In-silico immune cell deconvolution of the airway proteomes of infants with pneumonia reveals a link between reduced airway eosinophils and an increased risk of mortality

Charles J Sande[1†], Jacqueline M Waeni, James M Njunge[1], Martin N Mutunga[1], Elijah Gicheru[1], Nelson K Kibinge[1], Agnes Gwela[1]

1. KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

[†] Corresponding author email: csande@kemri-wellcome.org

Keywords: Respiratory syncytial virus, proteome, microbiome, secondary bacterial infection

Abstract

Rationale: Pneumonia is a leading cause of mortality in infants and young children. The immune responses in the airway that are associated with mortality are poorly understood. Studies of the cellular immunology of the infant airway have traditionally been hindered by the limited sample volumes available from the young, frail children who are admitted to hospital with pneumonia. This is further compounded by the relatively low frequencies of certain immune cell phenotypes that are thought to be critical to the clinical outcome of infection. To address this, we developed a novel in-silico deconvolution method for inferring the frequencies of immune cell phenotypes in the airway of children with different survival outcomes using proteomic data.

Methods: Using high-resolution mass spectrometry, we identified > 1,000 proteins expressed in the airways of children who were admitted to hospital with clinical pneumonia. 61 of these children were discharged from hospital and survived for more than 365 days after discharge, while 19 died during admission. We used unsupervised learning by random forest to derive protein classification markers that could be used to deconvolve individual immune cell phenotypes. We applied these phenotype-specific signatures to high-resolution mass spectrometry-based proteomic data obtained from airway samples collected at admission from infants and children who were discharged from hospital and survived for at least one year as well as those who died from pneumonia during the course of admission.

Main Results: We identified protein classification markers for 33 immune cell phenotypes. Eosinophil-associated protein markers were significantly elevated in airway secretions obtained from pneumonia survivors and were downregulated in children who subsequently died. To confirm these results, we analyzed clinical parameters from

>10,000 children who had been admitted with pneumonia in the previous 10 years. The results of this retrospective analysis mirrored airway deconvolution data and showed that survivors had significantly elevated eosinophils at admission compared to fatal cases of pneumonia.

Conclusions: Airway eosinophils appear to be a critical immune cell phenotype for pneumonia survival in infants and young children.

Introduction

Pneumonia is a leading cause of paediatric mortality word-wide. A recent study on the global burden of paediatric pneumonia conducted in seven countries found that viruses account for about 61% of all paediatric pneumonia infections, while about 27% of infection were attributed to bacterial pathogens, with RSV accounting for the largest etiological fraction of paediatric pneumonia[1]. More than 90% of the deaths that occur due to pneumonia in children under 5, occur in low resource settings, mainly due to the lack of paediatric intensive care facilities[2]. Very young infants especially those with comorbidities such as HIV and malnutrition have a poor survival prognosis following pneumonia infection. HIV-infected infants who develop a pneumonia infection are up to 10 times more likely to die from the infection than non-HIV infected children[3], while those with malnutrition are more than 3 times more likely to die after admission[4]. The damage to the lungs caused by severe pneumonia appears to persist even after discharge from hospital, with recent estimates showing that post-discharge mortality in African children previously admitted to hospital with pneumonia being 2.5 times greater than those discharged with other diagnosis[5].

The difference in the immunological response to pneumonia between children who succumb to infection and those who survive it is not clear. An improved understanding of the immunobiology of this elevated post infection mortality risk will be crucial in identifying prognostic biomarkers of poor outcome, which will be critical in guiding care decisions in the first critical hours after admission. Most studies on the mechanisms of severe pneumonia in infants have been done using blood samples[6], and whist these studies have provided significant insights into disease pathology, they might not fully

recapitulate the immune response to infection in the airway. In the case of RSV, there has only been one study in the last 60 years that examined the lung samples of children who died from pneumonia. Archived post-mortem lung tissue from three children who died in the pre-intensivist era in the USA in the 1930s and 1940s was evaluated and airway obstruction with inflammatory cells, fibrin, mucus, and fluid identified as a prominent feature in RSV lung infection[7]. Staining for immune cell populations was not technically feasible in these long term archival samples.

Due to the paucity of mechanistic data on the immunological response to pneumonia in the airway there is continuing interest in understanding the dynamics of airway-resident immune cells following pneumonia infection in infants. This effort has traditionally been limited by the unsuitability of routine airway sampling techniques for conventional cytometric analyses. Samples from nasopharyngeal washings and naso- and oropharyngeal swabbing are the most common methods of sampling the airways of sick children, and whilst ideal for molecular diagnostics, they are less suitable for phenotyping of airway resident immune cells using conventional cytometry techniques. Sampling of the airway by these methods typically results in limited cell yields and the cells that are recovered are generally highly enriched for granulocytes, precluding detailed characterisation of less abundant phenotypes using traditional flow cytometry-based tools[8]. Recent analysis of the cellular composition of upper airway by flow cytometry showed that the typical abundances of critical effector cells like T Cells, B Cells, Mast cells, Dendritic cells and NK cells to be less than 0.5% of all airway cells, while granulocytes were present at a median frequency of >90%[9].

A potential way to address these shortcomings is to monitor changes in marker proteins that are uniquely expressed by specific immune cell phenotypes and to use this information to infer the dynamics of the underlying cell types. We have recently described a high-resolution mass-spectrometry-based proteomics approach for characterizing the total proteome of the infant airway to a depth of more than 1,800 proteins[10,11]. The airway proteome characterized by this technique represents an unbiased snapshot of the underlying cell populations and can be leveraged to infer changes in the frequencies of the contributing cell phenotypes. Here, we describe an in-silico immune cell phenotype deconvolution approach where we use protein markers that are uniquely expressed by different immune cell phenotypes to infer the dynamics of those phenotypes in airways of children who survived or died following a pneumonia infection. Phenotype-specific markers were derived from a previously published data set containing the individual proteomes of purified immune cell populations (deconvolution data set) and these were then applied to airway proteome data from children with different pneumonia outcomes. Using this information, we identified an eosinophil related protein signature was elevated in the airways of children who survived pneumonia but that was downregulated in fatal pneumonia and in well controls. We subsequently validated these findings using a large retrospective pneumonia cohort of >10,000 children.

Results

We used high-resolution mass spectrometry to characterise the airway proteomes of 90 infants and children who met WHO criteria for the clinical syndromes of severe pneumonia[12]. Using a false discovery rate (FDR) of 5%, we identified >1,000 proteins in the airways of infants and children with different outcomes of pneumonia. In order to

resolve the immune cell populations that comprised the airway proteomes, we used a data set that had been previously published by Rieckmann et al.[13] to derive protein markers that could be used to distinguish between different immune cell populations. The data set contained the individual proteomes of different haematopoietic cell populations in different activation states (deconvolution data set – supplementary table 1b). Figure 1 is a graphical flowchart of the experimental and analytical design features of this study. We used ordination analysis by nonmetric multidimensional scaling (NMDS) to visualize differences in phenotype specific protein expression in the deconvolution data set in order to determine whether differences in overall protein expression could be used to resolve major immune phenotypes on the basis of protein expression alone. The NMDS analysis showed that major immune cell phenotypes including B-cells, T cells, natural killer (NK) cells, dendritic cells(DC), monocytes (MO), basophils, eosinophils and neutrophils could be distinctly segregated on the basis of differential protein expression (figure 2a). We then set out to identify individual protein markers that could be used to accurately distinguish major and sub immune phenotypes (sub phenotypes defined as lower functional hierarchies of the major phenotypes – e.g. plasmacytoid or myeloid DCs) within the deconvolution data set. Using random forest (RF) classification we identified protein classification features for 33 immune cell phenotypes (supplementary table 1b) which were differentially expressed between the phenotypes and that could potentially be used to disaggregate them (see methods for details of the RF procedure). We then examined the airway proteomes of children with pneumonia to determine whether any of the RF-derived classifiers from the deconvolution data set were also present in the infant airways. For some cell types such as monocytes/macrophages and neutrophils, a substantial proportion (>30%) of the RF classifiers from the deconvolution analysis were

also expressed in the airway, while for others (e.g NK cells), a lower proportion of the classifiers were identified in the airway (figures 2B,D,F & H). Supplementary table 1a contains a complete list of all the RF-derived classifiers for different immune cell phenotypes and sub phenotypes and their respective expression levels in the infant airway.

Next, we sought to determine the functional properties of the RF classifiers and specifically, whether classifiers derived from different immune cell phenotypes represented a significant enrichment of certain biological contexts. We reasoned that the RF classifiers for a particular phenotype would be related to the functional properties of that phenotype and that when they are subjected to an independent unsupervised enrichment analysis, the top contextual hit from that analysis would be the cell phenotype from which the classifiers were initially derived. For 20 of the 33 RF-classifier lists submitted for enrichR[14,15] enrichment analysis, the biological context that was most significantly enriched was the original cell phenotype that was used to derive the classifiers in deconvolution data set (examples in figures 2 C, E, G & I). For example, the input RF-classifier list for monocytes, returned that phenotype as the most enriched phenotype (enrichment score 500, p. value <0.0001 – figure 2I).

We then undertook a more detailed characterisation the RF classifier proteins of each phenotype to determine whether the combined expression profile of phenotype-specific proteins was sufficient to distinguish one phenotype from all others. For each classifier protein, the difference between its mean expression level within the phenotype from which it was derived was calculated relative to all other phenotypes and the difference

expressed as a fold change. Figure 3A shows an example of such a comparison for classical activated monocytes (MO_classical_activated). For this phenotype, some of its classifier proteins – for example SERPINB2- were expressed at significantly higher levels in activated monocytes relative to all other cell phenotypes (figure 3A). The median expression level of SERPINB2 in classical activated monocytes for example, was $>10^4$ fold higher in these cells, relative to all other phenotypes (figure 3B). Of the 64 protein classifiers identified in the RF analysis for this phenotype, 17 were detected in the airway samples from children with pneumonia (figure 3C), with SERPINB2 being detected at relatively high levels (median 37,309 MS reporter corrected intensity). For each phenotype, we then generated a phenotype classification profile where the expression levels of all classifier proteins of a particular phenotype were plotted relative to the expression level of the same proteins in all other phenotypes. An example of this analysis is shown in figure 3D, where the combined expression levels monocyte-derived RF classifiers were compared with all other phenotypes. The expression profile of these proteins was significantly higher in monocytes compared to all other proteins ($p<0.0001$). A comprehensive analysis of the combined expression profiles of the RF markers of all other phenotypes are presented in supplementary figure 1. We used t-SNE dimensional reduction analysis to visualise phenotype segregation on the basis of RF-classification markers (figure 4A). The results of this analysis showed that the RF classifiers could clearly resolve most of the immune cell types, with phenotypes such as plasma cells (B.plasma), pDC, mDC, eosinophils, basophils, neutrophils and different monocyte sub phenotypes being clearly distinguishable from the rest of the phenotypes. Some sub-phenotypes such as central memory CD4 T cells (T4.CM) and effector memory CD4 T cells (T4.EM), could not be clearly disaggregated on the basis of RF protein expression alone.

We then applied the RF-markers to the airway proteome data of children who survived or succumbed to a severe pneumonia infection in order to determine whether there were differences in the expression of immune phenotype markers by survival status. We noted that the expression levels of eosinophil-specific RF markers were significantly overexpressed in children who survived infection, relative to non-survivors ($P<0.0001$) and well controls (figure 4B). This was in contrast to neutrophils, whose RF markers did not vary by survival status (figure 4B). The RF classifiers for other phenotypes including naïve CD8 T cells (T8.naive) were also significantly elevated in pneumonia survivors compared to children who died within 48 hours of admission ($P<0.05$) and well controls while those of phenotypes and sub phenotypes of B cells, monocytes and dendritic cells were not significantly different between children who survived or succumbed to pneumonia (figure 4C). To validate these findings, we reviewed the clinical records of >10,000 children who had been admitted to Kilifi County Hospital over a >10 year period with clinical pneumonia and for whom haematological data (including blood frequencies of eosinophils and neutrophils) had been collected at admission. These data were stratified by survival status and the difference in the frequency of eosinophils and neutrophils was determined. The results of these analyses showed that the frequency of eosinophils in children who survived pneumonia in the retrospective validation cohort (N=10,859) was significantly higher than that of children who died (N=1,604) following admission with clinical pneumonia ($p=0.0004$) – Figure 5A. On the other the frequency of blood neutrophils was no different in children who died from pneumonia and those who survived it (figure 5B).

Discussion

We report on a new method of deconvolving immune cell populations that are resident in the airways of infants and children with different survival outcomes of severe pneumonia. The respiratory response to pneumonia is characterised by the recruitment of a broad array of immune cells, which contribute to the resolution of the infection, but in some instances, have been implicated in increased pathology. The study mucosal cellular immunity during very severe pneumonia has been hindered by a number of important hurdles including the availability of low sample volumes young pneumonia patients, as well as the relatively low abundance of immune cells that may be critical in directing the clinical course of pneumonia. In this study, we addressed these problems by using an unsupervised learning approach to identify and validate protein markers that could be used to deconvolve mixed immune cells. We then applied these markers to airway proteome data obtained from children with different survival outcomes of clinical pneumonia. Our results show that protein markers associated with eosinophils are elevated in the airways of survivors and are diminished in children who later died from infection. The airway levels of these eosinophil markers were no different between children who died and well controls, indicating a possible failure to mount an appropriate eosinophil response as a potential mechanism of pneumonia-related mortality. To validate the findings of the airway proteome analysis, we reviewed the hospitalisation records of >10,000 children who had been admitted to hospital in the previous 10 years with clinical pneumonia. The results of this retrospective analysis confirmed the observations made from analysis of the airway proteome and showed that children who died from pneumonia, had significantly lower blood eosinophil counts relative to those

who survived. In contrast, neutrophil levels were not different between survivors and non survivors in both the airway proteome and in the retrospective validation cohort.

Previous studies have shown that eosinophils are activated in the airway shortly after pneumonia infection appear to contribute significantly to airway recovery. An increase in the expression of eosinophil-related markers in children with severe pneumonia has been associated with a reduced requirement for supplemental oxygen[16], suggesting that these cells are a critical component in the host's response to infecting pathogens in the airway. In contrast, studies done using post mortem tissue from children who died of viral pneumonia, show that the presence of activated monocytes within the lungs of these children[7]. Although previous studies have generally failed to characterise using flow cytometry approaches, a substantial influx of eosinophils in the respiratory tract during viral pneumonia, eosinophil-related proteins including leukotriene C4, eosinophil-derived neurotoxin (RNASE2) and eosinophil cationic protein (ECP) are expressed at high levels during viral pneumonia[17-19]. Taken together with the results of the current study, these results demonstrate a clear role for eosinophils in an effective airway response to pneumonia and further highlight the power of using phenotype markers to infer the dynamic characteristics immune cells in the airway. The results of the current study highlights that the critical role of airway eosinophils might extend beyond just a rapid resolution of infection, but that these cells may be a crucial factor for survival in children with severe pneumonia. A potential limitation of our study is that it was carried out using samples collected from the upper airway and not the lung. While lung samples would have undoubtedly been more informative of responses at the site of disease, the collection of these samples is a highly invasive process and exposes children to

substantial additional risk without providing additional diagnostic value above nasopharyngeal or peripheral blood sampling. As a result, these samples are generally only available in settings with paediatric intensive care facilities, and even here, they are generally collected in children with atypically severe disease.

In addition to eosinophils, previous studies have shown infections that cause severe pneumonia such as RSV, trigger a strong cellular response, characterised by the influx of innate immune cells. The initial response to infection is characterised by the airway recruitment of neutrophils, which express markers such as CD11b (ITGAM) and neutrophil granule proteins such as neutrophil elastase (ELANE)[20,21]. Other innate immune cells such as NK cells which expressed granzyme B are recruited in the lower airway and can be detected in in the lungs of mechanically-ventilated children with very severe pneumonia[22,23]. In addition to these cells, both myeloid and plasmacytoid dendritic cells are typically recruited into the airways of children with pneumonia in the early stages of infection[22,24] and exhibit an activated proinflammatory phenotype[22]. Adaptive immune cells including CD4+ and CD8+ T cells are present in airway samples for children with viral pneumonia[25,26]. During infection, airway frequencies of granzyme B-secreting, activated CD8 T cells is greater in the airways of children with severe viral pneumonia[23]. In the present study, we found that protein markers of different phenotypes of CD8 T cells were significantly elevated in the airways of children who survived pneumonia relative to those who died or well controls. Although the frequencies of these cells were not available for validation in the retrospective cohort, the data suggests a role for these cells in pneumonia survival. Previous studies in mechanically ventilated children showed that the frequency of lung-associated T cells increased as children recovered from infection,

suggesting that these cells are an important compoentent of effective local immunity against pneumonia and that deficits in the airways of children with pneumonia is a significant risk factor for mortality. In summary, the present data identifies a new approach for characterising the cellular immune response to pneumonia in the airway and identified critical immune populations that appear to be critical for survival. Future studies should aim to replicate these findings in samples collected from the lower respiratory tract.

Methods

Study site and population

This study recruited 80 infants and children who were admitted to Kilifi County Hospital with clinical pneumonia, defined using World Health Organisation syndromic criteria. Nasal samples were collected from each child for proteomic analysis. The microbial etiology of pneumonia was determined using both blood cultures and using a 15-target multiplex PCR panel for the detection of respiratory syncytial virus (RSV - A & B), rhinovirus, parainfluenza virus (1, 2, 3 & 4) adenovirus, influenza (A, B & C), coronavirus (OC43 & e229), human metapneumovirus and *Mycoplasma pneumoniae*. Children with clinical pneumonia signs and a positive diagnostic result from any of these tests were included in the analysis. Children were stratified into the survival group if they were alive for at least 365 days after discharge(n=61) while the mortality group comprised of children who died within 72 hours of admission (n=19). In addition to these groups, we recruited 10 age-matched well controls as a comparator group. Written informed consent was sought from the parents and legal guardians of all children who were sampled in this study. Ethical approval for the conduct of this study was granted by the Kenya Medical Research Institute's Scientific and ethical research unit (SERU). All study procedures were conducted in accordance with Good Clinical Laboratory Practise (GCLP) standards.

Analysis of airway proteomes using mass spectrometry

Naso- and oropharyngeal swab samples were centrifuged at 17,000xg for 10 mins at 4°C to obtain cell pellets which were washed once using PBS and lysed by bead-vortexing for 10 minutes in cell lysis buffer (RLT, Qiagen, Germany). Proteins (as well as DNA and RNA)

were then extracted from the lysate using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen, Germany) following manufacturers instructions. The concentration of total protein obtained was determined using the Bradford assay (Bio-Rad, USA). Thirty micrograms (30µg) of total protein from each sample was then reduced with 10mM tris(2-carboxyethyl)phosphine (TCEP, Sigma-Aldrich, USA) at 55°C for 1h and subsequently alkylated with 18mM IAA (Sigma-Aldrich, USA) for 30 minutes at room temperature, while keeping the reaction protected from light. Proteins were precipitated overnight at -20°C with six volumes of pre-chilled (-20°C) acetone (Sigma-Aldrich, USA). The samples were centrifuged at 8,000xg for 10 minutes at 4°C to obtain the protein pellets and supernatants were discarded. The protein pellet was resuspended in 100µl of 50mM Triethylammonium bicarbonate (TEAB, Sigma-Aldrich, USA). Trypsin (Sigma-Aldrich, USA) was added to the protein samples at a trypsin-protein sample ratio of 1:10 and protein digestion was allowed to proceed overnight at 37°C with shaking. The peptide samples were randomly assigned to 10 individual batches: each containing nine patient samples and one pooled control sample. The pooled control sample consisted of a pool of peptides from all patient samples. The peptide samples derived from individual patients were then individually labelled using the TMT10plex mass tag kit (Thermo scientific, USA) according to manufacturer's instructions, with one isobaric tag being exclusively used to label the pooled control sample. The labelled peptides for each 10plex were subsequently combined to generate 10 individual pools. The labelled peptide pools were desalted using P10 C18 pipette ZipTips (Millipore, USA) according to the manufacturer's instructions. Eluted peptides were dried in a Speedvac concentrator (Thermo Scientific, USA). Peptides (8 µl) were loaded using a Dionex Ultimate 3000 nano-flow ultra-high-pressure liquid chromatography system (Thermo Scientific, USA) on to a

75µm x 2 cm C18 trap column (Thermo Scientific, USA) and separated on a 75µm x 50 cm C18 reverse-phase analytical column (Thermo Scientific) at heated at 40°C. For LFQ protein quantification; elution was carried out with mobile phase B (80% acetonitrile with 0.1% formic acid) gradient (4 to 30%) over 310 min at a flow rate of 0.25 µl/min. Each LC run was finished by washout with 98% B for 10 min and re-equilibration in 2% B for 30 min. Five blanks of 40 min each were run on the column between each injection comprising of two wash cycles with 90% B and an equilibration phase of 15 min to avoid sample carryover. Peptides were measured using a Q Exactive Orbitrap mass spectrometer (Thermo Scientific, USA) coupled to the chromatography system via a nano-electrospray ion source (Thermo Scientific). On the Q Exactive , the ms^1 settings for peptides were: Resolution, 70000; AGC target, 3e6; maximum IT, 120 ms; scan range, 400-1800 m/z; while the ms^2 settings for fragmentation spectra of peptides were: Resolution, 17000 (35000 for labelled peptides); AGC target, 5e4; maximum IT, 120 ms ; isolation window, 1.6 m/z. MS data were acquired by data dependent acquisition where the top 12 (15 for labelled peptides) most intense precursor ions in positive mode were selected for ms^2 Higher-energy C-trap dissociation fragmentation which were subsequently excluded for the next 45 s following fragmentation event. Charge exclusion was set to ignore peptide spectrum matches that were unassigned, singly charged, and those with ≥+8 charges. Raw mass spectrometer files were analysed by MaxQuant software version 1.6.0.1. by searching against the human Uniprot FASTA database (downloaded February 2014) using the Andromeda search engine..

Airway samples were centrifuged at 17,000xg for 10 mins at 4°C to obtain cell pellets which were washed once using PBS and lysed by bead-vortexing for 10 minutes in cell lysis

buffer (RLT, Qiagen, Germany). Proteins (as well as DNA and RNA) were then extracted from the lysate using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen, Germany) following manufacturers instructions. The concentration of total protein obtained was determined using the Bradford assay (Bio-Rad, USA). Thirty micrograms (30µg) of total protein from each sample was then reduced with 10mM tris(2-carboxyethyl)phosphine (TCEP, Sigma-Aldrich, USA) at 55°C for 1h and subsequently alkylated with 18mM IAA (Sigma-Aldrich, USA) for 30 minutes at room temperature, while keeping the reaction protected from light. Proteins were precipitated overnight at -20°C with six volumes of pre-chilled (-20°C) acetone (Sigma-Aldrich, USA). The samples were centrifuged at 8,000xg for 10 minutes at 4°C to obtain the protein pellets and supernatants were discarded. The protein pellet was resuspended in 100µl of 50mM Triethylammonium bicarbonate (TEAB, Sigma-Aldrich, USA). Trypsin (Sigma-Aldrich, USA) was added to the protein samples at a trypsin-protein sample ratio of 1:10 and protein digestion was allowed to proceed overnight at 37°C with shaking. The peptide samples were randomly assigned to 10 individual batches: each containing nine patient samples and one pooled control sample. The pooled control sample consisted of a pool of peptides from all patient samples. The peptide samples derived from individual patients were then individually labelled using the TMT10plex mass tag kit (Thermo scientific, USA) according to manufacturer's instructions, with one isobaric tag being exclusively used to label the pooled control sample. The labelled peptides for each 10plex were subsequently combined to generate 10 individual pools. The labelled peptide pools were desalted using P10 C18 pipette ZipTips (Millipore, USA) according to the manufacturer's instructions. Eluted peptides were dried in a Speedvac concentrator (Thermo Scientific, USA). Peptides (8 µl) were loaded using a Dionex Ultimate 3000 nano-flow ultra-high-pressure liquid

chromatography system (Thermo Scientific, USA) on to a 75µm x 2 cm C18 trap column (Thermo Scientific, USA) and separated on a 75µm x 50 cm C18 reverse-phase analytical column (Thermo Scientific) at heated at 40°C. For LFQ protein quantification; elution was carried out with mobile phase B (80% acetonitrile with 0.1% formic acid) gradient (4 to 30%) over 310 min at a flow rate of 0.25 µl/min. Each LC run was finished by washout with 98% B for 10 min and re-equilibration in 2% B for 30 min. Five blanks of 40 min each were run on the column between each injection comprising of two wash cycles with 90% B and an equilibration phase of 15 min to avoid sample carryover. Peptides were measured using a Q Exactive Orbitrap mass spectrometer (Thermo Scientific, USA) coupled to the chromatography system via a nano-electrospray ion source (Thermo Scientific). On the Q Exactive , the ms^1 settings for peptides were: Resolution, 70000; AGC target, 3e6; maximum IT, 120 ms; scan range, 400-1800 m/z; while the ms^2 settings for fragmentation spectra of peptides were: Resolution, 17000 (35000 for labelled peptides); AGC target, 5e4; maximum IT, 120 ms ; isolation window, 1.6 m/z. MS data were acquired by data dependent acquisition where the top 12 (15 for labelled peptides) most intense precursor ions in positive mode were selected for ms^2 Higher-energy C-trap dissociation fragmentation which were subsequently excluded for the next 45 s following fragmentation event. Charge exclusion was set to ignore peptide spectrum matches that were unassigned, singly charged, and those with ≥+8 charges. Raw mass spectrometer files were analysed by MaxQuant software version 1.6.0.1. by searching against the human Uniprot FASTA database (downloaded February 2014) using the Andromeda search engine.

Analysis of airway-resident immune cells using flow cytometry

1ml of nasopharyngeal and oropharyngeal swab samples obtained from children was centrifuged at 17,000xg for 7 minutes, after which 800µl of the supernatant was removed and discarded. The remaining 200µl were split into two aliquots of 100µl each. The first aliquot was used for neutrophil phenotyping assays and the other was used for neutrophil phagocytosis assays. 20µl of a pre-constituted cocktail of the following antibodies (from ThermoFisher) was used to label both aliquots – CD45, CD16, CD14, CD3, CD19, HLA-DR, CD66b, CD11b and a Live-dead marker. With the exception of the live/dead marker, all other antibodies were diluted 1:100 in FACS buffer. The live/dead marker was prepared at a 1:1000 dilution in FACS buffer. For the phagocytosis assay tube, 20µl of opsonised *Escherichia coli* (*E.coli*) was added to the tube (pHrodo Red E. coli BioParticles; ThermoFisher). The bacteria was initially prepared by mixing the *E.coli* strain with new-born calf sera followed by a 30-minute incubation at 37°C. After this step, both tubes were incubated at 37°C for 35 minutes. After the incubation, 20µl of a live-dead marker was added to each tube and incubated for 10 minutes at 37°C. The reaction was stopped by adding 500µl of 1X RBC lysis buffer followed by a 5-minute incubation. Cells in each tube were then spun down at 2,700xg for 1 minute and the supernatant discarded. Cells were then washed twice with FACS buffer, after which 350µl of FACS flow was added. Cells were then analysed immediately on a BD LSR Fortessa instrument. The following gating strategy was used to detect airway-resident neutrophils: Debris were excluded on the basis of their forward (FSC-A) and side (SSC-A) scatter characteristics, doublets were excluded using FSC-A versus FSC-H and dead cells were excluded using the live-dead marker. Data was analysed using FlowJo software.

Statistical data analysis

Data were analyzed in R. deconvolution analysis by random forest classification was done using the Boruta package, with a maximum of 300 runs. Protein classifiers that whose mean expression in the test phenotype was significantly greater than alternative phenotypes were taken forward for further analysis. T-SNE analysis was carried out using the Rtsne, with the iterations parameter set to a maximum of 300 and a perplexity value of 30. Analysis was carried out in two dimensional space. Cell type enrichment analysis was done one the enrichR platform. The input search term used for enrichment analysis was the RF protein classifier lists derived from random forest classification. All pairwise comparisons between the expression level of RF classifiers between phenotypes and frequency of immune cell types between survival states was done using t-tests on log10 normalised data. The deconvolution data set that was used to identify phenotype-specific protein classifiers was obtained from a previously published paper by Rieckmann et al.[13]

Data availability

The proteomics data reported in this paper are available at the ProteomeXchange Consortium database (Accession numbers: PXD009403 & JAMES) .

Acknowledgements

This study was supported by fellowship funding to CJS from the Wellcome Trust (WT105882MA). The funder played no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript

Conflicts of interest

The authors declare that they have no conflicting interests

Author contributions

CJS designed the study

CJS, JMN, MNM, ETG, AG conducted the experiments

CJS,NKK analysed the data

CJS wrote the manuscript

All authors reviewed and approved the manuscript

Figures

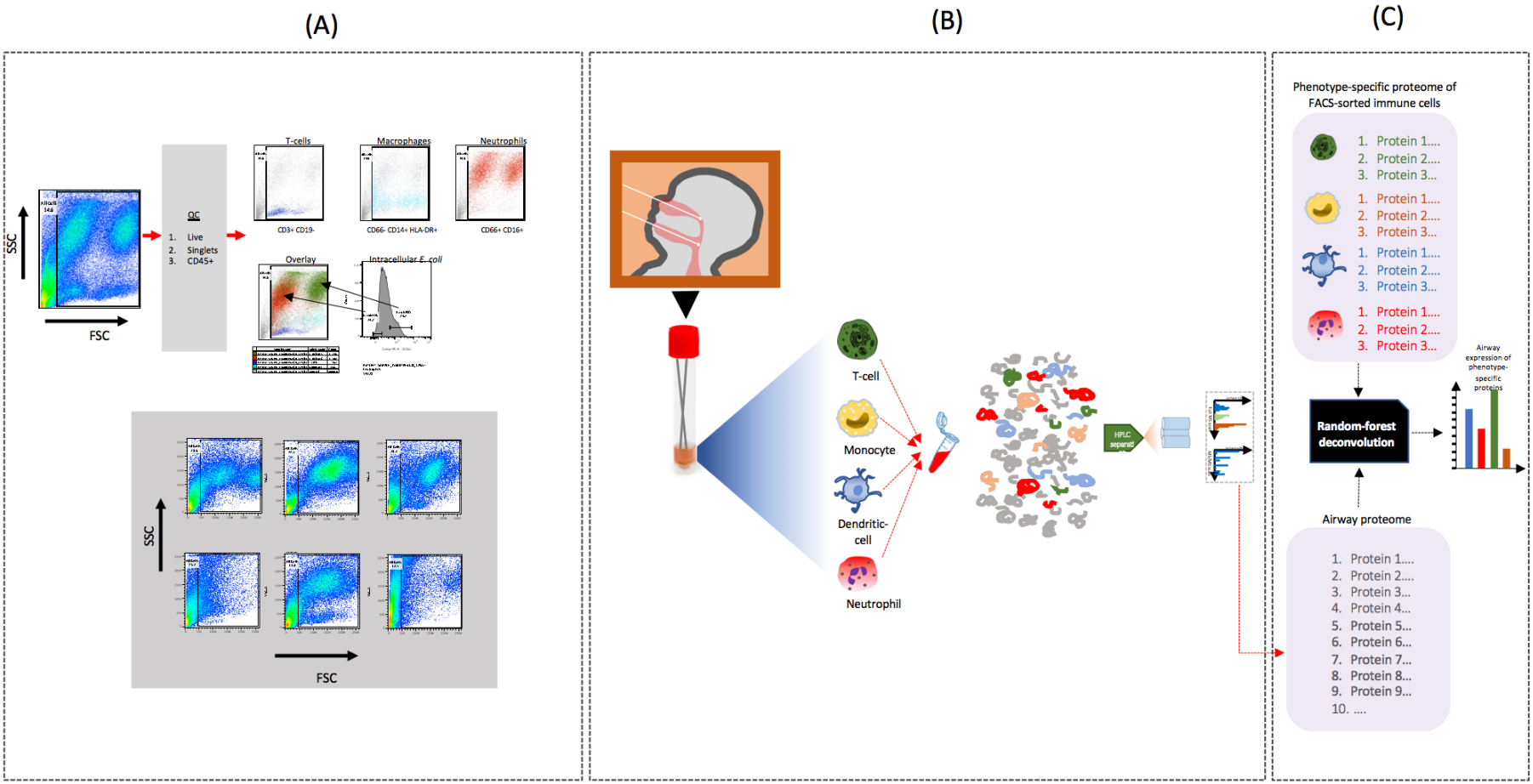
(A)
(B)
(C)
SSC
FSC
QC
1. Live
2. Singlets
3. CD45+
T-cells
CD3+ CD19-
Macrophages
CD66- CD14+ HLA-DR+
Neutrophils
CD66+ CD16+
Overlay
Intracellular E. coli
T-cell
Monocyte
Dendritic-cell
Neutrophil
HPLC separation
Phenotype-specific proteome of FACS-sorted immune cells
1. Protein 1....
2. Protein 2....
3. Protein 3...
Random-forest deconvolution
Airway expression of phenotype-specific proteins
Airway proteome
1. Protein 1....
2. Protein 2....
3. Protein 3...
4. Protein 4...
5. Protein 5...
6. Protein 6...
7. Protein 7...
8. Protein 8...
9. Protein 9...
10. ....

Figure 1: Analysis of the profile of airway-resident immune cells using flow cytometry and in-silico deconvolution. **(A)** Cells were obtained from naso- and oropharyngeal flocked swabs and processed by flow cytometry. Here the gating strategy is shown where initial gating for live cells, singlets and CD45 expression was carried out, followed by staining with fluorescent antibodies that were specific for different cell surface markers. CD3+ and CD19- cells were gated as T-cells, CD66-, CD14+ and HLA-DR- cells were gated as macrophages/monocytes while neutrophils were gated on the basis of CD66 and CD16 positivity. The functional activity of neutrophils was characterized by co-culturing the cells with E.coli particles that were labelled with pHRhodo. Neutrophils that contained phagocytosed particles are shown in green in the overlay dot plot. The bottom represented FSC and SSC plots of samples obtained from 6 different infants and demonstrates a high level of diversity in the immune cell populations present in the airways of children with pneumonia. **(B)** In silico deconvolution analysis was based on high resolution mass spectrometry data obtained from the airway. Cells were obtained from the naso and oropharyngeal sites using separate swabs, which were both eluted in a common transport media. As shown in A above, these samples contained a broad diversity of immune cells including T-cells, monocytes/macrophages and neutrophils. These samples were isolated from the sample by centrifugation and processed for mass spectrometry analysis using a standard protocol (see methods) and were subsequently analysed on a Q Exactive Orbitrap mass spectrometer. **(C)** In silico deconvolution of airway proteomes was conducted initially by identifying phenotype specific protein markers in a deconvolution data set using random forest classification. These markers

were then applied on the airway proteome data set, and used to infer relative frequencies immune cell phenotypes in the airways of infants with pneumonia.

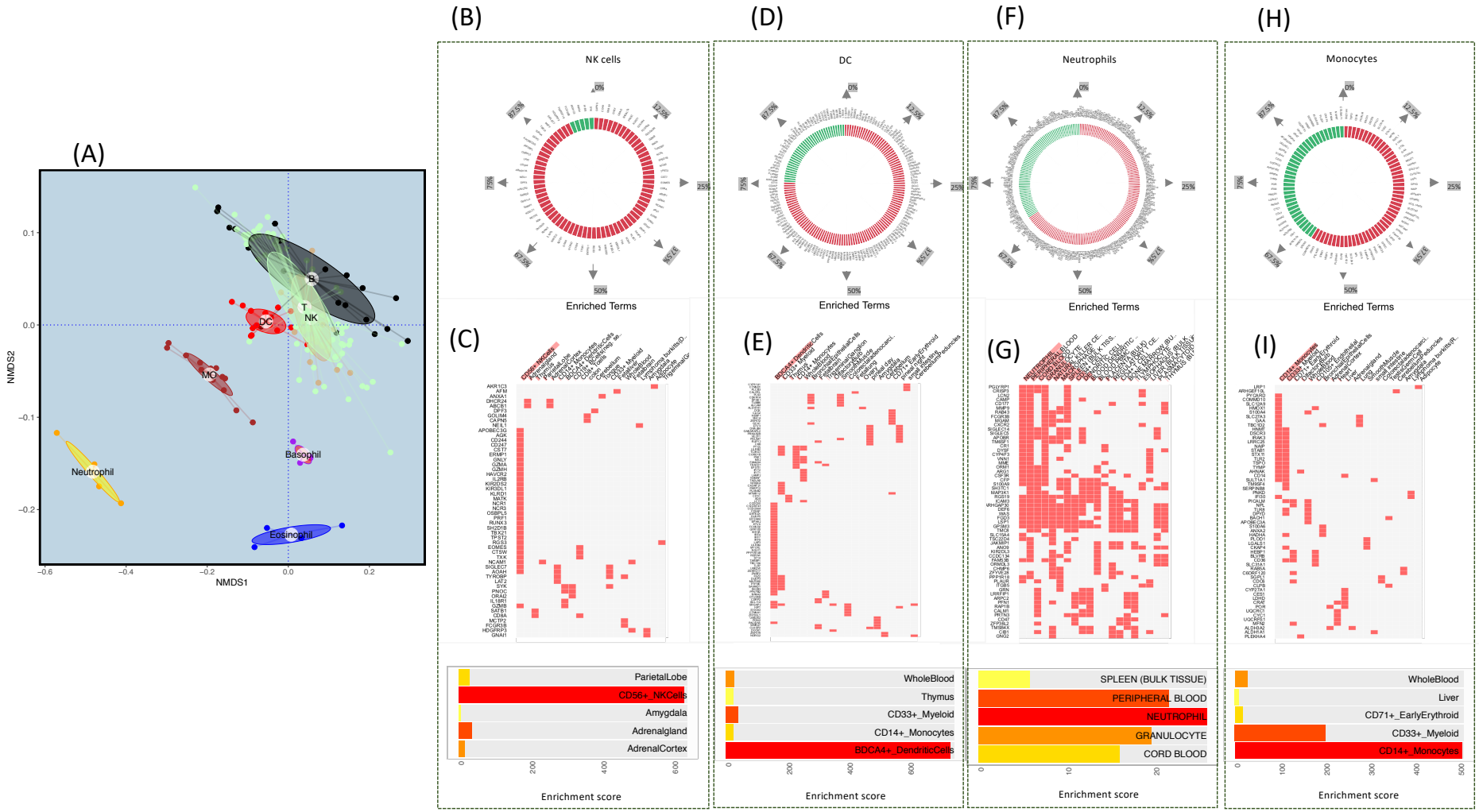
(A)
NMDS1
NMDS2
B
T
NK
DC
MO
Basophil
Neutrophil
Eosinophil
(B)
NK cells
Enriched Terms
ParietalLobe
CD56+_NKCells
Amygdala
Adrenalgland
AdrenalCortex
Enrichment score
(C)
(D)
DC
Enriched Terms
WholeBlood
Thymus
CD33+_Myeloid
CD14+_Monocytes
BDCA4+_DendriticCells
Enrichment score
(E)
(F)
Neutrophils
Enriched Terms
SPLEEN (BULK TISSUE)
PERIPHERAL BLOOD
NEUTROPHIL
GRANULOCYTE
CORD BLOOD
Enrichment score
(G)
(H)
Monocytes
Enriched Terms
WholeBlood
Liver
CD71+_EarlyErythroid
CD33+_Myeloid
CD14+_Monocytes
Enrichment score
(I)

Figure 2: Use of proteomics data to segregate immune cell phenotypes. **(A)** A nonmetric multidimensional scaling (NMDS) plot was used to visualize the separation of immune cell phenotypes at the to hierarchical level based on protein expression data derived from the deconvolution data set. We observed that expression profile of certain cell types including neutrophils, monocytes and basophils, resulted in a clear separation from other immune cell types. However others, such as B cells, T cells and NK cells did not exhibit clear partitioning on the 2D MDS space. To achieve better phenotype partitioning at the phenotype and sub-phenotype level, we used random forest (RF) classification to derive protein markers that could be used to distinguish specific phenotype (e.g. DCs) or sub-phenotype (pDCs or mDCs) from all others. We then examined the airway proteome data to determine whether these RF classifiers were detectable in the airway. An example of the results is shown in **(B,D,F & H)**, where the RF classifiers for NK cells, DCs, Neutrophils and Monocytes are shown labelled on the circular tile plot. The green tiles are the RF classifiers that were detected in the airway proteomes of children. We used enrichment analysis to determine whether respective RF-phenotype classifier lists represented a significant enrichment of certain cell phenotypes using an independent phenotype enrichment analysis on the enrichR platform. The clustergrams shown in **(C,E,G & I)** show the extent to which different immune cell phenotypes were statistically presented within RF classifier lists. In all examples shown here, the most highly enriched phenotype in the RF classifer list (first column of the clustergram) matched the phenotype that was used to derive the RF classifiers from the deconvolution data set. Each row on the clustergram represents a protein from the RF-classifier list, while each colored

box represents the fact that the protein is known to be expressed in the cell types listed on the columns. Horizonal panels at the bottom indicate enrichment scores for different RF classifiers.

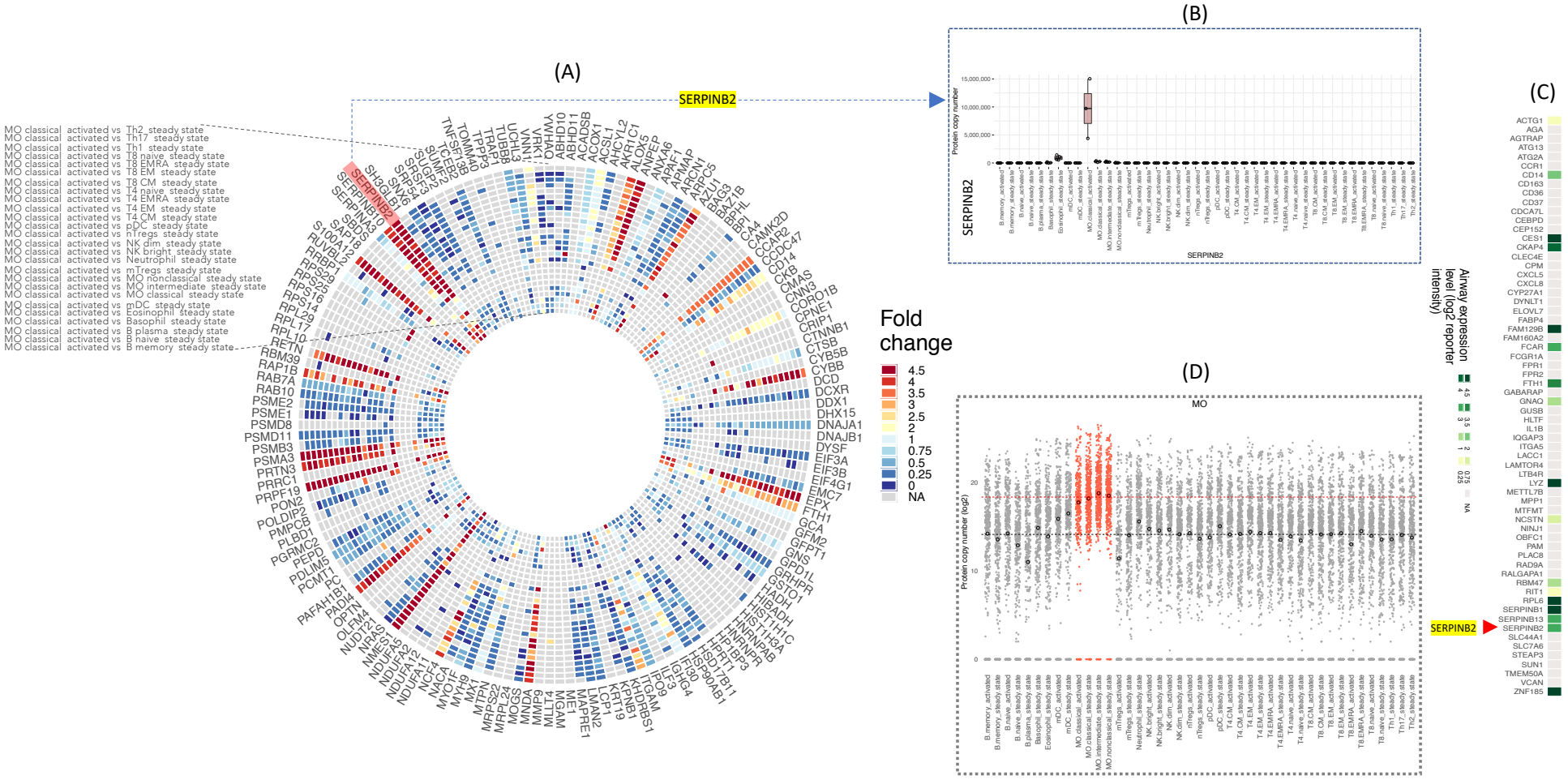
(A)
SERPINB2
MO classical activated vs Th2 steady state
MO classical activated vs Th17 steady state
MO classical activated vs Th1 steady state
MO classical activated vs T8 naive steady state
MO classical activated vs T8 EMRA steady state
MO classical activated vs T8 EM steady state
MO classical activated vs T8 CM steady state
MO classical activated vs T4 naive steady state
MO classical activated vs T4 EMRA steady state
MO classical activated vs T4 EM steady state
MO classical activated vs T4 CM steady state
MO classical activated vs pDC steady state
MO classical activated vs nTregs steady state
MO classical activated vs NK dim steady state
MO classical activated vs NK bright steady state
MO classical activated vs Neutrophil steady state
MO classical activated vs mTregs steady state
MO classical activated vs MO nonclassical steady state
MO classical activated vs MO intermediate steady state
MO classical activated vs MO classical steady state
MO classical activated vs mDC steady state
MO classical activated vs Eosinophil steady state
MO classical activated vs Basophil steady state
MO classical activated vs B plasma steady state
MO classical activated vs B naive steady state
MO classical activated vs B memory steady state
Fold change
4.5
4
3.5
3
2.5
2
1
0.75
0.5
0.25
0
NA
(B)
SERPINB2
Protein copy number
(C)
Airway expression level (log2 reporter intensity)
SERPINB2
(D)
MO
Protein copy number (log2)

Figure 3: Detailed analysis of the phenotype-specific RF classifiers identified from the deconvolution data set. **(A)** We used a circular heat map to visualize differential expression of RF classifier proteins in a selection of immune cell phenotypes. In this example, the relative expression levels of a subset of proteins on the wheels circumference were compared between classical activated monocytes and selected immune cell phenotypes. Each spoke of the wheel represents a single protein, and each segment on a spoke represents the difference (expressed as $\log_{10}$ fold change) between the expression level of the protein in classical activated monocytes and the listed cell types. An increase in the intensity of the red hue indicates greater expression in classical actibvated monocytes, relative to the comparator phenotype. **(B)** An example of the cross-phenotype expression of a classical activated monocyte-specific RF classifier (SERPINB2) is shown. SERPINB2 was expressed in this phenotype at significantly greater levels compared to alternative immune cell phenotypes. **(C)** The airway proteome data from paediatric pneumonia admissions was examined to determine whether any of the RF classifiers for classical activated monocytes were also detected in the airways of children. Each segment represents a single RF classifier derived from the deconvolution data set, while the RF classifiers that were also detected in the airway are depicted in a colored hue. An increase in the intensity of the green hue, corresponds to the median expression level of a particular protein the airway proteome data. SERPINB2 (red arrow), was expressed at relatively high levels in the infant airway. **(D)** The expression level of all monocyte-specific RF classifiers was plotted alongside the corresponding expression level of the same proteins in alternative immune cell phenotypes. Each dark black open circle denotes the median expression level of the RFs in particular phenotypes. The dashed red line indicates the median expression level of the RF classifier

proteins in monocytes sub-phenotypes, while the dashed black lines indicate the corresponding median expression of these protein in non-monocyte phenotypes. The expression of monocyte derived RF-classifiers was significantly higher in monocytes relative to other cell phenotypes. A detailed map of a similar comparison for the RF classifiers of all other phenotypes are shown in supplementary figure xxxx.

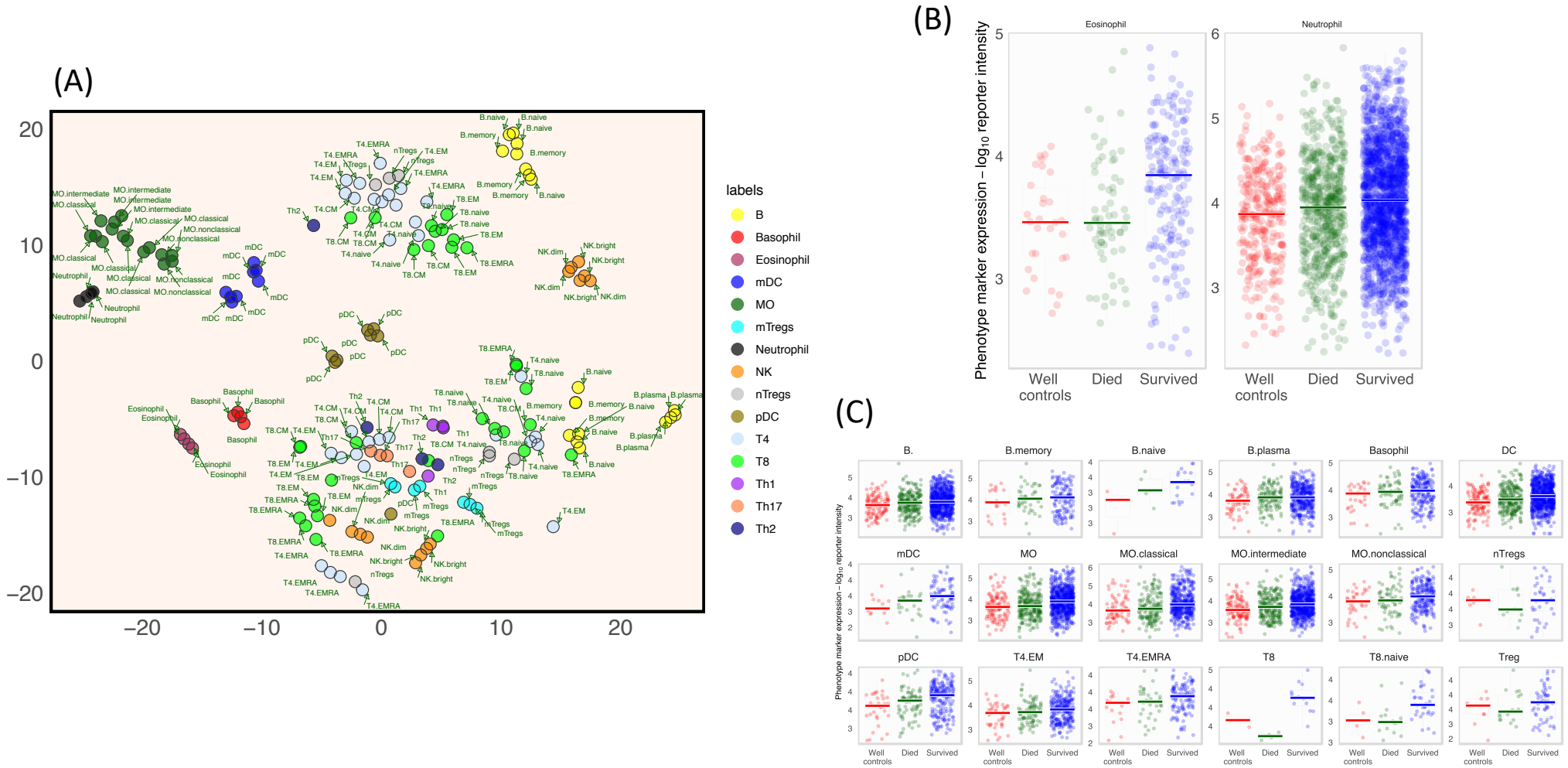

Figure 4: (A) Performance of all RF classifiers in segregating different immune phenotypes and sub phenotypes was visualized using a t-SNE plot. The markers are colored according by the major phenotypes and the labels indicate sub phenotypes. The RF classifiers were

able to clearly separate different sub phenotypes in the 2D t-SNE space. Neutrophils, eosinophils, monocytes, NK cells, dendritic cells, basophils among others were clearly separated in the t-SNE plot from all other phenotypes. There was less clear segregation for T-cell sub phenotypes, including CD8, CD4 and regulatory T-cell phenotypes. (B) the expression of eosinophil- and neutrophil-specific RF classifiers was compared between children with different survival outcomes of pneumonia as well as in well controls. Eosinophil RF markers were expressed at significantly higher levels in survivors compared with children who died and well controls. In contrast, there was no difference in the expression of neutrophil RF markers between these groups. (C) RF classifiers of other immune cell phenotypes were also compared by survival status. No significant difference in the RF expression of most phenotypes by survival status, with the exception of CD8 (T8) phenotypes and regulatory T cell (Treg) phenotypes.

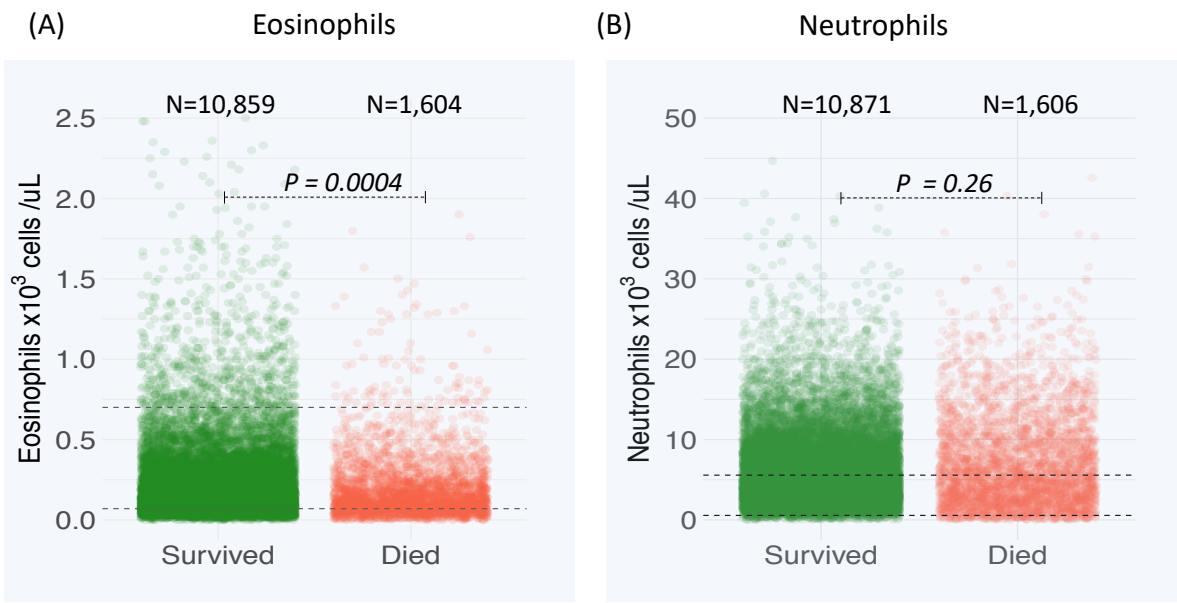

Figure 5: Validation of eosinophil association with survival in a large retrospective pneumonia cohort. (A) The frequency of eosinophils in the blood of a retrospective of >10,000 children with different survival outcocomes after admission. Eosinophil levels were significantly

elevated in paediatric pneumonia survivors relative to that of children who died. (B) In contrast, there was no significant difference in the blood neutrophil levels between survivors and non survivors.

References

1 Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *Lancet (London, England)* **394**, 757-779, doi:10.1016/s0140-6736(19)30721-4 (2019).

2 Shi, T. *et al.* Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet (London, England)*, doi:10.1016/s0140-6736(17)30938-8 (2017).

3 McNally, L. M. *et al.* Effect of age, polymicrobial disease, and maternal HIV status on treatment response and cause of severe pneumonia in South African children: a prospective descriptive study. *Lancet (London, England)* **369**, 1440-1451, doi:10.1016/s0140-6736(07)60670-9 (2007).

4 West, T. E., Goetghebuer, T., Milligan, P., Mulholland, E. K. & Weber, M. W. Long-term morbidity and mortality following hypoxaemic lower respiratory tract infection in Gambian children. *Bulletin of the World Health Organization* **77**, 144-148 (1999).

5 Ngari, M. M. *et al.* Mortality after Inpatient Treatment for Severe Pneumonia in Children: a Cohort Study. *Paediatric and perinatal epidemiology* **31**, 233-242, doi:10.1111/ppe.12348 (2017).

6 Mejias, A. *et al.* Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. *PLoS medicine* **10**, e1001549, doi:10.1371/journal.pmed.1001549 (2013).

7 Johnson, J. E., Gonzales, R. A., Olson, S. J., Wright, P. F. & Graham, B. S. The histopathology of fatal untreated human respiratory syncytial virus infection. *Mod Pathol* **20**, 108-119, doi:10.1038/modpathol.3800725 (2007).

8 Jochems, S. P. *et al.* Novel Analysis of Immune Cells from Nasal Microbiopsy Demonstrates Reliable, Reproducible Data for Immune Populations, and Superior Cytokine Detection Compared to Nasal Wash. *PloS one* **12**, e0169805, doi:10.1371/journal.pone.0169805 (2017).

9 Connors, T. J. *et al.* Developmental Regulation of Effector and Resident Memory T Cell Generation during Pediatric Viral Respiratory Tract Infection. *Journal of immunology (Baltimore, Md. : 1950)* **201**, 432-439, doi:10.4049/jimmunol.1800396 (2018).

10 Sande, C. J. *et al.* Untargeted analysis of the airway proteomes of children with respiratory infections using mass spectrometry based proteomics. *Scientific reports* **8**, 13814, doi:10.1038/s41598-018-32072-3 (2018).

11 Sande, C. J. *et al.* Airway response to respiratory syncytial virus has incidental antibacterial effects. *Nature communications* **10**, 2218, doi:10.1038/s41467-019-10222-z (2019).

12 World Health Organisation. Pocket book of hospital care for children. 80-81 (WHO, Geneva, 2013).

13 Rieckmann, J. C. *et al.* Social network architecture of human immune cells unveiled by quantitative proteomics. *Nature immunology* **18**, 583-593, doi:10.1038/ni.3693 (2017).

14 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).

15 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).

16 Lindemans, C. A. *et al.* Systemic eosinophil response induced by respiratory syncytial virus. *Clinical and experimental immunology* **144**, 409-417, doi:10.1111/j.1365-2249.2006.03084.x (2006).

17 Kim, H. H., Lee, M. H. & Lee, J. S. Eosinophil cationic protein and chemokines in nasopharyngeal secretions of infants with respiratory syncytial virus (RSV) bronchiolitis and non-RSV bronchiolitis. *Journal of Korean medical science* **22**, 37-42, doi:10.3346/jkms.2007.22.1.37 (2007).

18 Harrison, A. M., Bonville, C. A., Rosenberg, H. F. & Domachowske, J. B. Respiratory syncytical virus-induced chemokine expression in the lower airways: eosinophil recruitment and degranulation. *American journal of respiratory and critical care medicine* **159**, 1918-1924, doi:10.1164/ajrccm.159.6.9805083 (1999).

19 Dimova-Yaneva, D., Russell, D., Main, M., Brooker, R. J. & Helms, P. J. Eosinophil activation and cysteinyl leukotriene production in infants with respiratory syncytial virus bronchiolitis. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **34**, 555-558, doi:10.1111/j.1365-2222.2004.1918.x (2004).

20 Wang, S. Z. *et al.* Shedding of L-selectin and PECAM-1 and upregulation of Mac-1 and ICAM-1 on neutrophils in RSV bronchiolitis. *The American journal of physiology* **275**, L983-989, doi:10.1152/ajplung.1998.275.5.L983 (1998).

21 Wang, S. Z. *et al.* The apoptosis of neutrophils is accelerated in respiratory syncytial virus (RSV)-induced bronchiolitis. *Clinical and experimental immunology* **114**, 49-54, doi:10.1046/j.1365-2249.1998.00681.x (1998).

22 Kerrin, A. *et al.* Differential lower airway dendritic cell patterns may reveal distinct endotypes of RSV bronchiolitis. *Thorax* **72**, 620-627, doi:10.1136/thoraxjnl-2015-207358 (2017).

23 Bem, R. A. *et al.* Activation of the granzyme pathway in children with severe respiratory syncytial virus infection. *Pediatric research* **63**, 650-655, doi:10.1203/PDR.0b013e31816fdc32 (2008).

24 Gill, M. A. *et al.* Mobilization of plasmacytoid and myeloid dendritic cells to mucosal sites in children with respiratory syncytial virus and other viral respiratory infections. *The Journal of infectious diseases* **191**, 1105-1115, doi:10.1086/428589 (2005).

25 Heidema, J. *et al.* CD8+ T cell responses in bronchoalveolar lavage fluid and peripheral blood mononuclear cells of infants with severe primary respiratory syncytial virus infections. *Journal of immunology (Baltimore, Md. : 1950)* **179**, 8410-8417 (2007).

26 Everard, M. L. *et al.* Analysis of cells obtained by bronchial lavage of infants with respiratory syncytial virus infection. *Archives of disease in childhood* **71**, 428-432 (1994).