Nasopharyngeal gene expression, a novel approach to study the course of respiratory syncytial virus infection


Affiliations: 1Laboratory of Pediatric Infectious Diseases, Dept of Pediatrics, Radboud University Medical Center, Nijmegen, The Netherlands. 2Laboratory of Medical Immunology, Dept of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands.

Correspondence: Corné van den Kieboom, Laboratory of Pediatric Infectious Diseases, Dept of Pediatrics, Radboud University Medical Center, PO box 9101, Nijmegen 6500HB, The Netherlands. E-mail: Corne.vandenKieboom@radboudumc.nl

ABSTRACT Respiratory syncytial virus (RSV) causes mild infections in the vast majority of children. However, in some cases, it causes severe disease, such as bronchiolitis and pneumonia. Development of severe RSV infection is determined by the host response. Therefore, the main aim of this study was to identify biomarkers associated with severe RSV infection.

To identify biomarkers, nasopharyngeal gene expression was profiled by microarray studies, resulting in the selection of five genes: ubiquitin D, tetraspanin 8, mucin 13, b-microseminoprotein and chemokine ligand 7.

These genes were validated by real-time quantitative PCR in an independent validation cohort, which confirmed significant differences in gene expression between mildly and severely infected and between recovery and acute patients.

Nasopharyngeal aspirate samples are regularly taken when a viral respiratory tract infection is suspected. In this article, we describe a method to discriminate between mild and severe RSV infection based on differential host gene expression. The combination of pathogen detection and host gene expression analysis in nasopharyngeal aspirates will significantly improve the diagnosis and prognosis of respiratory tract infections.

Host gene expression analysis of nasopharyngeal aspirate samples can be used to predict RSV disease severity http://ow.ly/AMxpt
Introduction
Respiratory syncytial virus (RSV) is the most common cause of viral respiratory tract infections among hospitalised children. Symptoms range from a common cold to severely compromised respiratory function, causing 6–15% of the admitted children to require intensive care [1, 2]. It is difficult to predict the course of disease in infants, especially in children under the age of 3 months. Current clinical prediction rules are based on demographic criteria and clinical symptoms [3–10] and although they may improve clinical judgement [7, 8], they are not routinely used in daily practice. More objective and reproducible predictions may be achieved by using biomarkers [11–15].

Recent publications showed that gene expression profiles in peripheral blood mononuclear cells represent the subject’s health. RAMILLO et al. [16] were able to discriminate influenza A, Escherichia coli and Streptococcus pneumoniae infection with 95% accuracy, and distinguished E. coli from Staphylococcus aureus infection with 85% accuracy by analysing mRNA expression level in young patients. In the case of RSV infections, MEJAS et al. [17] showed, based on blood RNA profiles of infants, a classification of infants with RSV lower respiratory tract infection (LRTI) versus rhinovirus or influenza LRTI with 95% accuracy. Additional studies confirmed these observations and described the discovery of several biomarkers in blood by which different states of disease could be distinguished [18].

These findings may be of great value for clinical decision making. However, the studies, as described, were performed using blood samples, of which only limited volumes can be obtained from paediatric patients. Thus, it would be worthwhile to study whether other clinical materials could be used that reflect the host response and may harbour prognostic value. The nasopharynx is the point of entry of all respiratory viruses; moreover, a sample from this site is already taken to diagnose viral infections using real-time quantitative PCR (RT-qPCR) [19]. The use of nasopharyngeal aspirates (NPAs) for the measurement of host biomarker expression for prognostic purposes would be worthwhile to assess, especially as clinical laboratories would only need minimal adjustment to integrate this test into current practice. Knowledge with regards gene expression in the nasopharynx during RSV infections is limited and has been gained solely from animal models [20].

In this study, we assessed the possibility of measuring host gene expression in NPA taken from children with laboratory-confirmed RSV infections and aimed to identify markers with potential prognostic value for the course of disease.

Materials and methods

Ethics
The study was approved by the Committee on Research Involving Human Subjects of the Radboud University Medical Center (Nijmegen, the Netherlands). All clinical samples collected from patients in our cohorts were enrolled following informed consent from parents or guardians. The clinical information available to subjects being recruited for the study included a clear description of the risks and benefits of participation, indicating the required insurance and procedures, safe handling and protection of all data. All subjects were informed that there was no obligation to participate and that declining to participate or leaving the study had no adverse consequences.

Study design
Children <5 years of age with symptoms of viral LRTI were included. These symptoms included increased respiratory effort (e.g. tachypnoea and/or use of accessory respiratory muscles) and/or expiratory wheezing and/or crackles and/or apnoea. Retrospectively, the LRTI was confirmed by real-time quantitative PCR (RT-qPCR) on nasopharyngeal washes as described previously [19]. Patients with congenital or acquired immune deficiency, immunosuppressive medication (including >24 h of glucocorticosteroids) or severe psychomotor retardation were excluded.

All subjects were recruited at two hospitals in Nijmegen. Subjects admitted were screened for symptoms typical for viral respiratory tract infection, as described previously. When RSV infection was suspected, a NPA was collected within 24 h of admission (acute) and parents of hospitalised children were asked permission to draw a second NPA sample 4–6 weeks after admission (recovery). Medical history, demographics and clinical parameters were collected from questionnaires and medical records. Patients were classified into three different groups based on the degree of supportive care they needed during the course of disease: no supportive care; supplemental oxygen (when oxygen saturation was <93% for >10 min); or mechanical ventilation.

Microarrays were performed on samples from the cohort published by BRAND et al. [18] (n=30, 2006–2009) and subsequently selected markers were validated in the cohort of I.M.L. Ahout et al. (manuscript in preparation) (n=44, 2010–2012).
Sample collection
After informed consent, a NPA sample was collected by introducing a catheter, connected to a collection tube and an aspiration system, into the nasopharyngeal cavity. Then, saline (0.5–1.5 mL) was instilled into the catheter and, while slowly retracting the catheter, the nasopharyngeal fluid was aspirated in a collection tube. Afterwards, the catheter was flushed with saline and added to the collection fluid.

The samples were kept on ice and immediately transferred to the laboratory, where they were centrifuged (500 × g) for 10 min at 4°C to spin down the mucus and cells, after which the pellet was mixed with Trizol (Life Technologies, Carlsbad, CA, USA) and stored at -80°C. A RT-qPCR was performed using the supernatant to confirm viral aetiology of the disease [19].

Microarray
RNA was extracted from NPA in Trizol according to the manufacturer’s protocol. The final RNA pellet was washed twice with 1 mL of 75% ethanol, air dried and resuspended in 100 μL of RNase/DNase free water. Subsequently, a clean-up of the total RNA was performed with the RNeasy Minikit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. RNA integrity and quality for the microarray was assessed using a RNA 6000 Nano LabChip on an Bioanalyzer 2100 system (Agilent, Santa Clara, CA, USA).

RNA processing, target labelling and hybridisation to gene expression arrays was performed. Biotin-labelled cRNA was obtained using the One-Cycle Eukaryotic Target Labeling Assay (Affymetrix Inc., Santa Clara, CA). 15 μg of fragmented, biotin-labelled cRNA was hybridised to Affymetrix GeneChip Human Genome U133 Plus 2.0 arrays according to standard Affymetrix protocol.

Microarray data analysis
Quality control analyses were performed as previously described [21]. Scanned images were inspected for artefacts, percentage of present calls (<25%) and controls of RNA degradation. Background signal was removed using robust multichip analysis and probe intensity levels were quantile-normalised across arrays. Microarray analysis was performed using ArrayStar4 (DNAStar, Madison, WI, USA).

After the microarray data were obtained, comparisons were made on supportive care: the need for supportive care (comparison 1) and mechanical ventilation (comparison 2). p-values were corrected for false discovery rate and, subsequently, the significant genes with a four-fold change were selected, while probes directed at noncoding sequences were excluded [22].

The selected genes were not normally distributed according to the d’Agostino and Pearson omnibus normality test (α=0.05) (Prism 5; Graphpad Software, La Jolla, CA, USA). Therefore, a permutation test using MultiExperiment Viewer (tm4group; Dana-Farber Cancer Institute, Boston, MA, USA) with random group samples (10^5) was performed and the median fold change cut-off was set at 2.50 log. The overlapping genes between the two comparisons and the only upregulated gene were selected for validation by RT-qPCR.

Gene validation by RT-qPCR
Samples were defrosted on ice and RNA was extracted from NPA in Trizol according to the manufacturer’s protocol. The RNA pellet was washed twice with 1 mL of 75% ethanol, air dried, re-suspended in 100 μL of diethylpyrocarbonate-treated water. Subsequently, a clean-up was performed with the RNeasy Minikit according to the manufacturer’s instructions. The elutate was treated with TURBO DNA-free (Life Technologies); subsequently, cDNA synthesised by iScript (Biorad, Hercules, CA, USA) and stored at -20°C.

The RT-qPCR was conducted on a CFX96 (Biorad) using Taqman gene expression assays (Hs.0017774_m1, Hs.00610327_m1, Hs.00217239_m1, Hs.00738231_m1, Hs.00171147_m1 and Hs.02758991_g1; Life Technologies) and IQ Powermix (Biorad). The PCR programme consisted of an initial 5 min at 96°C, followed by 40 cycles of 15 s at 96°C and 45 s at 60°C. The data generated were analysed using CFX Manager 3.0 (Biorad) and the cut-off value was set by using the single threshold mode for quantification cycle (C_q) determination. The data were normalised by subtracting the C_q value of the housekeeping gene, i.e. glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (Hs.02758991_g1; Life Technologies).

Statistical analysis
Significance of the microarray analysis was calculated by a permutation test, with 10^5 random group samples using TMEV (tm4group). Significance of the RT-qPCR-measured gene expression was calculated by Mann–Whitney U-tests using Prism 5, where a two-sided value of p<0.05 was considered statistically significant.

Correlation was assessed by Spearman analysis using SPSS (IBM, Armonk, NY, USA), where a two-sided value of p<0.05 was considered statistically significant. For the patient characteristics, the Kruskal–Wallis
analysis was performed; in the case of significance (p<0.05), we subsequently performed Mann–Whitney U-tests for individual comparisons using SPSS.

Results
Microarray data analysis
In this study, the clinical value of nasopharyngeal gene expression was assessed by conducting microarrays on NPAs of 30 infants with a confirmed RSV infection, both mono- (RSV) and co-infections. The patient characteristics are displayed in table 1. No significant differences were found, except for the parameters on which the clinical categorisation was based and gestational age.

To select potential biomarkers, comparisons were made using microarray analysis with regards to the need for supportive care: supplemental oxygen (comparison 1) and the need for mechanical ventilation (comparison 2). The permutation test led to an initial selection of 53 targets (online supplementary table S1) for comparison 1 and 93 targets for comparison 2 (table S2). By the subsequent introduction of a median fold change cut-off of 2.5 log, 16 genes were selected that were significantly differentially expressed in comparison 1 (table S3) and 28 genes were selected in comparison 2 (table S4). The two selections were compared and four overlapping genes were found (ubiquitin D (UBD), tetraspanin 8 (TSPAN8), mucin 13 (MUC13) and β-microseminoprotein (MSP)), which all showed downregulation. In addition to these genes, chemokine ligand 7 (CCL7) was selected as the only upregulated differentially expressed gene in the whole microarray analysis (table 2).

Gene expression validation
In order to confirm differential expression of the genes selected by microarray analysis, RT-qPCR was used to validate these genes in an independent cohort. This cohort consisted of patients with RSV mono-infections to focus on the RSV specificity of the discovered genes. The final cohort consisted of 42 RSV-infected infants, of whom the patient characteristics are displayed in table 3. No significant differences were found between the groups, except for the parameters on which the clinical categorisation was based, age and RSV serotype.

To assess the overall sample quality, the expression of GAPDH in all samples was plotted (fig. 1). Generally, the amounts of RNA in the NPA samples were low; in a few samples, the amounts were too low to be able to measure gene expression, so these samples were excluded from further analysis.

These results clearly show a low but consistent expression in all samples, irrespective of the patient’s health status. RT-qPCR analysis of our selected genes using nasopharyngeal samples from the validation cohort confirmed that the expression of all selected genes was significantly different, when comparing the recovery

### TABLE 1 Microarray cohort patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>No supportive care</th>
<th>Supplemental oxygen</th>
<th>Mechanical ventilation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients n</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Age days</td>
<td>227 (57–279)</td>
<td>135 (46–337)</td>
<td>64 (41–161)</td>
<td>NS</td>
</tr>
<tr>
<td>Males n (%)</td>
<td>8 (80)</td>
<td>8 (80)</td>
<td>8 (80)</td>
<td>NS</td>
</tr>
<tr>
<td>Gestational age weeks</td>
<td>40 (37–41)</td>
<td>38 (35–40)</td>
<td>33 (33–35)</td>
<td>&lt;0.01***##</td>
</tr>
<tr>
<td>RSV load Cq</td>
<td>31 (25–33)</td>
<td>31 (24–33)</td>
<td>31 (28–32)</td>
<td>NS</td>
</tr>
<tr>
<td>days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of oxygen</td>
<td>0</td>
<td>3 [1–6]</td>
<td>9 [5–11]</td>
<td>&lt;0.001***###</td>
</tr>
<tr>
<td>therapy days</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001***###</td>
</tr>
<tr>
<td>Length of hospital</td>
<td>0 [0–3]</td>
<td>5 [2–7]</td>
<td>13 [7–16]</td>
<td>&lt;0.001***###</td>
</tr>
<tr>
<td>stay days</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001***###</td>
</tr>
</tbody>
</table>

Data are presented as median (interquartile range) unless otherwise stated. RSV: respiratory syncytial virus; Cq: quantification cycle; NS: nonsignificant; NA: not applicable. **: p<0.01 for no supportive care versus supportive care (supplemental oxygen or mechanical ventilation); ***: p<0.001 for no supportive care versus supportive care; **: p<0.01 for no mechanical ventilation (no supportive care or supplemental oxygen only) versus mechanical ventilation; ##: p<0.001 for no mechanical ventilation versus mechanical ventilation; **: p<0.01 for no supportive care versus mechanical ventilation; ###: p<0.001 for no supportive care versus mechanical ventilation; ##: p<0.01 for no supportive care versus mechanical ventilation; **: p<0.001 for no supportive care versus mechanical ventilation; *: p<0.05 for supplemental oxygen versus mechanical ventilation; **: p<0.01 for supplemental oxygen versus mechanical ventilation; ****: p<0.001 for no supportive care versus supplemental oxygen versus mechanical ventilation; +: p<0.05 for no supportive care versus supplemental oxygen; ++: p<0.01 for no supportive care versus supplemental oxygen.
with acute samples. In this analysis, the acute group consisted of pooled expression data from all three groups: no supportive care, supplemental oxygen and mechanical ventilation. In case of TSPAN8, MUC13, MSP and CCL7, the expression differed significantly in comparison 2 (the need for mechanical ventilation) (fig. 2). Subsequently, the correlation between the level of gene expression and known risk factors was assessed. Smoking at home, smoking during pregnancy, eczema, asthma and atopic family history did not show any correlation. For UBD only, a correlation with gestational age was found; however, gestational age is a well-known confounder, which does not exclude the value of UBD as severity marker [23].

**Discussion**

This study shows that nasopharyngeal aspirates taken to diagnose the viral cause of infection can also be used to identify biomarkers by microarray analysis that are associated with severe RSV infections. The value of these markers was confirmed in an independent patient cohort. To our knowledge, this is the first study in which nasopharyngeal host gene expression was studied in the context of infection.

Gene expression analysis was unforeseen at the time of collection and a large part the material was used for other studies; therefore, the quantity of the total material and the quality of the RNA isolated from the NPAs was suboptimal. The results of the GAPDH expression assay show that the absolute Cq values are high, but consistent.

**Table 2 Microarray cohort based gene selection**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene product</th>
<th>Supportive care</th>
<th>Mechanical ventilation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-value</td>
<td>Median fold change</td>
<td>p-value</td>
</tr>
<tr>
<td>UBD</td>
<td>Ubiquitin D</td>
<td>0.00271</td>
<td>-3.50</td>
</tr>
<tr>
<td>TSPAN8</td>
<td>Tetraspanin 8</td>
<td>0.00429</td>
<td>-3.39</td>
</tr>
<tr>
<td>MUC13</td>
<td>Mucin 13</td>
<td>0.00593</td>
<td>-3.54</td>
</tr>
<tr>
<td>MSP#</td>
<td>β-microseminoprotein</td>
<td>0.00168</td>
<td>-4.09</td>
</tr>
<tr>
<td>CCL7</td>
<td>Chemokine ligand 7</td>
<td>NA</td>
<td>2.70</td>
</tr>
</tbody>
</table>

NA: not applicable. #: multiple probes were found; the average p-value and median fold change are displayed.

**Table 3 Validation cohort patient characteristics**

<table>
<thead>
<tr>
<th></th>
<th>No supportive care</th>
<th>Supplemental oxygen</th>
<th>Mechanical ventilation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients n</td>
<td>7</td>
<td>21</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Age days</td>
<td>153 (102–325)</td>
<td>74 (55–179)</td>
<td>42 (28–76)</td>
<td>&lt;0.05*</td>
</tr>
<tr>
<td>Males</td>
<td>5 (71)</td>
<td>10 (48)</td>
<td>7 (50)</td>
<td>NS</td>
</tr>
<tr>
<td>Gestational age weeks</td>
<td>39 (35–40)</td>
<td>39 (37–40)</td>
<td>39 (37–40)</td>
<td>NS</td>
</tr>
<tr>
<td>RSV load Cq</td>
<td>22 (21–22)</td>
<td>25 (22–29)</td>
<td>24 (21–26)</td>
<td>NS</td>
</tr>
<tr>
<td>RSV serotype</td>
<td>6 (86)</td>
<td>11 (52)</td>
<td>14 (100)</td>
<td>&lt;0.01**</td>
</tr>
<tr>
<td>Duration of symptoms days</td>
<td>5 (4–5)</td>
<td>4 (3–5)</td>
<td>3 (2–5)</td>
<td>NS</td>
</tr>
<tr>
<td>Duration of supplemental oxygen days</td>
<td>0</td>
<td>3 (2–4)</td>
<td>9 (7–11)</td>
<td>&lt;0.001***,###</td>
</tr>
<tr>
<td>Length of hospital stay days</td>
<td>2 (0–6)</td>
<td>5 (4–8)</td>
<td>11 (10–13)</td>
<td>&lt;0.001***,###</td>
</tr>
</tbody>
</table>

Data are presented as median [interquartile range] or n (%), unless otherwise stated. RSV: respiratory syncytial virus; Cq: quantification cycle; NS: nonsignificant; NA: not applicable. *: p<0.05 for no supportive care versus supportive care [supplemental oxygen or mechanical ventilation]; **: p<0.01 for no supportive care versus supportive care (supplemental oxygen or mechanical ventilation); ***: p<0.001 for no supportive care versus supportive care (supplemental oxygen or mechanical ventilation); #: p<0.05 for no mechanical ventilation (no supportive care or supplemental oxygen only) versus mechanical ventilation; ###: p<0.001 for no mechanical ventilation (no supportive care or supplemental oxygen only) versus mechanical ventilation; **: p<0.01 for no supportive care versus supplemental oxygen; ***: p<0.001 for no supportive care versus supplemental oxygen; **: p<0.01 for no supportive care versus mechanical ventilation; **: p<0.001 for no supportive care versus mechanical ventilation; **: p<0.01 for supplemental oxygen versus mechanical ventilation; ***: p<0.001 for supplemental oxygen versus mechanical ventilation.

DOI: 10.1183/09031936.00085614
The microarray analysis resulted in the discovery of five significantly differentially expressed genes in the nasopharynx of infants with different degrees of severities of RSV infection. Remarkably, most genes are down-regulated in severely ill patients; this was confirmed with RT-qPCR. In addition, previously published microarray studies performed on gene expression profiles in peripheral blood of infants also showed a significant downregulation of genes [16, 17].

The method, as described in this study, with the aim to identify genes with potential prognostic value, has proven successful by the validation of TSPAN8, MUC13, MSP and CCL7. Although UBD did not discriminate between any acute group, it differed significantly between the recovered and acutely ill patients. This clearly underlines the value of our approach, in which microarray analysis was used to select differentially expressed genes in a discovery cohort, which were independently validated by qPCR and RT-qPCR in a separate cohort.

Besides CCL7, none of the genes identified in this study has previously been described to be associated with RSV. UBD participates in the ubiquitination process, labelling proteins for their degradation via the proteasome. MIELECH et al. [24] have shown that coronaviruses actively reduce ubiquitination of host cell proteins and observed their potential to modify the innate immune response. With regard to TSPAN8, YUE et al. [25] described its ability to override the adhesive features of CD151. On this note, CD151 is involved with the infectious entry of human papillomavirus [26], suggesting a significant infectious advantage for RSV when TSPAN8 is downregulated. Reduced amounts of MUC13 lead to decreased chemokine secretion in response to tumour necrosis factor-α. This pro-inflammatory activity of MUC13 suggests that disrupted or inappropriate expression of MUC13 could predispose to infectious disease [27]. The downregulation of MSP was previously described in the context of rhinovirus infection [28]. MYGATT et al. [29] reported virus-mediated silencing of MSP, which seems in accordance with our findings. Finally, significant upregulation of CCL7 has been observed in mice infected with rhinovirus [20] and in children during infection by different respiratory viruses, including RSV [30]. In addition, CCL7 was shown to be specifically upregulated when no mechanical ventilation was applied during RSV infection in mice and not differentially expressed when uninfected mice received mechanical ventilation [31], strongly suggesting that CCL7 upregulation is caused by RSV infection and is not a consequence of mechanical ventilation.

In conclusion, based on differential expression of TSPAN8, MUC13, MSP and CCL7 in readily available NPA samples, we can discriminate between severity of disease in RSV infected infants. This proves the potential of additional host gene expression analyses using regularly taken NPA samples, which, when combined with pathogen detection, would improve diagnosis of respiratory tract infections. Therefore, our findings demand further studies that might lead to the implementation of nasopharyngeal gene expression in diagnostics to guide clinical decisions in the context of respiratory tract infections.

References
FIGURE 2 Nasopharyngeal expression of a–c) ubiquitin D (UBD), d–f) tetraspanin 8 (TSPAN8), g–i) mucin 13 (MUC13), j–l) β-microseminoprotein (MSP) and m–o) chemokine ligand 7 (CCL7) in recovery versus acute (a, d, g, j and m), need for supportive care (b, e, h, k and n) and need for mechanical ventilation (c, f, i, l and o). Data are presented as median and interquartile range. Data were Mann–Whitney U-tested, two-sided (recovery, n = 23; acute, n = 42). ΔCq: glyceraldehyde 3-phosphate dehydrogenase-normalised quantification cycle. *: p < 0.05; **: p < 0.01; ***: p < 0.001.


