter referred to as the dichotomous etiologic classification). Children with a definite or probable bacterial infection, viral-bacterial co-infection or bacterial infection with an asymptomatic virus finding were combined into a bacterial infection group. Children with a definite or probable viral infection were combined into a viral infection group. Children with an infection of undetermined etiology or noninfectious disease were excluded from these analyses.

RNA isolation and quality control

Whole blood samples from case and control subjects were collected in Tempus Blood RNA tubes (Thermo Fisher Scientific) as instructed by the manufacturer. The samples collected between 2010 and 2012 were cryopreserved at −80 °C, and samples collected between 2016 and 2018 at −20 °C until RNA isolation. Total RNA was isolated using the Tempus Spin RNA Isolation Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. RNA quality was verified with the Agilent RNA 6000 Pico Kit (Agilent) on a 2100 Bioanalyzer (Agilent) instrument. An RNA integrity number of 7.5 or higher was considered adequate for RNA sequencing. The isolated RNA samples were stored frozen at −80 °C until RNA sequencing library preparation.

Preprocessing

The samples were arranged into seven libraries to enable bias correction[27], and processed by single-cell tagged reverse transcription (STRT) 2 with Globin-Lock[28]. Next, the sequences were processed with STRTprep pipeline[29] for the quality check and further analysis. The quality checks included assessing the number of spike-in reads, the ratio of mapped reads to spike-in reads, and the 5'-end capture rates of spike-ins and of protein-coding genes. In each library, outlier samples in any of the quality check parameters were considered disqualified. Following the exclusion of

disqualified samples, biases between the libraries were corrected, as described earlier[27].

Statistics and reproducibility

We estimated that a sample size of 200, with a prevalence of 15% for both definite bacterial and viral infections, would be sufficient for the planned analyses. During the study, we observed a lower rate of definite viral infections than anticipated, and therefore recruited children with a suspected viral infection to ensure the inclusion of at least 30 subjects in both these groups. Clinical characteristics of the study population was assessed using SPSS, version 29.0 (IBM).

Variable genes. The bias-corrected read counts were normalized using the spike-in normalization method[29,30]. Subsequently, variable genes were selected[29]. Genes that significantly fluctuated (adjusted fluctuation P values less than 0.05) compared to the technical variations modeled from the variation of spike-in RNA levels were considered as variable genes.

Overview of expression profile. The bias-corrected read counts were adjusted by adding an offset of one to avoid log(0), and subsequently normalized by spike-in normalization method[29,30]. The offset-normalized values were then subjected to a logarithmic transformation ($\log_2$) and scaled. Using the scaled, log-transformed, and normalized expression levels of the fluctuated genes, the optimal number of clusters in the samples was determined based on the total within sums of squares using the fviz_nbclust function in R's factoextra package. The samples were then clustered using R's k-means function. A heatmap with hierarchical clustering and sample annotation was illustrated using R's ComplexHeatmap package, with hierarchical sample clusters determined by R's k-means function. Principal component analysis (PCA) was performed using R's FactoMineR package, and uniform manifold approximation and projection (UMAP) was carried out using R's umap package.

Differential expression test. Differential expression was examined using SAMstrt[30] with 1000 permutation trials. The significance of the effect size was then compared to the technical variations observed in the spike-ins[29] to identify variable genes. Genes with adjusted $P < 0.05$ in both the differential expression and the effect size were considered significant.

Prediction. To predict bacterial infection based on the expression levels of two genes, logistic regression analysis was performed using the glm function in R. The accuracy of the predictions was evaluated using R's pROC package. The optimal sensitivity and specificity were estimated by Youden's method with 2000 bootstrap iterations through the ci.coords function, while the area under the curve (AUC) was calculated using DeLong's method through ci.auc function.

Quantitative real-time reverse transcription PCR (qRT-PCR)

To validate the results with qRT-PCR, RNA was reverse transcribed with SuperScript™ II reverse transcriptase (Invitrogen) according to the manufacturer's instructions. An ERCC RNA spike-in mix was included in each reaction as a control for normalization. Quantitative RT-PCR was performed using TaqMan® Gene Expression Assays (Applied Biosystems) using the following assay IDs: Hs00559362_m1 for *TSPO*, Hs00225345_m1 for *SECISBP2*, and Ac03459873_a1 for the ERCC control. qPCR reactions were run on Bio-Rad CFX384 thermocycler and analyzed with Bio-Rad CFX Manager 3.1 software. The accuracy of the predictions was confirmed as described above, but with using the $\log_2$-transformed normalized expression values.

Results

Characteristics and classification of study children

An RNA sample was obtained from 232 children hospitalized due to a suspected severe infection, and from 72 children participating in the STEPS study. Out of the 304 children, 36 were excluded due to insufficient sample volume or quality, resulting in a total of 268 children included in the analyses. Among these, 211 (78.7%) were hospitalized due to a suspected severe infection, 15 (5.6%) had a confirmed rhinoviral respiratory tract infection and were managed as outpatients in the STEPS study, and 42 (15.7%) were healthy children from the STEPS study (Fig. 1). The median ages of children with an acute infection and healthy children were 3.4 and 2.1 years (interquartile range [IQR], 1.6–7.9 and 1.2–2.1 years), respectively. The clinical characteristics, diagnoses, detected pathogens and biomarker levels of 226 children with an acute infection are shown in Supplementary data file 1. Three (1.3%) of them had an immunosuppressive condition, while 46 (20.4%) had another chronic condition without immunosuppression. Of the hospitalized children, 183 (86.7%) were febrile prior to admission, with a median duration of fever of 2 days. A respiratory virus was detected in a nasopharyngeal sample in 122 (61.0%) hospitalized children of the 200 studied, and in all 15 children managed as outpatients. Rhinovirus was the most frequent finding, detected in 77 (35.8%) children, followed by respiratory syncytial virus in 19 (8.8%), and human bocavirus in 19 (8.8%) children.

Out of 226 children with an acute infection, we determined 34 (15.0%) to have a definite bacterial infection, 29 (12.8%) a probable bacterial infection, 77 (34.1%) a viral-bacterial co-infection, 23 (10.2%) a bacterial infection and an asymptomatic virus finding, 41 (18.1%) a definite viral infection, 7 (3.1%) a probable viral infection, 12 (5.3%) an infection of undetermined etiology, and 3 (1.3%) a noninfectious disease (Fig. 1).

Clustering based on blood transcriptome

In our initial analysis, we compared the gene expression profiles of all study children based on the detailed etiologic classification. From the 268 samples, expression of a total of 15,789 protein-coding genes was detected, of which 6308 were found to be significantly variable. The peripheral blood transcriptome formed five clusters based on these variably expressed genes (Fig. 2a); however, the three main etiologic groups (definite bacterial infection, $n = 34$; definite viral infection, $n = 41$; and control children, $n = 42$) did not emerge as the primary determinants of the clusters. A PCA (Fig. 2b) and UMAP (Fig. 2c) suggested distinct expression patterns between healthy control children and children with a definite bacterial infection, but the expression profile of children with a definite viral infection was heterogeneous and not readily distinguishable from other groups. The clinical characteristics, diagnoses, and biomarker levels of children with a definite bacterial or viral infection are shown in Supplementary Table 1.

Characteristics of children in the discovery and validation groups

Next, we analyzed children based on the dichotomous (bacterial vs. viral) etiologic classification, dividing them into a discovery group of children with a respiratory tract infection and a validation group of children with a non-respiratory infection. In this way, we were able to decrease the heterogeneity of the clinical constellations in the discovery group and validate the performance of the best markers in children with a broad range of clinical presentations. The clinical characteristics, diagnoses, and biomarker levels are shown in Supplementary Table 2.

The discovery group consisted of 101 children (median age, 2.9 years, IQR, 1.7–6.3 years, 51.5% female) with a respiratory tract infection of either bacterial or viral etiology. Within this group, 72 (71.3%) were classified as having a bacterial infection, and 29 (28.7%) as a viral infection. Clinical diagnoses in the discovery group included pneumonia ($n = 66$) and upper or lower respiratory infection without pneumonia ($n = 35$). To assess the discriminatory potential of the identified classifier genes between bacterial and viral infections, a validation group was employed. This group comprised 109 children (median age, 5.0 years, IQR, 1.2–9.1 years, 50.5% female) with infections primarily not affecting the respiratory tract. Among them, 91 (83.5%) were classified as having a bacterial infection, and 18 (16.5%) as a viral infection. The most frequent clinical diagnoses in the validation group included pyelonephritis ($n = 36$), skin or soft tissue infection ($n = 28$), and suspected or microbiologically confirmed sepsis without focus ($n = 11$).

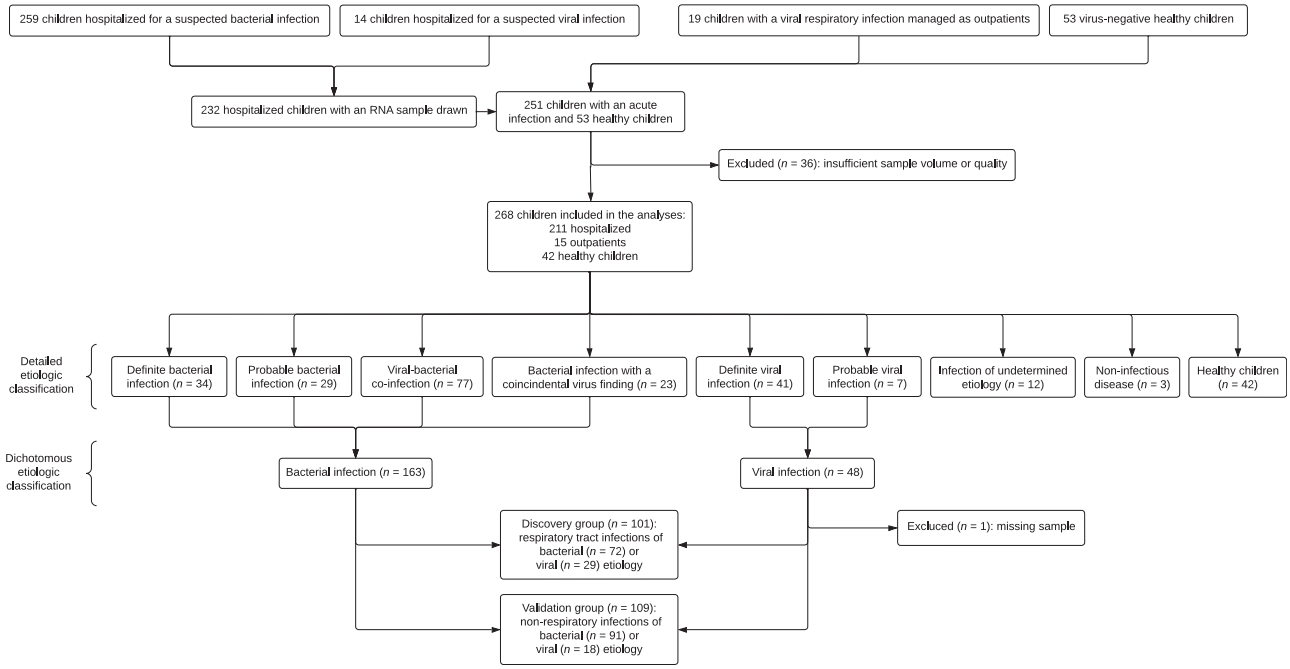

Fig. 1 | Participant flowchart. Of the 304 children with an RNA sample drawn, 36 were excluded due to insufficient sample volume or quality. A total of 268 children were included (211 hospitalized, 15 outpatients, and 42 healthy controls). Participants were classified into detailed etiologic categories, which were subsequently grouped into bacterial ($n = 163$) or viral ($n = 48$) infections for dichotomous classification. Children were further assigned to the discovery group ($n = 101$) or validation group ($n = 109$), with one excluded due to a missing RNA sample.

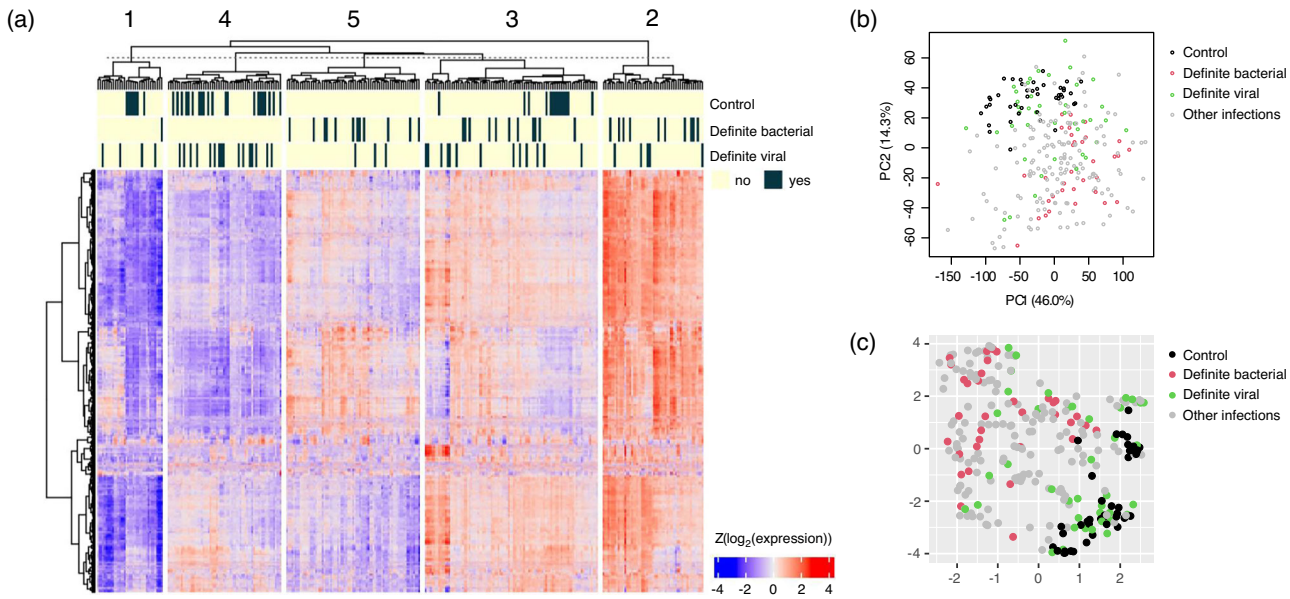

Fig. 2 | Overview of peripheral blood gene expression profiles in 42 healthy control children, 34 children with a definite bacterial infection, and 41 children with a definite viral infection. Overview of peripheral blood gene expression profiles using a heatmap with k-means clustering of the samples, principal component analysis (PCA), and uniform manifold approximation and projection (UMAP), shows distinct expression profiles between healthy control children ($n = 42$) and children with a definite bacterial infection ($n = 34$), but a heterogeneous expression pattern in children with a definite viral infection ($n = 41$). **a** Heatmap with k-means clustering of the samples represents patients as columns and RNA transcripts as rows. Expression intensity is indicated by color; RNA transcripts shown in red are upregulated and those in blue are downregulated. **b** PCA illustrates the distribution of samples based on their gene expression profiles, with circles representing individual samples. The first (PC1) and the second principal component (PC2) explain 46.0% and 14.3% of the total variance, respectively. **c** UMAP illustrates the distribution of samples, with the relative distances between points reflecting differences in gene expression patterns.

Differentiation between bacterial and viral infections by 2-transcript signatures

Sequencing of the transcripts in the discovery group ($n = 101$) identified 6147 genes, of which 2974 were differentially expressed between bacterial and viral respiratory infections. Among the differentially expressed genes, the best 2-transcript predictor for bacterial infection was the combination of *GAPDH* and *SHQ1* genes, achieving an AUC of 0.96 (95% confidence interval [CI], 0.93–0.99) (Supplementary data file 2; Figs. 3 and 4). This signature demonstrated a sensitivity of 88.9% and specificity of 100.0%. However, neither this combination nor four of the other top five signatures

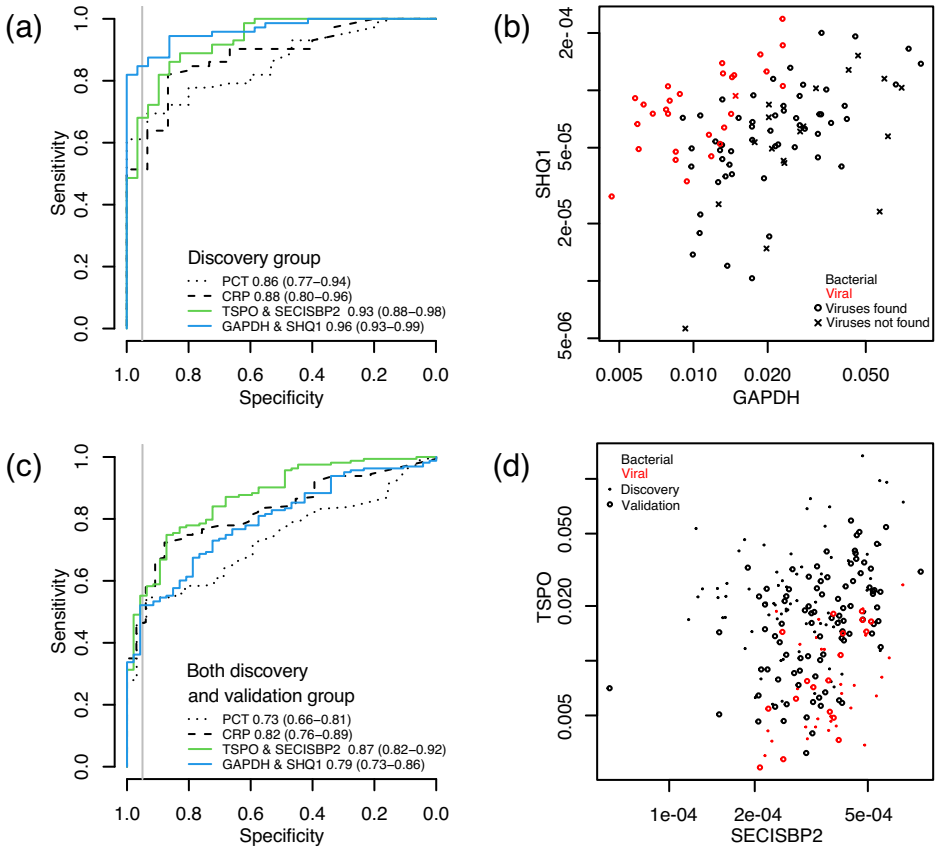

Fig. 3 | Receiver operating characteristic curves and scatter plots of normalized expression levels of the 2-transcript predictors for bacterial infection. In **a** receiver operating characteristic (ROC), curves are presented for the 2-transcript signatures *TSPO* and *SECISBP2*, as well as for *GAPDH* and *SHQ1*, along with bacterial biomarkers C-reactive protein (CRP) and procalcitonin (PCT) in the discovery group (respiratory tract infections; $n = 101$). Area under the curve (AUC) values and their corresponding 95% confidence intervals (CI) are provided. The gray vertical line indicates 95% specificity. In **b** a scatter plot illustrates the normalized expression levels of the 2-transcript signature *GAPDH* and SHQ1 in the discovery group ($n = 101$), where it was the best-performing combination. In **c** ROC curves are presented for the 2-transcript signatures *TSPO* and *SECISBP2*, as well as for *GAPDH* and *SHQ1*, and CRP and PCT in the combined discovery and validation groups ($n = 210$). AUC values and their corresponding 95% CI are provided. The gray vertical line indicates 95% specificity. In **d** a scatter plot illustrates the normalized expression levels of the 2-transcript signature *TSPO* and *SECISBP2* in the combined discovery and validation groups ($n = 210$), where it was the best-performing combination.

(all but one including the *SHQ1* gene) in the discovery group exhibited good discriminative value in the validation group, which consisted of children with different infection foci. In the combined discovery and validation groups, the AUC of *GAPDH* and *SHQ1* was 0.79 (95% CI, 0.73–0.86). In the validation group alone, the best-performing signature included the *C12orf75* and *TUSC2* genes, and exhibited an AUC of 0.88 (95% CI, 0.80–0.96) (Supplementary data file 2). However, like the most effective signatures in the discovery group alone, its discriminative value was lower when applied to both the discovery and validation groups together (AUC 0.81 [95% CI, 0.75–0.87]) (Supplementary data file 2).

Instead, a 2-transcript signature composed of *TSPO* and *SECISBP2* genes showed strong performance in the discovery group, validation group, and combined discovery plus validation groups, with AUCs of 0.93, 0.81 and 0.87 (95% CI, 0.88–0.98, 0.71–0.91, and 0.82–0.92), respectively (Supplementary data file 2; Figs. 3 and 4). The signature demonstrated a sensitivity and specificity of 87.5% and 89.7% in the discovery group, 71.4% and 88.9% in the validation group, and 76.7% and 87.2% in the combined groups. When prioritizing a higher sensitivity of 85.0%, aiming for a better rule-out value for a bacterial infection, the specificity was 70.2% in both groups combined. When combining both groups, four out of five best-performing 2-transcript combinations included *SECISBP2*, with two including *TSPO*. The expression of *TSPO* was upregulated in bacterial infections and downregulated in viral infections, whereas the reverse pattern was observed for *SECISBP2*. The signature proved to be more accurate than CRP and PCT, despite them being used in the etiologic classification of some of the children. In the combined discovery and validation groups, the AUC, sensitivity, and specificity for CRP were 0.82 (95% CI, 0.76–0.89), 73.0% and 87.9%, respectively, whereas for PCT, the respective values were 0.73 (95% CI, 0.66–0.81), 54.7% and 93.8% (Supplementary data file 2).

We further analyzed the performance of the *TSPO* and *SECISBP2* signature in children with microbiologically confirmed bacterial or viral infection without co-infection. The gene pair differentiated definite bacterial ($n = 34$) from definite viral infections ($n = 41$) with an AUC of 0.91 (95% CI, 0.84–0.97), sensitivity of 88.2% and specificity of 85.0% in the combined discovery and validation groups (Supplementary Table 3).

Additionally, we evaluated the performance of the 2-transcript signature proposed by Herberg et al.[12], which includes the *IFI44L* and *FAM89A* genes identified from children with a microbiologically confirmed bacterial or viral infection. In our discovery group, neither of these genes showed significant differential expression between bacterial and viral infections. The obtained AUC was low both in the discovery (0.63, 95% CI, 0.51–0.76) and in the combined discovery and validation groups (0.61, 95% CI, 0.52–0.71) (Supplementary data file 2).

Validation of the *TSPO* and *SECISBP2* signature by qRT-PCR

We internally validated the performance of the top-performing 2-transcript predictor for bacterial infection, *TSPO* and *SECISBP2*, by qRT-PCR in a subset of 90 samples. This subset included 61 samples from the discovery group (50 bacterial, 11 viral infections) and 29 samples from the validation group (16 bacterial, 13 viral infections). The gene pair distinguished bacterial from viral infections with an AUC of 0.82 (95% CI, 0.72–0.91) in the combined groups (Supplementary Table 4; Supplementary Figs. 1, 2 and 3). In comparison, CRP and PCT achieved AUCs of 0.86 (95% CI, 0.78–0.94) and 0.83 (95% CI, 0.74–0.92), respectively, in the combined dataset.

Discussion

In this study of children hospitalized with severe infections, the clustering of peripheral blood transcriptome could not be unequivocally explained by the etiology of infection. Although distinct expression patterns were observed between healthy children and children with a definite bacterial infection, the expression profile of children with a definite viral infection was heterogeneous. To address clinical heterogeneity, we analyzed respiratory and non-respiratory infections first separately (discovery and validation groups) and then together, categorized into bacterial and viral infections. The identified 2-transcript signature (*TSPO* and *SECISBP2*) differentiated between bacterial and viral

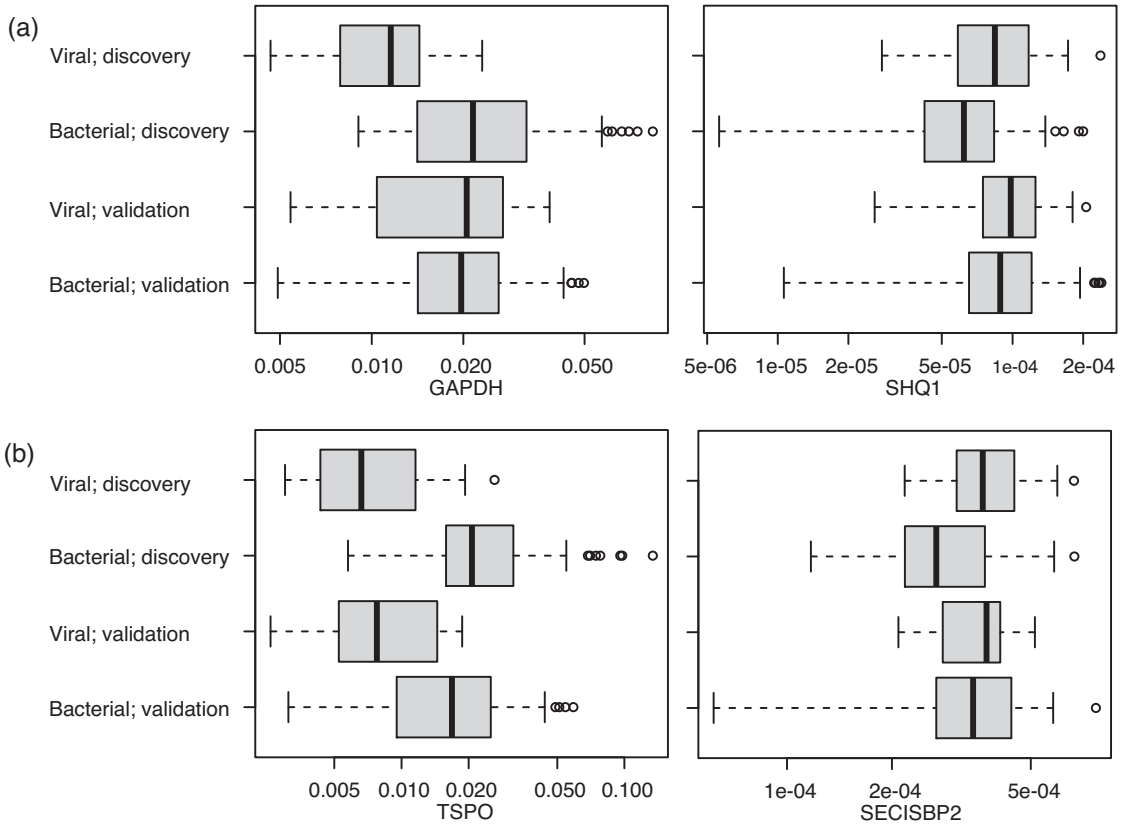

Fig. 4 | Boxplots of normalized expression levels of the 2-transcript predictors for bacterial infection. Boxplots represent the distribution of normalized expression levels of the 2-transcript predictor pairs **a** *GAPDH* and *SHQ1*, and **b** *TSPO and SECISBP2,* in the discovery (respiratory tract infections; $n = 101$) and validation (non-respiratory infections; $n = 109$) groups. In each boxplot, the central line denotes the median, the box bounds represent the interquartile range (IQR), and the whiskers extend to 1.5 × IQR or to the most extreme data point within that range. Outliers beyond the whiskers are shown individually.

infections with good accuracy in both the discovery and validation groups, despite the groups having different clinical manifestations.

The finding that the etiology of infection did not emerge as the main determinant of sample clustering can probably be explained by the inclusion of various infections with diverse foci and pathogens, as well as the wide age range of children, since leukocyte subpopulation counts change during childhood[31]. The distinct expression patterns between healthy children and hospitalized children with a definite bacterial infection are comprehensible. The heterogeneous expression profile of children with a definite viral infection may have resulted from this group consisting of children with various infectious diseases differing in viral etiology, clinical focus of infection, and illness severity.

Previous studies have identified mitochondrial translocator protein (*TSPO*) among transcripts distinguishing children with and without serious bacterial infections[13,32,33]. In adults with sepsis, *TSPO* expression correlated with disease severity and immune activation[34]. Conversely, *TSPO* was underexpressed in children with influenza compared to those with bacterial infections[35]. TSPO participates in mitochondrial bioenergetics, generation of reactive oxygen species (ROS), and regulation of inflammatory responses[36–38]. In immune cells, bacterial lipopolysaccharide stimulation increases *TSPO* expression, promoting the release of proinflammatory cytokines and ROS, possibly facilitating pathogen killing[39–41]. Beyond systemic infection, TSPO has been widely used as a biomarker of neuroinflammation, with animal studies linking its upregulation to sepsis-associated encephalopathy[42–46]. Pharmacological modulation of TSPO has demonstrated both anti-inflammatory and neuroprotective effects in preclinical models[47–54]. Taken together, our observation of *TSPO* expression as a biomarker for bacterial infection is biologically rational.

To our knowledge, no previous studies have described *SECISBP2* expression in the context of acute infections. The SECISBP2 protein regulates the incorporation of selenium into selenoproteins, which have antioxidant, immunomodulatory, and antiviral functions[55–59]. Selenium reduces the occurrence and severity of viral infections by enhancing T cell, natural killer cell and macrophage activity[60–62]. Its antioxidative effects counteract infection-induced oxidative stress[57–62]. Selenium deficiency, conversely, leads to oxidative stress, potentially enhancing viral replication, and has been associated with worse clinical outcomes during viral infections[63–67]. A plausible hypothesis for *SECISBP2* upregulation in viral infections observed in our study is the antiviral effect of selenium, mediated through SECISBP2-dependent mechanisms involving selenoproteins.

Noteworthy, with an AUC of 0.87 in the discovery plus validation groups, the identified 2-transcript signature outperformed CRP and PCT, despite their partial use in the etiologic classification. In the qRT-PCR validation within a subset of children, the AUC remained approximately similar (0.82). In a recent systematic comparison of published host gene expression signatures in discriminating between bacterial and viral infections, the signature performances varied widely, with median AUCs ranging from 0.55 to 0.96 for predicting bacterial infections and from 0.69 to 0.97 for viral infections[68]. Interestingly, diagnostic accuracy was lower in children under 11 years of age compared to adolescents and adults.

Previous RNA signature studies have predominantly focused on microbiologically confirmed single-etiology cases, often excluding complex patient populations such as those with probable or mixed etiologies[12,13,15–18,22–24,69]. However, one might argue that the latter patient populations are in the greatest need of novel diagnostic tests. While the discriminative accuracy has been excellent in rigorously defined cohorts (AUCs > 0.90; sensitivities 87–100%, and specificities 80–96%),

performance declines in probable etiologies or milder diseases[13,14,22,70,71]. Lydon et al. validated a transcript signature in adults and adolescents with acute respiratory illnesses, also in complex cases[20]. Among patients with suspected bacterial infections or viral-bacterial co-infections, the signature identified 82% and 71% of cases, respectively, as having either a bacterial infection or co-infection.

Interestingly, the 2-transcript signature by Herberg et al.[12] performed poorly in our study population, likely due to the high prevalence of viral-bacterial co-infections in our cohort. Interferon-stimulated genes (e.g., *IFI44L*) upregulated in viral infections may not effectively distinguish between viral infections and viral-bacterial co-infections, as both trigger antiviral responses. Furthermore, Herberg's signature was derived from children with severe, microbiologically confirmed infections, many requiring intensive care, whereas only one third of our children had microbiologically confirmed single-etiology infections, while another third presented with viral-bacterial co-infections. Additionally, we identified that 10% of children had a bacterial infection with an asymptomatic virus finding. Generally, RNA signature performance decreases when distinguishing between three or more phenotypes (e.g., bacterial, viral, non-infectious) compared to only two phenotypes[68].

We included clinically complex yet relevant populations, and validated the discovered 2-transcript signatures across different clinical presentations (respiratory vs non-respiratory infections). Our aim was to find out whether the identified signatures are applicable to a wide range of common infections encountered in hospitalized children. The top five gene pairs from the discovery group performed inadequately in the validation group, underscoring the need to test beyond a single syndrome. However, we identified *TSPO* and *SECISBP2* as a robust signature across presentations and in combined analyses.

Because this signature discriminated both bacterial infections and viral-bacterial co-infections from viral infections, it could help target antibiotic treatment. Given the novelty and preliminary nature of our finding, it should be further investigated and externally validated across diverse cohorts. Similar to other expression signatures, integration with clinical data would be needed. Furthermore, timely results are crucial to impact diagnostic decision-making. Our internal qRT-PCR validation is a pragmatic step towards implementable assays. Development of point-of-care tests for other validated transcript signatures has recently been reported[21,70,72]. The advantage of using gene expression analysis in measuring antiviral host response instead of direct detection of pathogens by PCR is the ability to differentiate between incidental and pathogenic virus findings, as well as to assess the potential coexisting bacterial infection. In a recent multicentre PERFORM study, extensive molecular testing for viruses in febrile children did not significantly improve the diagnostic assignment, as the identification of a virus had poor predictive value for excluding a bacterial infection[73].

A strength of our study is the relatively large, prospectively recruited cohort covering a wide pediatric age range and including children with comorbidities and complex phenotypes, thereby representing the real-world patient population of a tertiary care pediatric hospital. Another advantage is the systematic respiratory virus PCR testing, allowing for a detailed etiologic classification. As the number of children with a viral infection without bacterial involvement was initially lower than expected, we addressed this by recruiting an additional sample of children with either a suspected or confirmed viral infection. As a limitation of our study, healthy control children and a minor part of children with a viral infection were recruited from a previous study conducted at a different time point. As with all diagnostic performance studies, the lack of a golden reference standard for the etiologic classification of infectious diseases represents another limitation. We adopted a pragmatic approach in identifying and validating the signatures within a heterogeneous and complex patient population. This introduced uncertainty in the etiological classification and may have reduced the performance of the signature. However, it simultaneously enhances the generalizability of the results. We preferred a pragmatic study design, since there is a need to distinguish between not only the confirmed bacterial and viral infections but also between the less straightforward cases, which are common among hospitalized children.

Conclusion

In conclusion, our study demonstrated a good performance of a 2-transcript signature, *TSPO* and *SECISBP2*, in discriminating between bacterial and viral infections in children primarily hospitalized for suspected severe infections. It also provides insights into the utility of transcriptomics in acute pediatric infectious diseases with complex phenotypes. Heterogeneity of patient groups and frequency of viral-bacterial co-infections challenges novel diagnostic approaches. While the performance of our signature derived from a group of children with respiratory tract infections could be replicated in the validation group consisting of children with non-respiratory infections, and was internally validated by qRT-PCR in our own dataset, our discovery is preliminary and should be further investigated and externally validated in different clinical settings.

Data availability

Individual participant data used in this study are not publicly available because that would compromise research participant privacy. The de-identified data and the full analysis source code are available upon reasonable request from the corresponding author (E-mail: ruut.piri@utu.fi) pending approval of our institution, and may require a data use agreement between the parties.

Code availability

The STRTprep pipeline[29] for STRT RNA-Seq data preprocessing is available at https://github.com/shka/STRTprep/tree/v3dev. Parts of the code used for downstream analysis of STRT RNA-seq data are available as follows: https://gist.githubusercontent.com/shka/204e85b8a7a25ba8d264f5481f13c292/raw/362c47e384c26c53bccb7d6d771b1b23bbe9a397/spikein_normalization.R for spike-in normalization; https://gist.githubusercontent.com/shka/d5bdcdbe91df7e6565bdb6d2ed72805e/raw/a11e93113a9741a3e141ac7b7d94d99584315a74/test_fluctuation.R for identification of fluctuated genes; and https://gist.githubusercontent.com/shka/7ceff193189b0f914408b3e58f9fc69b/raw/da0327be4756b6a967fd3f4ab0e0b39de0177cdd/automation_SAMstrt_fluctuation.R for differential expression analysis. CFX Manager 3.1 software can be purchased from Bio-Rad. Full R scripts for downstream analyses of STRT RNA-seq and qRT-PCR data, as well as SPSS syntax used for the statistical analyses, are available upon reasonable request, together with the participant data (see data availability section).

Received: 26 August 2024; Accepted: 24 December 2025;
Published online: 09 April 2026

References

1. Jansen, R. et al. Frequent detection of respiratory viruses without symptoms: toward defining clinically relevant cutoff values. *J. Clin. Microbiol.* **49**, 2631–2636 (2011).
2. Rhedin, S. et al. Clinical utility of PCR for common viruses in acute respiratory illness. *Pediatrics* **133**, e538–e545 (2014).
3. Byington, C. L. 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.* **61**, 1217–1224 (2015).
4. Relman, D. A. New technologies, human-microbe interactions, and the search for previously unrecognized pathogens. *J. Infect. Dis.* **186**, 254–258 (2002).
5. Iroh Tam, P., Bernstein, E., Ma, X. & Ferrieri, P. Blood culture in evaluation of pediatric community-acquired pneumonia: a systematic review and meta-analysis. *Hosp. Pediatr.* **5**, 324–336 (2015).
6. Van Den Bruel, A. et al. Diagnostic value of laboratory tests in identifying serious infections in febrile children: systematic review. *BMJ* **342**, d3082 (2011).

7. Huang, D. T. et al. Procalcitonin-guided use of antibiotics for lower respiratory tract infection. *N. Engl. J. Med.* **379**, 236–249 (2018).
8. Engelmann, I. et al. Diagnosis of viral infections using myxovirus resistance protein A (MxA). *Pediatrics* **135**, e985–e993 (2015).
9. Rhedin, S. et al. Myxovirus resistance protein A for discriminating between viral and bacterial lower respiratory tract infections in children: the TREND study. *Clin. Microbiol. Infect.* **28**, 1251–1257 (2022).
10. Piri, R. et al. Myxovirus resistance protein A as a marker of viral cause of illness in children hospitalized with an acute infection. *Microbiol. Spectr.* **10**, e0203121 (2022).
11. Piri, R. et al. Evaluation of a novel point-of-care blood myxovirus resistance protein A measurement for the detection of viral infection at the pediatric emergency department. *J. Infect. Dis.* https://doi.org/10.1093/infdis/jiae367 (2024).
12. Herberg, J. A. et al. Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children. *JAMA* **316**, 835–845 (2016).
13. Mahajan, P. et al. Association of RNA biosignatures with bacterial infections in febrile infants aged 60 days or younger. *JAMA* **316**, 846–857 (2016).
14. Kaforou, M., Herberg, J. A., Wright, V. J., Coin, L. J. M. & Levin, M. Diagnosis of bacterial infection using a 2-transcript host RNA signature in febrile infants 60 days or younger. *JAMA* **317**, 1577–1578 (2017).
15. Ramilo, O. et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **109**, 2066–2077 (2007).
16. Zaas, A. K. et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe* **6**, 207–217 (2009).
17. 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).
18. Sweeney, T. E., Wong, H. R. & Khatri, P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci. Transl. Med.* **8**, 346ra91 (2016).
19. Sampson, D. L. et al. A four-biomarker blood signature discriminates systemic inflammation due to viral infection versus other etiologies. *Sci. Rep.* **7**, 1–17 (2017).
20. Lydon, E. C. et al. Validation of a host response test to distinguish bacterial and viral respiratory infection. *EBioMedicine* **48**, 453–461 (2019).
21. Tsalik, E. L. et al. Discriminating bacterial and viral infection using a rapid host gene expression test. *Crit. Care Med.* **49**, 1651–1663 (2021).
22. Habgood-Coote, D. et al. Diagnosis of childhood febrile illness using a multi-class blood RNA molecular signature. *Med* **4**, 635–654 (2023).
23. 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).
24. Ravichandran, S. et al. VB10, a new blood biomarker for differential diagnosis and recovery monitoring of acute viral and bacterial infections. *EBioMedicine* **67**, 103352 (2021).
25. Toivonen, L. et al. Blood MxA protein as a marker for respiratory virus infections in young children. *J. Clin. Virol.* **62**, 8–13 (2015).
26. Toivonen, L. et al. Rhinovirus infections in the first 2 years of life. *Pediatrics* **138**, e20161309 (2016).
27. Katayama, S. et al. Guide for library design and bias correction for large-scale transcriptome studies using highly multiplexed RNAseq methods. *BMC Bioinform.* **20**, 1–9 (2019).
28. Ezer, S. et al. Generation of RNA sequencing libraries for transcriptome analysis of globin-rich tissues of the domestic dog. *STAR Protoc.* **2**, 100995 (2021).
29. Krjutškov, K. et al. Single-cell transcriptome analysis of endometrial tissue. *Hum. Reprod.* **31**, 844–853 (2016).
30. Katayama, S., Töhönen, V., Linnarsson, S. & Kere, J. SAMstrt: statistical test for differential expression in single-cell transcriptome with spike-in normalization. *Bioinformatics* **29**, 2943–2945 (2013).
31. Erkeller-Yuksel, F. M. et al. Age-related changes in human blood lymphocyte subpopulations. *J. Pediatr.* **120**, 216–222 (1992).
32. Wong, H. R. et al. Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum. *Crit. Care Med.* **37**, 1558–1566 (2010).
33. Huang, L., Qiao, L., Zhu, H., Jiang, L. & Yin, L. Genomics of neonatal sepsis: has-miR-150 targeting BCL11B functions in disease progression. *Ital. J. Pediatr.* **44**, 145 (2018).
34. Ming, T. et al. Integrated analysis of gene co-expression network and prediction model indicates immune-related roles of the identified biomarkers in sepsis and sepsis-induced acute respiratory distress syndrome. *Front. Immunol.* **13**, 897390 (2022).
35. Herberg, J. A. et al. Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. *J. Infect. Dis.* **208**, 1664–1668 (2013).
36. Batarseh, A. & Papadopoulos, V. Regulation of translocator protein 18kDa (TSPO) expression in health and disease states. *Mol. Cell. Endocrinol.* **327**, 1–12 (2010).
37. Papadopoulos, V. et al. Translocator protein (18 kDa): new nomenclature for the peripheral-type benzodiazepine receptor based on its structure and molecular function. *Trends Pharmacol. Sci.* **27**, 402–409 (2006).
38. Betlazar, C., Middleton, R. J., Banati, R. & Liu, G. J. The translocator protein (TSPO) in mitochondrial bioenergetics and immune processes. *Cells* **9**, 512 (2020).
39. Blevins, L. K., Crawford, R. B., Azzam, D. J., Guilarte, T. R. & Kaminski, N. E. Surface translocator protein 18 kDa (TSPO) localization on immune cells upon stimulation with LPS and in ART-treated HIV^{+} subjects. *J. Leukoc. Biol.* **110**, 123–140 (2021).
40. Veenman, L., Shandalov, Y. & Gavish, M. VDAC activation by the 18 kDa translocator protein (TSPO), implications for apoptosis. *J. Bioenerg. Biomembr.* **40**, 199–205 (2008).
41. Banati, R. B. et al. Positron emission tomography and functional characterization of a complete PBR/TSPO knockout. *Nat. Commun.* **5**, 5452 (2014).
42. Rupprecht, R. et al. Translocator protein (18 kDa) (TSPO) as a therapeutic target for neurological and psychiatric disorders. *Nat. Rev. Drug Discov.* **9**, 971–988 (2010).
43. Colasanti, A. et al. In vivo assessment of brain white matter inflammation in multiple sclerosis with 18F-PBR111 PET. *J. Nucl. Med.* **55**, 1112–1118 (2014).
44. Hatori, A. et al. PET imaging of lung inflammation with [18F]FEDAC, a radioligand for translocator protein (18 kDa). *PLoS ONE* **7**, e45065 (2012).
45. Giridharan, V. V. et al. A crosstalk between gut and brain in sepsis-induced cognitive decline. *J. Neuroinflammation* **19**, 114 (2022).
46. Giga, H. et al. Pharmacological and genetic inhibition of translocator protein 18 kDa ameliorated neuroinflammation in murine endotoxemia model. *Shock* **56**, 142–149 (2021).
47. Jin, G. lin et al. Koumine regulates macrophage M1/M2 polarization via TSPO, alleviating sepsis-associated liver injury in mice. *Phytomedicine* **107**, 154484 (2022).
48. Leaver, K. R. et al. Effects of translocator protein (18 kDa) ligands on microglial activation and neuronal death in the quinolinic-acid-injected rat striatum. *ACS Chem. Neurosci.* **3**, 114–119 (2012).
49. Barron, A. M. et al. Ligand for translocator protein reverses pathology in a mouse model of Alzheimer's disease. *J. Neurosci.* **33**, 8891–8897 (2013).
50. Simon-O'Brien, E., Gauthier, D., Riban, V. & Verleye, M. Etifoxine improves sensorimotor deficits and reduces glial activation, neuronal degeneration, and neuroinflammation in a rat model of traumatic brain injury. *J. Neuroinflammation* **13**, 203 (2016).

51. Leva, G. et al. The translocator protein ligand XBD173 improves clinical symptoms and neuropathological markers in the SJL/J mouse model of multiple sclerosis. *Biochim. Biophys. Acta Mol. Basis Dis.* **1863**, 3016–3027 (2017).
52. Ma, B. et al. Translocator protein agonist Ro5-4864 alleviates neuropathic pain and promotes remyelination in the sciatic nerve. *Mol. Pain* **14**, 1744806917748019 (2018).
53. Monga, S., Nagler, R., Amara, R., Weizman, A. & Gavish, M. Inhibitory effects of the two novel TSPO ligands 2-Cl-MGV-1 and MGV-1 on LPS-induced microglial activation. *Cells* **8**, 486 (2019).
54. Dominguini, D. et al. The protective effect of PK-11195 on cognitive impairment in rats survived of polymicrobial sepsis. *Mol. Neurobiol.* **58**, 2724–2733 (2021).
55. Copeland, P. R., Fletcher, J. E., Carlson, B. A., Hatfield, D. L. & Driscoll, D. M. A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J.* **19**, 306–314 (2000).
56. Rayman, M. P. Selenium and human health. *Lancet* **379**, 1256–1268 (2012).
57. Avery, J. C. & Hoffmann, P. R. Selenium, selenoproteins, and immunity. *Nutrients* **10**, 1203 (2018).
58. Guillin, O. M., Vindry, C., Ohlmann, T. & Chavatte, L. Selenium, selenoproteins and viral infection. *Nutrients* **11**, 2101 (2019).
59. Fairweather-Tait, S. J. et al. Selenium in human health and disease. *Antioxid. Redox Signal.* **14**, 1337–1383 (2011).
60. Bae, M. & Kim, H. Mini-review on the roles of vitamin C, vitamin D, and selenium in the immune system against COVID-19. *Molecules* **25**, 5346 (2020).
61. Hoffmann, F. K. W. et al. Dietary selenium modulates activation and differentiation of CD4+ T cells in mice through a mechanism involving cellular free thiols. *J. Nutr.* **140**, 1155–1161 (2010).
62. Kiremidjian-Schumacher, L., Roy, M., Harvey, W., Martin, C. & Stotzky, G. Supplementation with selenium and human immune cell functions. I. Effect on lymphocyte proliferation and interleukin 2 receptor expression. *Biol. Trace Elem. Res.* **41**, 103–104 (1994).
63. Beck, M. et al. Selenium deficiency increases the pathology of an influenza virus infection. *FASEB J.* **15**, 1481–1483 (2001).
64. Beck, M. A. et al. Benign human enterovirus becomes virulent in selenium-deficient mice. *J. Med. Virol.* **43**, 166–170 (1994).
65. Moghaddam, A. et al. Selenium deficiency is associated with mortality risk from COVID-19. *Nutrients* **12**, 1–13 (2020).
66. Khan, M., Dilawar, S., Ali, I. & Rauf, N. The possible role of selenium concentration in hepatitis B and C patients. *Saudi J. Gastroenterol.* **18**, 106–110 (2012).
67. Stone, C. A., Kawai, K., Kupka, R. & Fawzi, W. W. Role of selenium in HIV infection. *Nutr. Rev.* **68**, 671–681 (2010).
68. Bodkin, N. et al. Systematic comparison of published host gene expression signatures for bacterial/viral discrimination. *Genome Med.* **14**, 1–14 (2022).
69. Ko, E. R. et al. Host-response transcriptional biomarkers accurately discriminate bacterial and viral infections of global relevance. *Sci. Rep.* **13**, 22554 (2023).
70. Ko, E. R. et al. Prospective validation of a rapid host gene expression test to discriminate bacterial from viral respiratory infection. *JAMA Netw. Open* **5**, e227299 (2022).
71. Barral-Arca, R., Pardo-Seco, J., Martinón-Torres, F. & Salas, A. A 2-transcript host cell signature distinguishes viral from bacterial diarrhea and it is influenced by the severity of symptoms. *Sci. Rep.* **8**, 4–10 (2018).
72. Pennisi, I. et al. Discrimination of bacterial and viral infection using host-RNA signatures integrated in a lab-on-chip platform. *Biosens. Bioelectron.* **216**, 114633 (2022).
73. Shah, P. et al. Relationship between molecular pathogen detection and clinical disease in febrile children across Europe: a multicentre, prospective observational study. *Lancet Reg. Health Eur.* **26**, 100682 (2023).

Acknowledgements

We thank Auli Saarinen for skilled technical assistance and CSC–IT Center for Science, Finland, for computational resources. This work was supported by the University of Turku (to R.P.); Research funds from specified government transfers to hospital district of Southwest Finland (to R.P. and V.P.); the Foundation for Pediatric Research (to R.P. and V.P.); Maud Kuistila Memorial Foundation (2021-0105B to R.P.); and Päivikki and Sakari Sohlberg Foundation (230055 to R.P.). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author contributions

Concept and design: R.P., J.L., L.I., and V.P. Methodology: R.P., M.V., J.L., J.K., S.K., and V.P. Acquisition and interpretation of data: all authors. Statistical analyses: R.P., S.K. Drafting of the manuscript: R.P. Critical revision of the manuscript for important intellectual content: all authors. Figures: R.P., S.K. Supervision: S.K., V.P. (equally). All authors approved the final version of the manuscript to be published.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-025-01370-z.

Correspondence and requests for materials should be addressed to Ruut Piri.

Peer review information *Communications Medicine* thanks Meirav Mor and the other anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

[1]Department of Paediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland. [2]Immunogenetics Laboratory, Institute of Biomedicine, University of Turku, Turku, Finland. [3]Department of Clinical Microbiology, Turku University Hospital, Turku, Finland. [4]InFLAMES Research Flagship Center, University of Turku, Turku, Finland. [5]Folkhälsan Research Center, Helsinki, Finland. [6]Stem Cells and Metabolism Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [7]Department of Medicine, Karolinska Institutet, Huddinge, Sweden. [8]These authors jointly supervised this work Shintaro Katayama Ville Peltola. ✉e-mail: ruut.piri@utu.fi