# Combining the host transcriptome and molecular pathogen diagnostics to better understand febrile illness in children

Thesis submitted for the degree of Doctor of Philosophy

Dr Priyen Shah

Supervisors: Dr Jethro Herberg & Dr Myrsini Kaforou

Section of Paediatric Infectious Diseases
Department of Infectious Diseases
IMPERIAL COLLEGE LONDON

September 2022

# Statement of Originality

All work presented in this thesis, unless otherwise stated and referenced is my own. Any work performed by other is clearly declared and referenced where appropriate.

## Copyright Declaration

‘The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.’

# Acknowledgements

During my PhD I was fortunate to receive the support from a number of exceptional people. First and foremost, I would like to acknowledge the huge impact of my primary supervisor, Dr Jethro Herberg. I will always be immensely grateful that Jethro took the risk of appointing me as a clinical research fellow, despite my limited research background. Since then he has supported me through my early research projects, my PhD application and in the transition to returning to clinical practice. I am eternally thankful for his efforts and encouragement that have allowed me to work in this fantastic department and on this exciting project, which has completely changed my career trajectory.

I would also like to acknowledge Dr Myrsini Kaforou, my secondary supervisor of this PhD. Myrsini somehow managed to hide her intimidating knowledge of bioinformatics behind an approachable and welcoming smile and I have learnt a great deal from her. She has helped me to develop from having very basic computer science skills to a confident bioinformatician with a passion for learning and developing new techniques.

During my time at Imperial, I am grateful to have met a number of inspiring colleagues and friends. I would like to thank everyone in the Mike Levin Group for being so welcoming and helpful throughout my PhD. It has truly been a wonderful team to work with. In particular, I would like to acknowledge Heather Jackson, who has always been a willing sounding board for ideas and was always willing to help diagnose issues with my code as I was getting to grips with using R.

I would also like to acknowledge the PERFORM and DIAMONDS consortia and the team at Micropatholgy Ltd, who I got to know throughout my time at Imperial College. I always enjoyed the stimulating discussions we had, and left our meetings with more ideas and

excitement! In particular I wanted to thank Sobhia Mustafa, Sarah Darnell at Imperial College London and Marie Voice at Micropathology Ltd with whom I worked closely throughout my PhD.

Finally, I am enormously grateful for the unconditional love and support of my family and friends. I am thankful to my wife Reina and my dear children Saayan and Arahya who have been incredibly patient and been my partners on this journey. I am thankful to my parents, who have always nurtured my love for science and exploration of the truth.

## Abstract

Background: The significant morbidity and mortality associated with infectious diseases if untreated and the rise of antimicrobial resistance has prompted the development of novel diagnostics for infectious diseases. These diagnostics have performed well at identifying patients with confident diagnosis of bacterial or viral infection, but perform less well when the aetiology is unclear. The introduction of molecular diagnostics has shown there is more pathogen data than what is revealed by conventional diagnostics. In this study I explore whether molecular diagnostic data, when analysed in conjunction with gene expression data can help us to understand febrile illnesses more clearly.

Methods: I analysed molecular diagnostic pathogen data from 4611 febrile children and 1061 controls recruited to the PERFORM study, an EU funded study recruiting patients with fever or suspected infection across Europe and supplemented this with host gene expression data from a 12-gene signature.

Results: Pathogen detection did not always align with the patient's clinical course, with bacteria detected in patients presumed to have viral disease and controls and viruses detected in patients presumed to have viral disease and controls. In addition, we found that pathogen co-detection was common and non-random.

Gene expression analysis of a 12-gene signature showed that the detection of some pathogens and syndromes resulted in a "bacterial-like" response, others resulted in a "viral-like" response, while many resulted in a mixed picture, with both bacterial and viral-like

responses in different patients. The host response was also influenced by site of infection and the presence of co-detections.

Conclusions: This study shows the utility of gene expression studies in understanding infections. Further analysis of additional pathogen data, and expression data from a wider set of genes at multiple timepoints, will allow for better understanding of when pathogen detection requires intervention. This could form the foundation for novel diagnostics for infectious diseases.

# Table of Contents

**STATEMENT OF ORIGINALITY** ..... **1**

**COPYRIGHT DECLARATION** ..... **2**

**ACKNOWLEDGEMENTS** ..... **3**

**ABSTRACT** ..... **5**

**LIST OF ABBREVIATIONS** ..... **10**

**LIST OF TABLES** ..... **11**

**LIST OF FIGURES** ..... **13**

**1. INTRODUCTION** ..... **20**

1.1 Viral Infections ..... 20

1.2 Bacterial Infections ..... 21

1.3 Diagnosis of Infectious Disease ..... 22

1.4 Impact of the Overuse of Antimicrobials ..... 23

1.5 Molecular Diagnostic Approaches ..... 24

1.6 Co-Infections ..... 25

1.7 Novel Diagnostic Techniques ..... 26

*1.7.1 Transcriptomics ..... 27*

*1.7.2 Protein Biomarkers ..... 28*

**1.8 PROJECT AIMS** ..... **30**

**2. METHODS** ..... **31**

2.1 Patient Recruitment ..... 31

2.2 Clinical Assignment ..... 33

2.4 Molecular Diagnostic Testing ..... 36

2.5 Statistical Methods ..... 38

**3. PATHOGENS DETECTED IN EUROPEAN FEBRILE CHILDREN USING MOLECULAR DIAGNOSTICS** ..... **38**

3.1 Introduction ..... 38
*3.1.1 Aims ..... 38*
3.2 Methods ..... 39
3.3 Results ..... 40
*3.3.1 Patient Recruitment ..... 40*
*3.3.2 Pathogens detected by local investigations and by centralized molecular pathogen tests ..... 44*
*3.3.3 CMT in febrile cases vs controls ..... 45*
*3.3.4 CMT in Definite Bacterial vs Definite Viral groups ..... 50*
*3.3.4 Quantitative CMT ..... 53*
***3.3.5 Combination of CMT and local diagnostics ..... 54***
3.4 Discussion ..... 56
**4. PATTERNS OF PATHOGEN DETECTION AND CO-DETECTION ..... 62**
4.1 Introduction ..... 62
*4.1.1 Pathogen Detection in Different Clinical Syndromes ..... 62*
*4.1.2 Patterns of Pathogen Co-Detection and Their Impact on Infection ..... 62*
*4.1.3 Aims ..... 63*
4.2 Methods ..... 63
4.3 Results ..... 64
*4.3.1 Pathogen Detection in Clinical Syndromes ..... 64*
*4.3.2 Co-detection of Pathogens ..... 78*
4.4 Discussion ..... 82
**5. TRANSCRIPTOMIC SIMILARITIES BETWEEN IGA VASCULITIS AND KAWASAKI DISEASE ..... 84**
5.1 Introduction ..... 84
5.2 Methods ..... 85
*5.2.1 Patient groups ..... 85*
*5.2.2 Ethical Approval ..... 86*
*5.2.3 Statistical Analysis ..... 87*
5.3 Results ..... 89

*5.3.1 Transcriptome of IgAV patients* ..... *90*

*5.3.2 Transcriptome of KD patients* ..... *96*

*5.3.3 Comparative Analysis of IgAV and KD* ..... *96*

*5.3.4 Differences between IgAV and KD* ..... *99*

5.4 DISCUSSION ..... 100

**6. GENE EXPRESSION ANALYSIS OF FEBRILE CHILDREN ..... 103**

6.1 INTRODUCTION ..... 103

*6.1.1 Host Response to Infection with Bacteria and Viruses* ..... *103*

*6.1.2 The 12-Gene Signature for Distinguishing Bacterial from Viral Disease* ..... *103*

*6.1.2 Using the 12-gene signature to investigate infectious diseases* ..... *107*

*6.1.3 Aims* ..... *108*

6.2 METHODS ..... 108

6.3 RESULTS ..... 110

*6.3.1 Gene expression in controls* ..... *112*

*6.3.2 Expression in cases, DB and DV* ..... *112*

*6.3.3 Gene expression in different clinical syndromes* ..... *115*

*6.3.4 Expression (in response) to pathogen detection* ..... *119*

6.4 DISCUSSION ..... 126

**7. OVERALL CONCLUSIONS AND DISCUSSION ..... 131**

**8. REFERENCES ..... 134**

# List of Abbreviations

| | |
|---|---|
| AMR | Antimicrobial Resistance |
| CI | Confidence Interval |
| CMT | Centralised Molecular Test |
| CMV | Cytomegalovirus |
| CRP | C-reactive Protein |
| Ct | Cycle Time |
| DB | Definite Bacterial |
| DEG | Differentially Expressed Genes |
| DNA | Deoxyribonucleic acid |
| DRS | Disease Risk Score |
| DV | Definite Viral |
| EBV | Epstein-Barr Virus |
| ECDC | European Centre for Disease Control |
| EDTA | Ethylenediaminetetraacetic acid |
| EU | European Union |
| FC | Febrile Children |
| HHV | Human Herpes Virus |
| HSV | Herpes Simplex Virus |
| IBI | Invasive Bacterial Infection |
| IQR | Inter-Quartile Range |
| KD | Kawasaki Disease |
| LFC | Log Fold Change |
| LRTI | Lower Respiratory Tract Infection |
| Mis-C | Multisystem Inflammatory Syndrome in Children |
| OR | Odds Ratio |
| PB | Probable Bacterial |
| PERFORM | Personalised Risk Assessment in Febrile Illness to Optimise Real-life Management across the European Union |
| PPBI | Predicted Probability of Bacterial Infection |
| PV | Probable Viral |
| RNA | Ribonucleic acid |
| RSV | Respiratory Syncytial Virus |
| TB | Tuberculosis |
| TRAIL | TNF-related apoptosis-inducing ligand |
| U.K. | United Kingdom |
| URTI | Upper Respiratory Tract Infection |
| UTI | Urinary Tract Infection |

# List of Tables

Table 1: Participating European recruiting sites in PERFORM. ..... 33

Table 2: List of syndrome classifications, organised by syndrome group. Patients were assigned to at least one clinical syndrome, but multiple syndromes could be selected where appropriate. If none of the listed clinical syndromes were applicable, the ICD10 code of an alternative syndrome could be documented. ..... 35

Table 3: Investigation of potential confounding factors. Potential confounding factors were investigated for their impact on the odds of a child being a case vs control and being classified as definite bacterial vs definite viral. The impact (odds ratio), standard error of the mean and p value for these are show in the table. Age was the only significant confounder for both comparisons. ..... 40

Table 4: Clinical characteristics of patients. Ethnicity recorded was the self-reported ethnicity. Maximum CRP and neutrophil count refer to the highest value for each patient during that illness episode. The control group included hospitalized patients with no fever and with conditions considered to be not due to infection or inflammatory disorders. Controls included 35 patients with critical illness and 54 patients with primary or secondary immunodeficiency. ..... 43

Table 5: Frequency of pathogen detection in throat swabs by centralized molecular tests in relation to the clinically-assigned diagnostic category. The frequency of each pathogen's detection by CMT in relation to the patient diagnostic category is shown. The overall cases column is the sum of all detections for that virus, expressed as a percentage of all cases tested. The controls column shows detections for that virus expressed as a percentage of all controls tested. ..... 46

Table 6: Frequency of virus detection in blood by centralized molecular tests in relation to the clinically-assigned diagnostic category. The frequency of each pathogen's detection in relation to the patient diagnostic category is shown. The overall cases column is the sum of all detections for that virus, expressed as a percentage of all cases tested. The controls column shows detections for that virus expressed as a percentage of all controls tested. ..... 47

Table 7: Frequency of bacterial/fungal detection in blood by centralized molecular tests in relation to the clinically-assigned diagnostic category. The frequency of each pathogen's detection in relation to the patient diagnostic category is shown. The overall cases column is the sum of all detections for that pathogen,

expressed as a percentage of all cases tested. The controls column shows detections for that pathogen expressed as a percentage of all controls tested. ..... 48

Table 8: Clinical characteristics and initial laboratory findings for patients with Kawasaki disease, IgA Vasculitis, febrile controls and healthy controls. Abbreviations: IQR = interquartile range. Aneurysmal coronary arteries were defined as those with z-score of ≥5. ..... 89

Table 9: Top DEGs (by log fold change and B-H adjusted p-value) between IgAV and healthy controls. ..... 92

Table 10: Documented functions and tissue expression for DEGs with discordant, direction of expression between IgAV and KD. All these genes are significantly overexpressed in KD compared to controls and significantly underexpressed in IgAV compared to controls. ..... 97

Table 11: Table comparing demographic, syndrome and phenotype data between patients with both CMT and gene expression data and the larger group of patients with CMT data available. ..... 111

# List of Figures

Figure 1: Correlation between antibiotic consumption and antibiotic resistance in European countries. Figure created using publicly available data from the European Centre for Disease Control (ECDC). Resistance rates to common pathogens were calculated by pooling published resistance rates to streptococcal and staphylococcal pathogens. ..... 24

Figure 2: Performance of the bacterial/viral disease risk score in patients with confirmed viral or bacterial aetiology and patients with unknown aetiology. (Herberg, Jethro A et al. "Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children." JAMA vol. 316,8 (2016) 835-45) ..... 28

Figure 3: Clinical investigative algorithm to assign patients to individual diagnostic categories. All patients were classified to one of these phenotypes. ..... 36

Figure 4: Patient flow and numbers from recruitment to final diagnostic categories. Children with fever, history of recent fever or suspicion of infection of sufficient severity to warrant blood tests were assigned to a presumptive diagnostic group based on clinical findings on presentation. After results of all microbiological investigations, biochemical and hematological investigations and imaging were available, patients were assigned to final diagnostic categories using the algorithm in Figure 2. ..... 41

Figure 5: Frequency of detection of pathogen DNA or RNA by research molecular tests in cases and controls. Donut plots show the breakdown of the total number of positive molecular pathogen identifications from the research study in cases and controls for respiratory pathogens found in throat swabs, viruses detected in blood, and bacteria and fungi detected in blood. ..... 44

Figure 6: Odds ratios for molecular pathogen identifications comparing febrile children and non-febrile controls. Square symbols show the mean and bars and whiskers the 95% confidence intervals. Odds Ratio's above 1 (to right of the dotted line; blue arrow) indicate increased detection of pathogen in cases. OR less than 1 (left of dotted line; red arrow) indicate increased pathogen detection in controls. ..... 49

Figure 7: Odds ratios for molecular pathogen identifications comparing definite bacterial and definite viral groups. Square symbols show the mean and bars and whiskers the 95% confidence intervals. Odds Ratio's above 1 (to right of the dotted line; blue arrow) indicate increased detection of pathogen in DB; OR less than 1 (left of dotted line; red arrow) indicate increased pathogen detection in the viral group. ..... 51

Figure 8: Odds ratios for molecular pathogen identifications in combined Definite and Probable Bacterial groups compared to Definite and Probable Viral groups. OR above 1 indicates more frequent detection in a patient with bacterial infection and OR below 1 indicates increased detection in those with viral infection. Whiskers around the OR represent the 95% confidence interval. ..... 52

Figure 9: Quantitative molecular detection by increasing viral load quartile – blood viruses in relation to cases vs controls. Odds ratio with 95% CIs for detection of blood viruses in cases versus controls at increasing levels of viral load. ..... 53

Figure 10: Quantitative molecular detection by increasing viral load quartile – blood viruses in relation to Definite Viral vs Definite Bacterial diagnostic categories. Odds ratio for detection of blood viruses in Definite Viral versus Definite Bacterial cases, at increasing levels of viral load. ..... 54

Figure 11: Proportion of patients in each diagnostic category with pathogen detected. Blue bars show pathogens detected by local investigations; green bars show detection by both local and centralized molecular testing (CMT); orange bars show pathogen detection only by CMT. Purple bar shows patients without sterile site bacterial identification but classified as DB on the basis of mucosal detection of specific pathogens. Combined coloured bar height indicates proportion of all patients in each diagnostic category with pathogens detected. Individual panels show detection of respiratory viruses (left), blood viruses (middle) and bacterial and fungal pathogens (right). ..... 55

Figure 12 Sankey diagram showing patient reassignment from final diagnostic category based on local investigations alone, to a final diagnostic group based on addition of CMT. The height of each coloured bar represents number of patients in each group, and the curves crossing groups indicate patients reassigned based on CMT results. Addition of CMT resulted in an addition of 87 (3.2%) of patients to the definite bacterial group and 53 (2.0%) of patients to the definite viral group. ..... 56

Figure 13: Treemaps showing pathogen detections and clinical syndromes in patients with upper respiratory tract infections. Size of each rectangle indicates the number of detections of each pathogen or the number of patients with each syndrome in patients with URTIs. ..... 65

Figure 14: Odds ratios of pathogen detection in patients with upper respiratory tract infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed. ..... 66

Figure 15: Odds ratios for pathogen detection in patients with upper respiratory tract infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed. ..... 67
Figure 16: Treemaps showing pathogen detections and clinical syndromes in patients with lower respiratory tract infections. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen. ..... 68
Figure 17: Odds ratios of pathogen detection in patients with lower respiratory tract infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed. ..... 69
Figure 18: Odds ratios for pathogen detection in patients with lower respiratory tract infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed. ..... 70
Figure 19: Treemaps showing pathogen detections and clinical syndromes in patients with gastrointenstinal infection. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen. ..... 71
Figure 20: Treemaps showing pathogen detections and clinical syndromes in patients with central nervous system infections. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen. ..... 72
Figure 21: Odds ratios of pathogen detection in patients with central nervous system infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed. ..... 73
Figure 22: Odds ratios for pathogen detection in patients with central nervous system infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed. ..... 73
Figure 23: Treemaps showing pathogen detections and clinical syndromes in patients with urinary tract infections. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen. 74
Figure 24: Treemaps showing pathogen detections and clinical syndromes in patients with musculoskeletal infections. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen. ..... 75

Figure 25: Treemaps showing pathogen detections and clinical syndromes in patients with skin and soft tissue. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen. 76

Figure 26: Proportion of febrile children and controls with no pathogens detected, a single pathogen detection and more than one pathogen detected. ..... 78

Figure 27: Odds ratios for co-detection. The odds ratios for viral-viral (blue), bacterial-bacterial (green) and bacterial-viral co-detections in patients, compared to detection of each pathogen in isolation. Only odds ratios that were statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed. As each pathogen combination could exist in 2 places on this plot, data positions have been arranged to maximise clarity. ..... 80

Figure 28: Breakdown of patient samples and data analysis plan including the number of patients recruited with Kawasaki disease (KD), IgA Vasculitis (IgAV), healthy controls (HC) and febrile controls (FC). There were 93 patients recruited from the U.K. (1 IgAV, 61 HC and 31 FC) and 124 from the U.S.A. (77KD, 17 IgAV and 30 FC). ..... 86

Figure 29: PCA plot of patients with RSV, Influenza or adenovirus, grouped by Site. A: Before and B: After COCONUT co-normalisation. ..... 87

Figure 30: Comparison of cell abundance estimation using CIBERSORT and clinical haematology results for A: Neutrophils and B: Lymphocytes. ..... 88

Figure 31: Primary component analysis identified one clear outlier which was removed from the analysis ..... 90

Figure 32: Comparison between IgAV and healthy controls. A: Heatmap of the top 200 differentially expressed genes (DEGs) between IgAV (green) and healthy controls (grey). B: Volcano plot showing gene expression in IgAV vs healthy controls. DEGs with the highest log fold change (green) and the lowest p-value (blue) are labelled. Genes that fall into both the highest log fold change and lowest p-value groups are red. More information about these genes can be found in Supplementary Table 1. C: Pathway analysis of DEGs in IgAV vs healthy controls using Ingenuity Pathway Analysis Software. The proportion of genes in each pathway that are DEG in IgAV and the -log(B-H corrected p-value) can be seen on the x-axes while the z-score (which indicates direction of pathway regulation) is represented by the colour of each bar (red=high; blue=low). ..... 91

Figure 33: Comparison between KD and healthy controls. A: Volcano plot showing the top DEGs for KD vs healthy controls. Genes with the highest fold change are highlighted in green, genes with the lowest p-value are blue and those that fall into both groups are red. B: Pathway analysis in KD vs healthy controls using Ingenuity

Pathway Analysis Software. The proportion of DEGs in each pathway and the log(B-H corrected p-value) can be seen on the x-axes while the z-score (which indicates direction of pathway regulation) is represented by the colour of each bar (orange=high; blue=low). ..... 95

Figure 34: Concordant/Discordant Genes and boxplots highlighting difference in expression levels of DEGs. The axes of the concordant/discordant plot represent the $log_2$ fold change and the colours represent genes that are significant in KD vs HC only (blue), IgAV vs HC only (green) and both (purple). 4,930 (97%) genes showed concordant expression and 3,901 (79%) of these had a greater fold change in KD vs HC compared to IgAV vs HC. 154 (3%) genes showed opposite direction of change in expression, of which 129 (83%) were significantly differentially expressed in KD only, 18 (13%) were significantly differentially expressed in IgAV only and 7 (4%) were significantly differentially expressed in both. Boxplots show the gene expression profiles of these 7 genes, with expression in healthy controls (HC) shown in grey, IgAV shown in green and KD shown in blue. Median expression in each group is shown, with box margins representing interquartile range. ..... 98

Figure 35: Pathway comparison of KD and IgAV. A: Comparison of significantly differentially expressed pathways in IgAV and KD, compared to healthy controls and organised using hierarchical clustering. Pathways with a positive z-score are represented in orange and pathways with a negative z-score are blue. Pathways for which a z-score could not be calculated are grey. B: Pathway analysis of the 64 DEGs which were either significantly differentially expressed in IgAV vs HC but not KD vs HC or were significantly differentially expressed in both IgAV and KD vs HC with opposite direction of change. Upregulated genes and pathways in IgAV are highlighted in red and downregulated gene and pathways in IgAV in blue. Yellow links between genes indicate that the state of the downstream molecule is inconsistent with the state of the upstream regulator. In IgAV pathways of apoptosis of kidney cells and vascular smooth muscle cells are inhibited. The gene PLAC8 is one of the 7 genes that show significant orthogonal expression in IgAV and KD (Fig 33). ..... 99

Figure 36: Validation performance of the 12-gene signature and accompanying model. The statistical model transformed gene expression of the 12 genes into a single numerical output that represents the predicted probability of bacterial infection. This plot shows the median and IQR predicted probability of bacterial infection for patients in each phenotype. Figure created by R Womersley (unpublished). ..... 107

Figure 37: Consort diagram showing the number of febrile children and controls with both CMT and gene expression data available. ..... 111

Figure 38: Expression of the 7 bacterial genes and 5 viral genes in controls with no pathogens detected (green) and one or more pathogens detected (orange) by CMT. Expression is expressed as Log2 fold change (LFC) in relation to housekeeping genes. ..... 112

Figure 39: Comparison of gene expression of the 7 bacterial genes and 5 viral genes between cases (yellow) and controls (green). In this analysis all cases are grouped together, irrespective of their cause of illness. Expression is expressed as Log2 fold change (LFC) in relation to housekeeping genes..... 113

Figure 40: Density plot showing the distribution of predicted probabilities of bacterial infection in febrile children. ..... 113

Figure 41: Predicted probability of bacterial infection (PPBI) in febrile children assigned to definite bacterial (green) and definite viral (orange) phenotypes and controls (blue) calculated using a support vector machines model developed by Rebecca Womersley (unpublished). A: Density plot showing distribution PPBIs in DB and DV patients and controls. B: Distribution of PPBIs in all controls, controls with bacteria and viruses detected and controls with no pathogen detections. ..... 115

Figure 42: Boxplots showing predicted probability of bacterial infection (PPBI) in patients with each clinical syndrome, organised by syndrome group. The black vertical line, grey box and grey horizontal lines represent the median, IQR and 1.5*IQR PPBIs for each syndrome. Outliers outside the horizontal line representing 1.5*IQR are omitted for clarity. The green and orange bars show the confidence interval of PPBI for DB and DV patients respectively when the model was applied to them. For clarity, only syndromes detected in more than 20 patients are displayed. It was possible for a patient to be assigned more than one syndrome and in these cases patients may be represented more than once in this plot. ..... 118

Figure 43: Boxplot showing median and IQR predicted probability of bacterial infection (PPBI) in febrile children with detection of viruses and no bacterial co-detections. This boxplot has been superimposed on a density plot showing distribution of PPBI in DB (green) and DV (orange) patients. ..... 119

Figure 44: Boxplots showing median and IQR predicted probabilities of bacterial infection for respiratory viruses. Only respiratory viruses detected in more than 20 patients with both CMT and gene expression results available are displayed. Boxplots are overlayed on a density plot showing PPBI in DV (orange) and DB (green) patients. ..... 120

Figure 45: Boxplots showing median and IQR predicted probability of bacterial infection (PPBI) when each blood virus is detected. Boxplots are superimposed on a density plot showing the distribution of PPBIs in DV

(orange) and DB (green) patients. Only blood viruses that were detected more than 20 times in patients with both CMT and gene expression data are displayed. ..... 120

Figure 46: Scatterplots showing predicted probability of bacterial infection at increasing viral load for viruses detected in blood. For enterovirus and parechovirus, the inverse of the Ct value is used as a surrogate marker for pathogen load. The optimal viral load cut-off limit to maximise odds of detection in cases vs controls for each pathogen in shown by the red dashed line. ..... 121

Figure 47: Boxplots showing median and IQR predicted probability of bacterial infection before and after the application of viral load cut-off limits. ..... 122

Figure 48: Boxplot showing median and IQR predicted probability of bacterial infection in febrile children with detection of bacteria and no viral co-detections. This boxplot has been superimposed on a density plot showing distribution of probabilities in DB (green) and DV (orange) patients. ..... 122

Figure 49: Boxplots showing median and IQR predicted probability of bacterial infection (PPBI) when each bacteria is detected. Boxplots are superimposed on a density plot showing the distribution of PPBIs in DV (orange) and DB (green) patients. Only bacteria that were detected more than 20 times in patients with both CMT and gene expression data are displayed. ..... 123

Figure 50: Boxplots showing median and IQR of the PPBI score for patients with respiratory viruses when detected with and without bacterial co-detections. Only viruses detected in more than 20 patients are displayed. .... 124

Figure 51: Boxplots showing median and IQR of blood viruses when detected with and without bacterial co-detections. Only viruses detected in more than 20 patients are displayed. ..... 125

Figure 52: Boxplots showing median and IQR of bacteria when detected with and without viral co-detections. Only bacteria detected in more than 20 patients are displayed. ..... 126

# 1. Introduction

Fever and suspected infectious disease are the most common reason for a child to present to emergency departments in the U.K., and the proportion of attendances due to infectious diseases has been increasing over the last 10 years[1]. Infectious diseases are the 4th highest acquired cause of mortality in children in the U.K[2]. and the leading cause of death in children under 5 worldwide[3]. In fact, the World Health Organization estimates that over 60% of deaths of children under the age of 5 on the planet are due to infectious diseases[4].

The vast majority of infectious diseases in children in the U.K. are caused by bacteria and viruses and can be broadly classified into three groups: those whose aetiology can primarily be attributed to one or more bacteria, those caused by one or more viruses and those caused by a combination of both viruses and bacteria. A minority of infections are caused by other pathogens, including parasites and fungi, but these patients often have a distinctive disease history, travel history or clinical syndrome and generally do not pose the same diagnostic challenge as bacterial and viral infections.

## 1.1 Viral Infections

Most infections in the U.K. are caused by viruses. These are predominantly self-limiting illnesses, but can have variable clinical presentations, from mild, trivial illnesses that can be managed without intervention, to severe systemic presentations, with raised inflammatory markers, mimicking bacterial infections[5,6]. Although the majority result in upper or lower respiratory tract symptoms[5,7], viral infections can cause an array of clinical syndromes

affecting a number of body systems, from gastroenteritis[8] and hepatitis[9] to meningitis and encephalitis[10].

In addition to the primary acute illness, early infection with a number of viruses can lead to changes in the host immune and inflammatory response and have been implicated in inducing predisposition to other conditions. These include acute syndromes such as Mis-C[11] and long-term diseases such as including type 1 diabetes mellitus[12], asthma[13] and eczema[14]. More recently, viral infections have been linked to lingering non-specific syndromes that last such as chronic fatigue syndrome[15,16] and long covid[17].

Viral infections, or the host response to viral infections, can lead to significant changes in a patient's microbiome or immune response, opening the door to further disease. In animal models, influenza and RSV infection has been shown to result in significant changes in the gut microbiome and this in turn has been shown to impair the innate immune response to subsequent infection[18]. Although the precise mechanism is unclear, infection with influenza and other viruses has been shown to increase susceptibility in humans to secondary bacterial infection, possibly via a combination of immune dysregulation and damage to the epithelial lining of the respiratory tract[18–22].

## 1.2 Bacterial Infections

The widespread introduction of vaccines against *H.influenzae* type B, *S.pneumoniae* and *N.meningitidis* group B, as well as improvements in hygiene and personal health have led to a decline in the incidence of invasive bacterial infections (IBIs) over the last 20 years, and they now account for approximately 10% of children presenting to emergency departments

with fever or suspected infection[23,24]. Although host defences alone are sufficient to combat some bacterial infections, many will require prompt therapy with antimicrobial agents to prevent significant morbidity and mortality. If left untreated, some bacterial infections such as bacterial meningitis are reported to have a mortality rate close to 100%.[25] Like viral infections, bacteria can cause infection in a number of different body systems and present with a diverse range of clinical syndromes, from isolated fever without any clear focus of infection, to local infections of the skin or soft tissue and systemic illnesses such as bacteraemia or meningitis[23,26–28].

## 1.3 Diagnosis of Infectious Disease

Due to the high morbidity and mortality associated with delayed identification IBIs, much of the initial clinical approach to infectious diseases is aimed at confirming or ruling-out bacterial infection and the need for immediate antimicrobial therapy. Conventional diagnostics for confirming bacterial infection are based on the isolation of bacteria, cultured from a sterile site such as blood or cerebrospinal fluid. However, sterile site cultures often require incubation for more than 24 hours. Furthermore, it is often not possible or not appropriate to obtain samples of sufficient volume or from the site of infection and, even when a sufficient sample is obtained using best practice techniques, there is a contamination rate of up to 3%[29]. These factors contribute to the low positivity rate and diagnostic yield of blood cultures of 4.1% and 2% respectively[30].

Although the presence of a viral infection does not preclude the presence of a concurrent bacterial infection, viral identification is important to aid clinicians to understand and predict the natural history and prognosis of a child's illness. In addition, it allows for measures to be

put in place to limit spread of highly contagious viruses to staff and other children, including those with impaired immunity. Before the introduction of molecular methods, cell culture and antigen detection methods were the primary modalities of viral diagnostics. Cell culture methods were limited to viruses that were known to support growth and could take up to 2 weeks for results to be available and antigen detection techniques lack sensitivity and can only be used where the appropriate assays existed, which excluded common viruses including rhinovirus and coronaviruses[31,32].

Limitations in these diagnostic techniques for infectious diseases, means that no causative organism or organisms are identified in more than 50% of children with suspected infection[33]. As a result, to safeguard against the implications of delaying antimicrobial treatments for bacterial infections as outlined above, up to a third of children starting on antibiotics may not need them[34].

## 1.4 Impact of the Overuse of Antimicrobials

Over the last decade it has become clear that the use of antimicrobials in patients where a bacterial infection cannot be excluded due to the inadequacies of current diagnostics is not without risk. Antimicrobial resistance (AMR) has been identified by the World Health Organisation as one of the major health concerns of the 21st century, with reports suggesting that it accounts for more that 1 million deaths worldwide each year and resistance to all antibiotics in clinical use has now been observed[35,36]. Overuse of antibiotics where they are not necessary, such as in treating viral infections, has been identified as a key driver for AMR (Figure 1) and reducing this practice by developing rapid and accurate diagnostics, is part of the World Health Organisation AMR strategy[36].

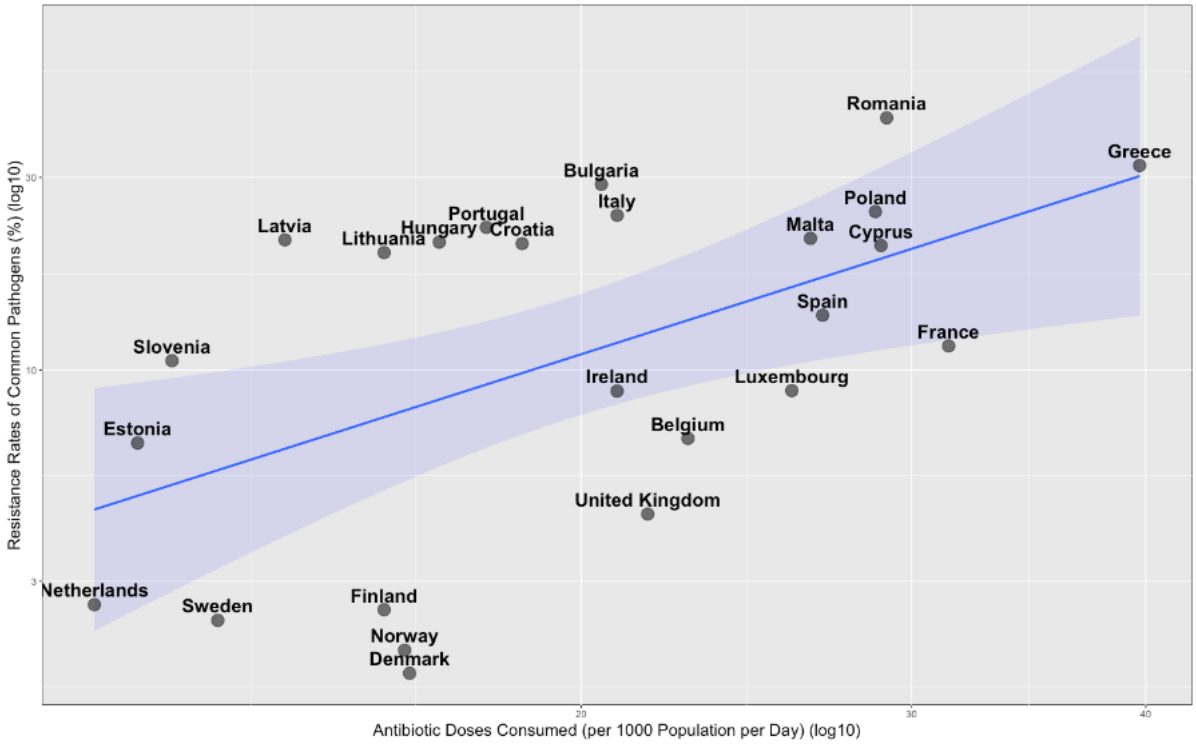



*Figure 1: Correlation between antibiotic consumption and antibiotic resistance in European countries. Figure created using publicly available data from the European Centre for Disease Control (ECDC). Resistance rates to common pathogens were calculated by pooling published resistance rates to streptococcal and staphylococcal pathogens.*

At the individual level, antibiotics have been shown to disrupt the host microbiota, resulting in wide-ranging effects including vitamin K deficiency, development of conditions such as obesity, asthma, allergies, inflammatory bowel disease and autoimmune diabetes and even increased susceptibility to future infections[37,38].

## 1.5 Molecular Diagnostic Approaches

The widespread availability of molecular diagnostics in infectious diseases have revolutionised clinical practice. The most common molecular diagnostic technique used in infectious diseases involve genome detection, such as real-time PCR, where sequences of pathogen DNA or RNA are detected in patient samples and results can be available within a matter of hours[39–41]. Other molecular diagnostic approaches are based on the identification

and analysis of circulating cell-free DNA in plasma using next-generation downstream sequencing techniques[42] or combine molecular diagnostics with mass spectrometry to identify single nucleotide polymorphisms (SNPs), which can be used for pathogen typing and identifying patterns of antimicrobial resistance[43].

In infectious disease, molecular methods are most commonly used in patients with respiratory disease, where they have been shown to produce rapid results with high sensitivity and can detect viruses for which no other clinically available testing modality exists[31]. In children with bronchiolitis or acute wheeze, molecular diagnostic methods increase viral detection rates from 40% to 95%[44].

However, due to the high sensitivity of molecular diagnostics, it is difficult to distinguish a clinically relevant pathogen detection from sub-clinical colonisation or latent disease. This is highlighted by studies reporting positive bacterial detections from healthy controls[45].

With the exception of a small number of blood bacterial and viral assays and specialist investigations, this has limited the clinical use of molecular diagnostics to the detection of respiratory viruses in patients with respiratory illnesses.

## 1.6 Co-Infections

The widespread use of molecular diagnostics has revealed that infectious diseases are more complicated than previously thought. The historical view that infections are either bacterial or viral nature is being replaced by a model involving intricate interactions between one or multiple pathogens and host responses. More than one clinically relevant pathogen has

been reported to be detected in between 10% and 84% of children, depending on age group and site of infection[19,21,32,44,45].

Initial reports on co-infection proposed indirect relationships between organisms, as seen in secondary bacterial infection following influenza infection, where the innate immune response is hampered by the primary influenza infection, allowing for secondary bacterial infection, usually with bacteria that commonly colonise the upper respiratory tract[18,21,46]. Direct interaction between organisms has also been described, for example RSV has been shown to enhance binding of *S.pneumoniae* and *H.influenzae* in the development of lower respiratory tract infections[47] and bacterial surface polysaccharides have been shown to bind to viral particles to enhance their adherence and thermal stability in the development of gastrointestinal infections[47,48].

Many direct relationships have now been described including bacteria promoting or inhibiting future infections by viruses or other bacteria and viruses promoting or inhibiting future infections by bacteria or other viruses[19,47–50].

Due to these reports of the potential direct and indirect mechanisms by which micro-organisms can cause disease, in this thesis I assume all micro-organism detections as potentially pathogenic and henceforth refer to them all as pathogens.

## 1.7 Novel Diagnostic Techniques

Due to the high incidence of co-infections, and the difficulty in discerning whether a pathogen detection represents active disease, molecular diagnostic techniques are not

sufficient to impact the overuse of antimicrobial agents. Much research is directed at developing novel diagnostic techniques to better identify patients where antimicrobial agents can be safely withheld.

### 1.7.1 Transcriptomics

Gene expression analysis has been introduced in clinical specialties such as oncology to aid diagnosis and researchers have explored the utility of these techniques to determine the underlying aetiology of infections. RNA expression studies have shown that gene expression can be used to accurately identify patients with specific infections, including TB[51], influenza[52], RSV[52] and malaria[53] and inflammatory conditions, such as Kawasaki disease[54].

Previous studies have used the expression levels of just 2 genes: FAM89A and IF44L, combined to create a "Disease Risk Score" (DRS) to successfully distinguish children with bacterial infections from those with viral infections.[55] Further studies have shown that this DRS differentiates between bacterial and viral infections in neonates[56] and children with diarrhoea.[57] Although this DRS is able to differentiate patients with proven bacterial or viral aetiology, and can segregate some patients with unclear aetiology into likely bacterial and likely viral infections, there is a large cohort of patients with unclear aetiology and indeterminate disease risk scores, where the signature is unable to define the illness (Figure 2).

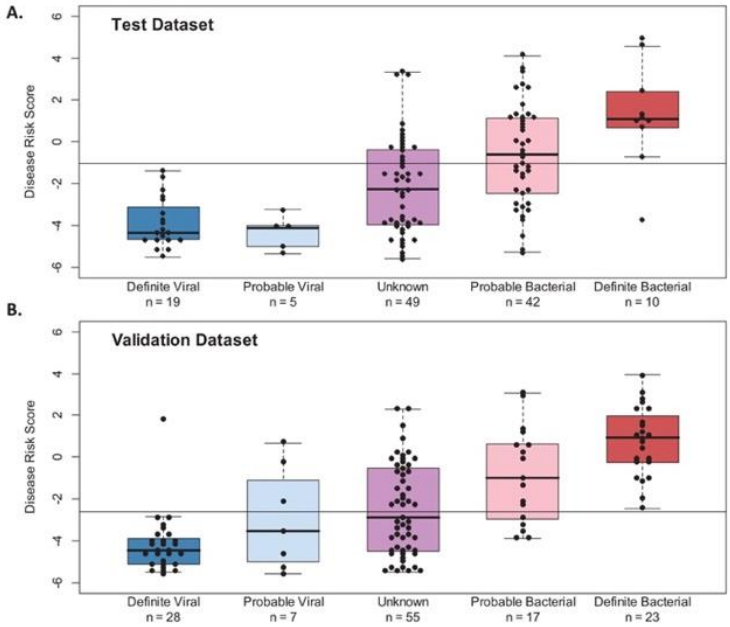



*Figure 2: Performance of the bacterial/viral disease risk score in patients with confirmed viral or bacterial aetiology and patients with unknown aetiology. (Herberg, Jethro A et al. "Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children." JAMA vol. 316,8 (2016) 835-45)*

Other gene expression signatures have replicated this pattern, with good separation of patients who have proven bacterial or proven viral infections and superior performance to other laboratory tests[58], but poorer performance in patients with uncertain aetiology or co-infections.[59,60]

### 1.7.2 Protein Biomarkers

Protein biomarkers have the potential to be developed into quick and relatively inexpensive tests and may not require specialist equipment to be incorporated into a hospital laboratory set-up. Many biomarkers have been studied, including interleukin 6 and procalcitonin. Along with C-reactive protein, these biomarkers have been shown to be effective at identifying patients with bacterial infections in some scenarios, but not others[61,62]. Models involving the analysis of multiple protein biomarkers have been developed, which show considerable

promise, such as the MeMed BV test which utilises CRP, IP-10 and TRAIL to distinguish bacterial from viral infection[63]. However, these novel multi-biomarkers remain expensive and their performance in real-world settings and patients with unclear aetiology is yet to be determined[64].

### 1.7.3 mRNA Detection

Quantitative multi-mRNA diagnostics have the potential to be fast and accurate tools to distinguish between bacterial and viral infections[65]. Initial reports show superiority in sensitivity and specificity for the identification of patients with bacterial infections compared to PCT[66].

### 1.7.4 Clinical Metagenomics

The combination of clinical metagenomics and next generation sequencing has makes the analysis of the entire genetic content of a clinical sample possible. Reports have highlighted the clinical utility of these techniques in diagnostics, identification of novel pathogens, identification of antimicrobial resistance and public health measures[67,68]. The clinical translation of metagenomics has been limited by cost, turnaround time and difficulties in interpretation of results[67].

### 1.7.5 Biosensors

Biosensors are a relatively new technology whose development has been accelerated during the Covid-19 pandemic. Initial reports suggest they would be capable of providing quick and accurate detection of pathogens, but further investigation and validation of these techniques are needed[69,70].

## 1.8 Project Aims

I believe that these novel diagnostic approaches have thus far failed to result in a diagnostic test that can accurately identify patients where antimicrobials can be safely withheld because their discovery is based on an incomplete representation of infectious diseases. Molecular diagnostics have shown the importance of co-infections in the development and natural history of infectious diseases and revealed that the development of infections is more complex than previously thought. Transcriptomic studies can investigate the host response to infection and shed light on the interplay between different pathogens and their roles in febrile illness. This may be a necessary foundation for the development of novel diagnostics that are accurate in patients with bacterial and viral mono-infections and those with more complicated infections involving multiple pathogens. To do this I will investigate these four aims:

1) Apply broad-ranging molecular diagnostics to febrile children and healthy controls to present a more complete picture of the pathogens present in febrile illnesses.
2) Examine patterns of pathogen detection and co-detection in patients with infections at different sites or involving different body systems.
3) Demonstrate that analysis of the transcriptome can be used to investigate differences in the host response to diseases that have an overlapping clinical picture.
4) Incorporate host transcriptomic data to the molecular diagnostic results of febrile children and healthy controls to explore the biological relevance of differences in host gene expression in the presence of different pathogens or combinations of pathogens to better understand their role in causing febrile illnesses.

# 2. Methods

These methods cover patient recruitment, clinical assignment and molecular pathogen detection as part of the PERFORM study (Personalised Risk Assessment in Febrile Illness to Optimise Real-Life Management across the European Union; www.perform2020.org), applicable to most of the analyses performed. Where patient recruitment or research techniques deviated from these methods, they are described in the relevant chapter alongside details of any specific analyses.

## 2.1 Patient Recruitment

Patients were recruited as part of the PERFORM study, funded by the European Union's Horizon 2020 programme under GA number 668303. Patients were recruited from 17 hospitals across 9 European countries between April 2016 and December 2020. The recruiting sites and countries involved are shown in Table 1. Children aged under 18 years of age presenting with fever, recent history of fever, or suspected infection who were considered ill enough to warrant blood tests, were recruited from emergency departments, paediatric inpatient wards, and intensive care units. Non-febrile children attending hospital with non-infectious conditions, who were undergoing blood sampling were recruited as controls if they had no febrile symptoms or vaccinations within the previous 3 weeks.

Recruited febrile children (FC) underwent clinical assessment and investigation as part of their routine clinical care, according to best practice at each institution, including blood and urine cultures, microscopy, and cultures of cerebrospinal fluid (where clinically indicated),

and detection of bacteria and viruses in throat swabs or nasopharyngeal aspirates. Serological studies and chest radiographs and other imaging were undertaken at the discretion of treating clinicians. Serial clinical findings and results from all investigations were recorded on an electronic standardized case report form and stored on a centralized database.

| Country | Participating site | Ethics Approval Number |
|---|---|---|
| Spain | Hospital Clínico Universitario de Santiago de Compostela | 2016/331 |
| United Kingdom | St Mary's Hospital, Imperial College Healthcare NHS Trust, London<br>Alder Hey Children's Hospital, Liverpool<br>Great North Children's Hospital, Newcastle-upon-Tyne<br>John Radcliffe Hospital, Oxford<br>Royal Alexandra Children's Hospital, Brighton | 16/LO/1684 |
| Netherlands | Sophia's Children's Hospital, Rotterdam<br>Academic Medical Centre, Amsterdam<br>Radboudumc, Nijmegen | NL58103.091.16 |
| Greece | P. and A. Kyriakou Children's Hospital, Athens | 415/13.06.16 |
| Austria | Medizinische Universität Graz, Graz | 28-518 ex 15/16 |
| Slovenia | University Medical Centre Ljubljana | 0120-483/2016-3 |
| Latvia | Children's Clinical University Hospital, Riga | 1/16-07-14 |
| Germany | Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University, Munich | 699-16 |
| Switzerland | University Children's Hospital, Universität Bern, Bern | 2016-01835 |

*Table 1: Participating European recruiting sites in PERFORM.*

## 2.2 Clinical Assignment

Once all clinical information, including medical records, blood results, radiology, microbiology and other clinical investigations were reviewed by a local phenotyping team made up of clinical and research professionals, patients were assigned one or more clinical syndromes and a single overall phenotype. I undertook regular phenotyping sessions at St Mary's Hospital, Imperial College London. The aim of assigning phenotype was to identify subgroups of recruited patients that could confidently be assigned as having a primarily bacterial or viral infection. With this aim in mind, phenotyping was performed with caution to avoid overconfidence of assignment, and it is therefore likely that there were patients phenotyped as having less certain or unknown aetiology that have bacterial or viral disease.

To ensure standardised assignment across recruiting centres, phenotyping teams were provided with clear definitions to select one of the clinical syndromes in Table 2, with the option to document alternative syndromes where required and an algorithm to use when deciding the overall phenotype (Figure 3).

To maximise the accuracy of phenotype assignment, representatives from each recruiting site were invited to regular phenotyping webinars and quality control checks were performed using clinical data entered onto the central database to ensure that phenotype assignment followed the phenotyping algorithm. I led a number of quality control checks within the research consortium to ensure accuracy of the data entered

| Syndrome Group | Clinical Syndrome | | |
|---|---|---|---|
| Upper respiratory tract infection (URTI) | Croup | Pharyngitis | Stomatitis (non-HSV) |
| | Mastoiditis | Tonsilitis | Undefined URTI |
| | Otitis Media | Sinusitis | |
| Lower respiratory tract infection (LRTI) | Asthma exacerbation | Pleural Effusion | Undefined LRTI |
| | Bronchiolitis | Pneumonia | |
| | Empyema | Viral-induced wheeze | |
| Gastrointestinal and intra-abdominal infections | Appendicitis | Mesenteric adenitis | |
| | Diarrhoeal illness | Peritonitis | |
| Central nervous system infections | Encephalitis | Meningitis | |
| Urinary tract infections (UTIs) | Lower UTI | Pyelonephritis | |
| Musculoskeletal infections | Discitis | Osteomyelitis | Septic arthritis |
| | Myositis | Reactive arthritis | Transient synovitis |
| Skin and soft tissue infections | Abscess | Infected eczema | Periorbital cellulitis |
| | Cellulitis | Lymphadenitis | Wound infection |
| | Impetigo | Necrotising fasciitis | |
| Pathogen specific syndromes | Flu-like illness | Lyme disease | Roseola |
| | Glandular fever | Malaria | Scarlet fever |
| | Hand-foot-mouth | Mumps | Staphylococcal scalded skin |
| | HSV stomatitis | Pertussis | Varicella |
| Other infections | Conjunctivitis | Endocarditis | Line infection (central) |
| Sepsis syndromes | Bacteraemia | Purpura fulminans | Toxic shock syndrome |

*Table 2: List of syndrome classifications, organised by syndrome group. Patients were assigned to at least one clinical syndrome, but multiple syndromes could be selected where appropriate. If none of the listed clinical syndromes were applicable, the ICD10 code of an alternative syndrome could be documented.*

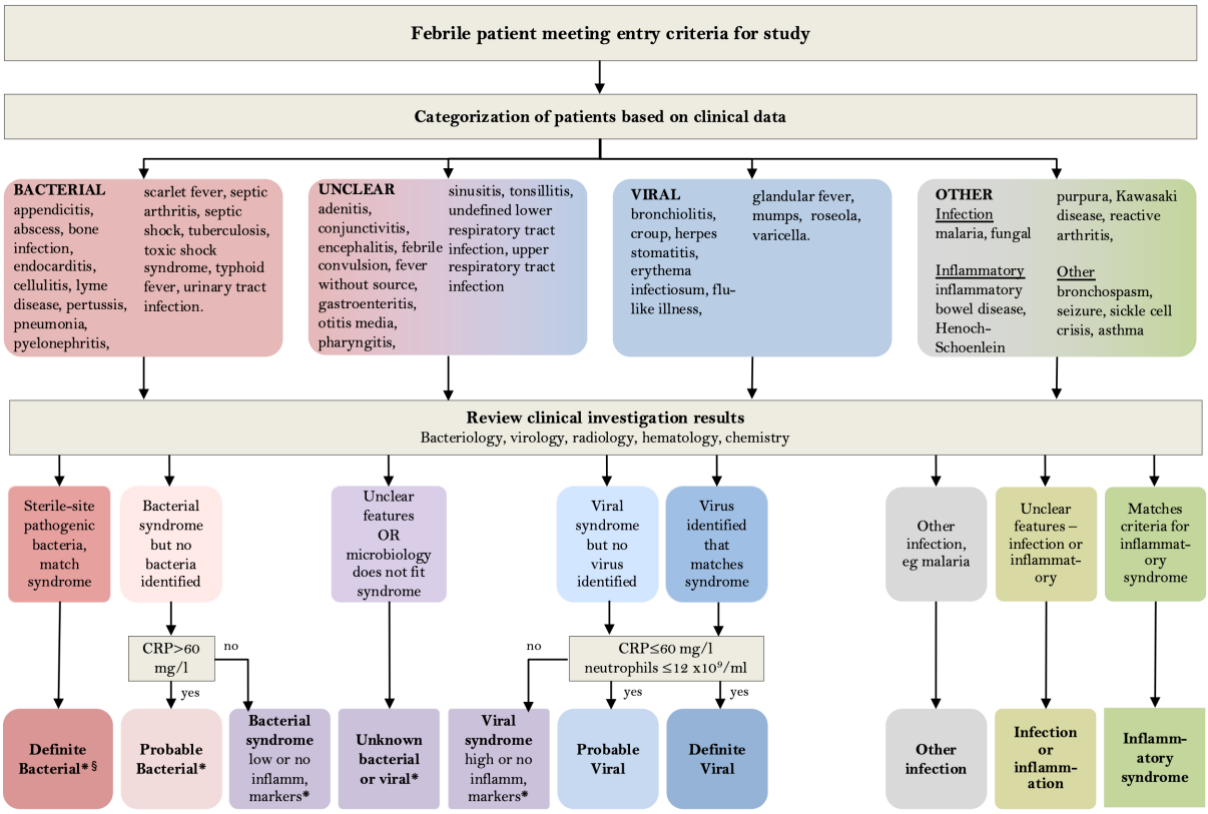



*Figure 3: Clinical investigative algorithm to assign patients to individual diagnostic categories. All patients were classified to one of these phenotypes.*

**Denotes diagnostic categories in which viral co-infection may or may not be present.*

*§Pathogens normally only detected on mucosal surfaces or diagnosed serologically (M. tuberculosis, B. pertussis, M. pneumoniae, Borrelia species, Campylobacter and Salmonella) can be included in the Definite Bacterial category if the clinical presentation is consistent with the pathogen detected, but are analysed separately for some analyses.*

## 2.4 Molecular Diagnostic Testing

Patient respiratory swab and EDTA whole blood samples were sent to Micropathology Ltd, Coventry, U.K. for centralised molecular testing.

At Micropathology Ltd Respiratory swabs were thawed from -80ºC to room temperature and vortexed thoroughly and a 200µL aliquot of eNat® medium was transferred to a 96-well lysis

plate for nucleic acid extraction. Blood samples were -80ºC to room temperature and mixed thoroughly. The sample was split into 2 x 200µL aliquots. One of these aliquots was enzymatically treated by adding 50µL of pre-lysis buffer (lysozyme/lysostaphin) to ensure complete release of bacterial and fungal DNA. The other 200µL aliquot, which was to be used for virological assessment was not pre-treated prior to extraction. Total nucleic acid was extracted in 96-well format and stored at -20ºC prior to diagnostic testing.
Respiratory samples were analysed using the commercially available NxTAG™ Respiratory Pathogen Panel (RPP) assay (Luminex® Corporation), and results were reported in a binary positive/negative format.

Blood samples were process using in-house developed assays with highly species-specific pathogen/probe sets. For adenovirus, CMV, EBV, HHV6a, HHV6b, HHV7 and parvovirus assays, results were reported as viral loads in DNA copies/ml with no values reported below the assay's theoretical limit of detection (100 DNA or RNA copies/ml, equivalent to 1 viral genome/test). Results from bacterial assays for *N.meningitidis, H.influenzae,* streptococci (incl. *S.pneumoniae)*, *S.aureus, S.maracescens, A.baumannii, P.aeruginosa, K.kingae, K.pneumoniae, E.coli,* Enterococci and Enterobacteradeae, fungal targets for aspergillus and candida species and viral assays for enterovirus and parechovirus were reported as cycle times (Ct). To avoid false positive results from contaminating environmental organisms, positive results with a high Ct (low bacterial/fungal load) were repeated for confirmation.

## 2.5 Statistical Methods

All analysis was performed using the open-source statistical software R version 3.6.0 (R Foundation for Statistical Computing)[71] and RStudio version 1.2.1335 (www.rstudio.com). The statistical techniques used for each analysis are described in the relevant chapters.

# 3. Pathogens Detected in European Febrile Children using Molecular Diagnostics

## 3.1 Introduction

To gain a better understanding febrile illnesses in children, it is first necessary to investigate which pathogens are present, and how they relate to the child's clinical course. In this analysis I assume that all detected micro-organisms have the potential to have a direct or indirect impact on the child's illness and are henceforth referred to as pathogens. Molecular diagnostics have been shown to provide additional pathogen data, including detections of pathogens that are not fully understood. To use this data to guide clinical management, or in the development of novel diagnostics, the relationship between pathogen detection and disease must be explored.

### 3.1.1 Aims

In this study I aim to:

1. Describe the pathogens detected in febrile children and healthy controls presenting to healthcare services across Europe.
2. Relate pathogen detection to presumed aetiology in febrile children.

3. Examine if factors such as pathogen load influence the confidence by which an illness is classed as a bacterial or viral infection.
4. Explore whether the addition of pathogen data from molecular diagnostics could be used to increase the number of children with confident assignment as bacterial or viral infection.

## 3.2 Methods

The results from centralised molecular tests (CMTs) were assessed in relation to the locally assigned diagnostic categories, by calculating odds ratios (ORs) with 95% confidence intervals (CIs) for the detection of each organism in both bacterial vs viral diagnostic categories, and in cases vs controls. I used a multiple logistic regression model with iteratively re-weighted least squares using the R package glmnet[72]. Age, gender, ethnicity, country of recruitment, immunodeficiency status and clinical syndrome were investigated as potential confounding factors (Table 3). Age was identified as the only significant factor and the logistic regression model was therefore designed to adjust for age as a confounder. In addition, to reflect the differences in epidemiology and investigation of febrile illnesses in infants compared to older children, ORs with 95% CIs were calculated for the detection of each organism in cases vs controls and bacterial vs viral diagnostic categories in children under 1 and children over 1 separately. For blood viral assays, where pathogen load counts were available, ORs with 95% CIs were calculated at increasing quartiles of viral load for the detection of each virus in cases vs controls, and in definite bacterial vs viral diagnostic categories. The ability of a positive virus test to exclude a bacterial infection was calculated for each individual virus, and for the aggregated detection of any virus in blood or respiratory samples.

| Comparison | Coefficient | Impact | Std Error | p |
|---|---|---|---|---|
| | | | | |
| **Case vs Control** | Age | -0.0047 | 0.00029 | $<2e^{-16}$ |
| **Case vs Control** | Gender | 0.035 | 0.033 | 0.3 |
| **Case vs Control** | Ethnicity | -0.339 | 0.242 | 0.16 |
| **Case vs Control** | Country | 0.089 | 0.058 | 0.13 |
| **Case vs Control** | Immunodeficiency | -0.154 | 1.97 | 0.94 |
| | | | | |
| **Definite Bacterial vs Definite Viral** | Age | 0.0033 | 0.00066 | $5.9e^{-7}$ |
| **Definite Bacterial vs Definite Viral** | Gender | -0.021 | 0.081 | 0.8 |
| **Definite Bacterial vs Definite Viral** | Ethnicity | -0.902 | 1.01 | 0.37 |
| **Definite Bacterial vs Definite Viral** | Country | -0.093 | 0.171 | 0.59 |
| **Definite Bacterial vs Definite Viral** | Immunodeficiency | -0.299 | 0.193 | 0.12 |
| **Definite Bacterial vs Definite Viral** | Clinical Syndrome | 0.119 | 0.141 | 0.39 |

*Table 3: Investigation of potential confounding factors. Potential confounding factors were investigated for their impact on the odds of a child being a case vs control and being classified as definite bacterial vs definite viral. The impact (odds ratio), standard error of the mean and p value for these are show in the table. Age was the only significant confounder for both comparisons.*

## 3.3 Results

### 3.3.1 Patient Recruitment

Centralised molecular test (CMT) results were reported on 4 611 patients recruited between April 2016 and December 2019 as part of the PERFORM study. 4 199 patients had CMT on blood samples, 2 762 on throat swabs, with CMT results available on both blood and throat swabs in 2 350 patients. In addition, CMT results were reported on 1061 paediatric controls of which 1002 had CMT on blood samples, 485 on throat swabs and 426 on both (Figure 4).

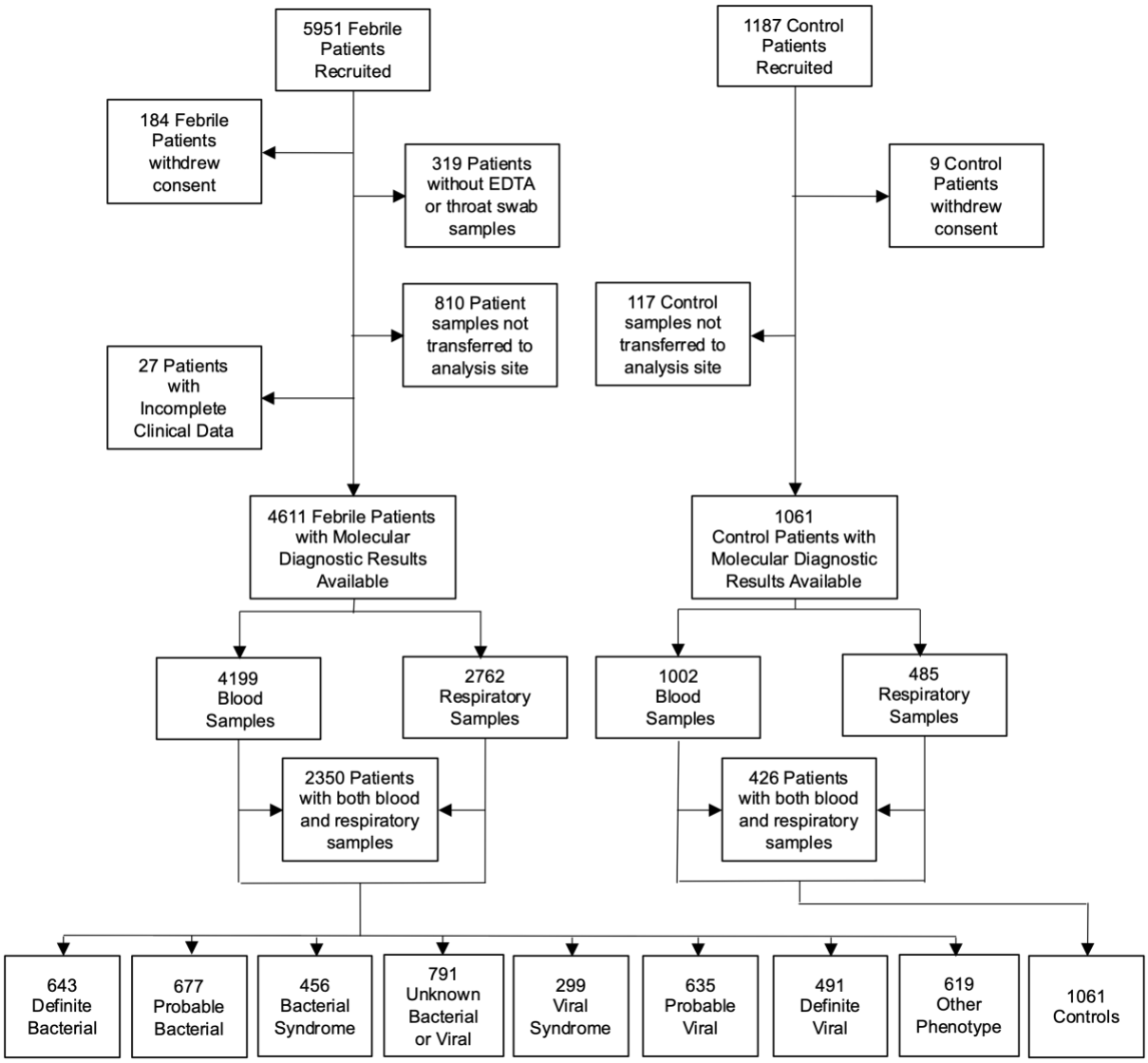



*Figure 4: Patient flow and numbers from recruitment to final diagnostic categories. Children with fever, history of recent fever or suspicion of infection of sufficient severity to warrant blood tests were assigned to a presumptive diagnostic group based on clinical findings on presentation. After results of all microbiological investigations, biochemical and hematological investigations and imaging were available, patients were assigned to final diagnostic categories using the algorithm in Figure 2.*

The diagnostic categories assigned to the 4 611 patients by their local recruiting team, using the available clinical and investigation data and the study phenotyping algorithm (Figure 3), comprised of 643 (14%) patients with a DB phenotype, 677 (15%) patients with a PB

phenotype, 491 (10%) with a DV phenotype and 635 (14%) patients with a PV phenotype.

There were 1546 (34%) patients categorised as having an infection of less certain aetiology.

| | Definite Bacterial (N=643) | Probable Bacterial (N=677) | Bacterial Syndrome (N=456) | Unknown Bacterial or Viral (N=791) | Viral Syndrome (N=299) | Probable Viral (N=635) | Definite Viral (N=491) | Other Phenotype (N=619) | Control (N=1061) | Overall (N=5672) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Gender M (%)** | 348 (54.1%) | 375 (55.4%) | 253 (55.5%) | 453 (57.3%) | 175 (58.5%) | 352 (55.4%) | 262 (53.4%) | 340 (54.9%) | 601 (56.6%) | 3159 (55.7%) |
| **Age (months)** | | | | | | | | | | |
| Median [IQR] | 58.9 [15.0-131] | 59.9 [26.5-128] | 63.3 [26.2-128] | 55.2 [19.8-111] | 38.3 [16.1-86.5] | 43.3 [14.8-96.3] | 39.7 [9.32-86.4] | 104 [38.9-161] | 122 [63.2-167] | 64.5 [22.7-134] |
| **Ethnicity** | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Afro Caribean | 1 (0.2%) | 1 (0.1%) | 3 (0.7%) | 6 (0.8%) | 1 (0.3%) | 4 (0.6%) | 2 (0.4%) | 8 (1.3%) | 3 (0.3%) | 29 (0.5%) |
| Asian | 38 (5.9%) | 31 (4.6%) | 16 (3.5%) | 65 (8.2%) | 22 (7.4%) | 33 (5.2%) | 46 (9.4%) | 37 (6.0%) | 34 (3.2%) | 322 (5.7%) |
| Black African | 13 (2.0%) | 17 (2.5%) | 5 (1.1%) | 38 (4.8%) | 11 (3.7%) | 12 (1.9%) | 17 (3.5%) | 30 (4.8%) | 15 (1.4%) | 158 (2.8%) |
| Middle-Eastern | 15 (2.3%) | 11 (1.6%) | 14 (3.1%) | 21 (2.7%) | 9 (3.0%) | 12 (1.9%) | 20 (4.1%) | 16 (2.6%) | 16 (1.5%) | 134 (2.4%) |
| White European | 536 (83.4%) | 559 (82.6%) | 393 (86.2%) | 598 (75.6%) | 230 (76.9%) | 527 (83.0%) | 366 (74.5%) | 467 (75.4%) | 911 (85.9%) | 4587 (80.9%) |
| Mixed | 10 (1.6%) | 14 (2.1%) | 7 (1.5%) | 19 (2.4%) | 10 (3.3%) | 13 (2.0%) | 16 (3.3%) | 9 (1.5%) | 17 (1.6%) | 115 (2.0%) |
| Other | 30 (4.7%) | 44 (6.5%) | 18 (3.9%) | 44 (5.6%) | 16 (5.4%) | 34 (5.4%) | 24 (4.9%) | 52 (8.4%) | 65 (6.1%) | 327 (5.8%) |
| **Country** | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Austria | 44 (6.8%) | 77 (11.4%) | 33 (7.2%) | 64 (8.1%) | 14 (4.7%) | 19 (3.0%) | 42 (8.6%) | 34 (5.5%) | 165 (15.6%) | 492 (8.7%) |
| Germany | 7 (1.1%) | 21 (3.1%) | 4 (0.9%) | 17 (2.1%) | 11 (3.7%) | 15 (2.4%) | 5 (1.0%) | 21 (3.4%) | 13 (1.2%) | 114 (2.0%) |
| Greece | 72 (11.2%) | 37 (5.5%) | 9 (2.0%) | 88 (11.1%) | 23 (7.7%) | 130 (20.5%) | 59 (12.0%) | 47 (7.6%) | 35 (3.3%) | 500 (8.8%) |
| Latvia | 71 (11.0%) | 106 (15.7%) | 52 (11.4%) | 37 (4.7%) | 13 (4.3%) | 100 (15.7%) | 43 (8.8%) | 31 (5.0%) | 131 (12.3%) | 584 (10.3%) |
| Netherlands | 96 (14.9%) | 78 (11.5%) | 57 (12.5%) | 54 (6.8%) | 49 (16.4%) | 116 (18.3%) | 50 (10.2%) | 102 (16.5%) | 114 (10.7%) | 716 (12.6%) |
| Slovenia | 52 (8.1%) | 62 (9.2%) | 27 (5.9%) | 16 (2.0%) | 25 (8.4%) | 36 (5.7%) | 17 (3.5%) | 19 (3.1%) | 23 (2.2%) | 277 (4.9%) |
| Spain | 79 (12.3%) | 67 (9.9%) | 57 (12.5%) | 121 (15.3%) | 30 (10.0%) | 54 (8.5%) | 68 (13.8%) | 47 (7.6%) | 218 (20.5%) | 741 (13.1%) |
| Switzerland | 16 (2.5%) | 13 (1.9%) | 18 (3.9%) | 16 (2.0%) | 12 (4.0%) | 19 (3.0%) | 2 (0.4%) | 15 (2.4%) | 10 (0.9%) | 121 (2.1%) |
| Taiwan | 20 (3.1%) | 0 (0%) | 0 (0%) | 7 (0.9%) | 1 (0.3%) | 2 (0.3%) | 3 (0.6%) | 3 (0.5%) | 0 (0%) | 36 (0.6%) |
| U.K. | 186 (28.9%) | 216 (31.9%) | 199 (43.6%) | 371 (46.9%) | 121 (40.5%) | 144 (22.7%) | 202 (41.1%) | 300 (48.5%) | 352 (33.2%) | 2091 (36.9%) |
| **Max CRP (mg/L)** | | | | | | | | | | |
| Median [IQR] | 102 [36.4-202] | 146 [100-220] | 20.5 [7.00-37.0] | 35.9 [13.1-79.4] | 67.0 [23.3-101] | 10.0 [4.00-24.0] | 10.3 [4.00-26.0] | 19.5 [4.00-78.0] | NA | 35.0 [9.00-102] |
| **Max Neutrophils (10^9/L)** | | | | | | | | | | |
| Median [IQR] | 10.1 [5.87-15.4] | 12.8 [8.55-18.9] | 7.09 [4.52-11.4] | 6.68 [3.50-11.3] | 12.5 [7.50-15.3] | 4.92 [2.89-7.80] | 4.46 [2.53-7.20] | 6.61 [3.81-11.2] | NA | 7.55 [4.10-12.4] |
| **Required Admission** | 514 (79.9%) | 558 (82.4%) | 351 (77.0%) | 534 (67.5%) | 214 (71.6%) | 266 (41.9%) | 350 (71.3%) | 427 (69.0%) | 232 (21.9%) | 3446 (60.8%) |
| **PICU Admission** | 129 (20.1%) | 118 (17.4%) | 25 (5.5%) | 116 (14.7%) | 58 (19.4%) | 24 (3.8%) | 47 (9.6%) | 109 (17.6%) | 35 (3.3%) | 661 (11.7%) |
| **Ventilated** | 62 (9.6%) | 78 (11.5%) | 12 (2.6%) | 93 (11.8%) | 47 (15.7%) | 15 (2.4%) | 39 (7.9%) | 64 (10.3%) | 0 (0%) | 410 (7.2%) |
| **Primary or Secondary Immunodeficiency** | 69 (10.7%) | 62 (9.2%) | 27 (5.9%) | 160 (20.2%) | 20 (6.7%) | 40 (6.3%) | 51 (10.4%) | 93 (15.0%) | 57 (5.4%) | 579 (10.2%) |
| **Upper Respiratory Tract Infection** | 46 (7.2%) | 169 (25.0%) | 83 (18.2%) | 213 (26.9%) | 88 (29.4%) | 289 (45.5%) | 105 (21.4%) | 39 (6.3%) | 0 (0%) | 1032 (18.2%) |
| **Lower Respiratory Tract Infection** | 97 (15.1%) | 242 (35.7%) | 83 (18.2%) | 170 (21.5%) | 93 (31.1%) | 88 (13.9%) | 103 (21.0%) | 27 (4.4%) | 0 (0%) | 903 (15.9%) |
| **Meningitis or Encephalitis** | 42 (6.5%) | 5 (0.7%) | 3 (0.7%) | 21 (2.7%) | 17 (5.7%) | 15 (2.4%) | 59 (12.0%) | 10 (1.6%) | 0 (0%) | 172 (3.0%) |
| **Sepsis** | 157 (24.4%) | 17 (2.5%) | 3 (0.7%) | 9 (1.1%) | 2 (0.7%) | 0 (0%) | 0 (0%) | 1 (0.2%) | 0 (0%) | 189 (3.3%) |
| **Musculoskeletal Infection** | 33 (5.1%) | 11 (1.6%) | 21 (4.6%) | 7 (0.9%) | 0 (0%) | 11 (1.7%) | 13 (2.6%) | 29 (4.7%) | 0 (0%) | 125 (2.2%) |
| **Soft Tissue Infection** | 55 (8.6%) | 92 (13.6%) | 174 (38.2%) | 27 (3.4%) | 3 (1.0%) | 13 (2.0%) | 7 (1.4%) | 12 (1.9%) | 0 (0%) | 383 (6.8%) |
| **Gastrointestinal Infection** | 84 (13.1%) | 62 (9.2%) | 47 (10.3%) | 62 (7.8%) | 28 (9.4%) | 69 (10.9%) | 44 (9.0%) | 29 (4.7%) | 0 (0%) | 425 (7.5%) |
| **Urinary Tract Infection** | 204 (31.7%) | 55 (8.1%) | 12 (2.6%) | 11 (1.4%) | 2 (0.7%) | 0 (0%) | 0 (0%) | 3 (0.5%) | 0 (0%) | 287 (5.1%) |
| **Other Diagnosis** | 81 (12.6%) | 34 (5.0%) | 39 (8.6%) | 40 (5.1%) | 50 (16.7%) | 82 (12.9%) | 160 (32.6%) | 72 (11.6%) | 0 (0%) | 558 (9.8%) |
| **Recieved Antibiotics** | 613 (95.3%) | 671 (99.1%) | 438 (96.1%) | 631 (79.8%) | 196 (65.6%) | 306 (48.2%) | 348 (70.9%) | 357 (57.7%) | 0 (0%) | 3560 (62.8%) |

*Table 4: Clinical characteristics of patients. Ethnicity recorded was the self-reported ethnicity. Maximum CRP and neutrophil count refer to the highest value for each patient during that illness episode. The control group included hospitalized patients with no fever and with conditions considered to be not due to infection or inflammatory disorders. Controls included 35 patients with critical illness and 54 patients with primary or secondary immunodeficiency.*

*PICU: pediatric intensive care unit; numbers show the median and interquartile range (IQR); CRP: C-reactive protein.*

The clinical details of the patients assigned to each diagnostic group and controls are shown in Table 4. A high proportion of children received antibiotics, across all phenotypes (96% for DB, 99% for PB, 71% for DV, 48% for PV and 82% for infections with less confident aetiology).

### 3.3.2 Pathogens detected by local investigations and by centralized molecular pathogen tests

The spectrum of pathogens detected by CMT in febrile children and controls are shown in Figure 5. The most frequent CMT identifications for FC were, in decreasing order, rhinovirus, adenovirus and RSV (respiratory pathogens); HHV7, HHV6b and EBV (blood viruses); and *N. meningitidis*, *S. aureus* and *K. pneumoniae* (blood bacteria).




*Figure 5: Frequency of detection of pathogen DNA or RNA by research molecular tests in cases and controls. Donut plots show the breakdown of the total number of positive molecular pathogen identifications from the research study in cases and controls for respiratory pathogens found in throat swabs, viruses detected in blood, and bacteria and fungi detected in blood.*

### 3.3.3 CMT in febrile cases vs controls

#### *3.3.3.1 Respiratory viruses*

1 257 of 2 762 cases (46%) and 121 of 485 (22%) controls had at least one virus detected in throat swabs. Viruses were more commonly detected in FC than controls (OR 1.37 95% CI: 1.21-1.54). Compared to controls, FC had increased frequency of influenza viruses A and B, parainfluenza 1 and 4, bocavirus, rhinovirus, coronavirus OC43, RSV, enterovirus, and adenovirus (Figure 6; Table 5).

#### *3.3.3.2 Blood viruses*

Viruses were detected in blood less commonly in cases than in controls (cases: 2 708/4 175 (65%); controls: 723/999 (72%); OR: 0.85, 95% CI: 0.79 – 0.92). Only CMV and adenovirus were more commonly detected in cases than controls, while HHV7 and HHV6b were more common in controls (Figure 6; Table 6).

#### *3.3.3.3 Blood bacteria and fungi*

Bacterial and fungal DNA was detected in blood at low frequency in both cases and controls (cases: 380/4120 (9.2%); controls: 79/996 (7.9%); OR 1.06, 95% CI 0.92 – 1.20). Odds ratios were overlapping for detecting all bacterial genomes in cases and controls (Figure 6; Table 7).

| Pathogen | Definite Bacterial (n=417) | Probable Bacterial (n=434) | Bacterial Syndrome (n=270) | Unknown Bacterial or Viral (n=425) | Viral Syndrome (n=173) | Probable Viral (n=401) | Definite Viral (n=280) | Other Phenotype (n=362) | Overall Cases (n=2762) | Controls (n=485) |
|---|---|---|---|---|---|---|---|---|---|---|
| Adenovirus | 17 | 41 | 14 | 38 | 28 | 26 | 13 | 14 | 191 (6.04%) | 10 (2.37%) |
| Bocavirus | 10 | 23 | 17 | 30 | 12 | 21 | 14 | 11 | 138 (4.36%) | 8 (1.9%) |
| C.pneumoniae | 3 | 3 | 0 | 1 | 1 | 2 | 0 | 2 | 12 (0.38%) | 3 (0.71%) |
| Coronavirus 229E | 8 | 9 | 3 | 9 | 5 | 14 | 12 | 2 | 62 (1.96%) | 17 (4.03%) |
| Coronavirus HKU1 | 1 | 1 | 0 | 2 | 2 | 3 | 2 | 3 | 14 (0.44%) | 3 (0.71%) |
| Coronavirus NL63 | 5 | 7 | 2 | 4 | 1 | 5 | 5 | 5 | 34 (1.07%) | 4 (0.95%) |
| Coronavirus OC43 | 5 | 7 | 6 | 4 | 3 | 13 | 9 | 4 | 51 (1.61%) | 4 (0.95%) |
| Enterovirus | 11 | 7 | 12 | 32 | 18 | 33 | 19 | 6 | 138 (4.36%) | 2 (0.47%) |
| Influenza A | 8 | 11 | 10 | 15 | 8 | 28 | 34 | 7 | 121 (3.83%) | 4 (0.95%) |
| Influenza B | 3 | 8 | 1 | 7 | 5 | 21 | 26 | 9 | 80 (2.53%) | 4 (0.95%) |
| M.pneumoniae | 25 | 6 | 8 | 5 | 1 | 3 | 0 | 2 | 50 (1.58%) | 2 (0.47%) |
| Metapneumovirus | 3 | 17 | 3 | 11 | 8 | 10 | 11 | 2 | 65 (2.06%) | 5 (1.18%) |
| Parainfluenza 1 | 3 | 5 | 1 | 4 | 5 | 9 | 3 | 2 | 32 (1.01%) | 2 (0.47%) |
| Parainfluenza 2 | 1 | 2 | 3 | 6 | 1 | 5 | 2 | 3 | 23 (0.73%) | 3 (0.71%) |
| Parainfluenza 3 | 2 | 12 | 5 | 10 | 4 | 18 | 3 | 5 | 59 (1.87%) | 4 (0.95%) |
| Parainfluenza 4 | 4 | 3 | 5 | 5 | 1 | 4 | 4 | 4 | 30 (0.95%) | 2 (0.47%) |
| Rhinovirus | 82 | 101 | 43 | 107 | 39 | 64 | 52 | 58 | 546 (17.26%) | 71 (16.82%) |
| RSV | 4 | 10 | 12 | 31 | 22 | 27 | 31 | 11 | 148 (4.68%) | 8 (1.9%) |
| L.pneumophila | 0 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 7 (0.22%) | 2 (0.47%) |
| Total Detections | 390 | 548 | 292 | 644 | 330 | 616 | 480 | 302 | 1801 (56.94%) | 158 (37.44%) |

*Table 5: Frequency of pathogen detection in throat swabs by centralized molecular tests in relation to the clinically-assigned diagnostic category. The frequency of each pathogen's detection by CMT in relation to the patient diagnostic category is shown. The overall cases column is the sum of all detections for that virus, expressed as a percentage of all cases tested. The controls column shows detections for that virus expressed as a percentage of all controls tested.*

| Pathogen | Definite Bacterial (n=569) | Probable Bacterial (n=618) | Bacterial Syndrome (n=406) | Unknown Bacterial or Viral (n=724) | Viral Syndrome (n=274) | Probable Viral (n=568) | Definite Viral (n=454) | Other Phenotype (n=562) | Overall Cases (n=4175) | Controls (n=999) |
|---|---|---|---|---|---|---|---|---|---|---|
| Adenovirus | 5 | 24 | 8 | 23 | 20 | 16 | 10 | 7 | 113 (2.3%) | 6 (0.67%) |
| CMV | 16 | 14 | 3 | 17 | 5 | 13 | 13 | 11 | 92 (1.87%) | 5 (0.55%) |
| EBV | 102 | 125 | 60 | 107 | 42 | 84 | 98 | 67 | 685 (13.94%) | 146 (16.19%) |
| Enterovirus | 15 | 10 | 11 | 35 | 20 | 55 | 40 | 9 | 195 (3.97%) | 13 (1.44%) |
| HHV6a | 5 | 2 | 1 | 4 | 1 | 5 | 0 | 5 | 23 (0.47%) | 5 (0.55%) |
| HHV6b | 148 | 176 | 127 | 185 | 87 | 189 | 93 | 119 | 1124 (22.87%) | 286 (31.71%) |
| HHV7 | 225 | 270 | 169 | 277 | 100 | 223 | 140 | 237 | 1641 (33.39%) | 528 (58.54%) |
| Parechovirus | 2 | 5 | 2 | 6 | 4 | 6 | 7 | 3 | 35 (0.71%) | 4 (0.44%) |
| Parvovirus | 37 | 32 | 19 | 31 | 13 | 33 | 23 | 23 | 211 (4.29%) | 43 (4.77%) |
| Total Detections | 1,110 | 1,316 | 800 | 1,370 | 584 | 1,248 | 848 | 962 | 4119 (83.82%) | 1036 (114.86%) |

*Table 6: Frequency of virus detection in blood by centralized molecular tests in relation to the clinically-assigned diagnostic category. The frequency of each pathogen's detection in relation to the patient diagnostic category is shown. The overall cases column is the sum of all detections for that virus, expressed as a percentage of all cases tested. The controls column shows detections for that virus expressed as a percentage of all controls tested.*

| *Pathogen* | *Definite Bacterial (n=567)* | *Probable Bacterial (n=614)* | *Bacterial Syndrome (n=404)* | *Unknown Bacterial or Viral (n=706)* | *Viral Syndrome (n=268)* | *Probable Viral (n=564)* | *Definite Viral (n=448)* | *Other Phenotype (n=549)* | *Overall Cases (n=4120)* | *Controls (n=996)* |
|---|---|---|---|---|---|---|---|---|---|---|
| A.baumanii | 1 | 6 | 3 | 2 | 1 | 4 | 3 | 4 | 24 (0.48%) | 10 (1.06%) |
| E.aerogenes | 3 | 2 | 1 | 4 | 0 | 1 | 2 | 0 | 13 (0.26%) | 2 (0.21%) |
| E.cloacae | 5 | 5 | 1 | 2 | 1 | 2 | 2 | 1 | 19 (0.38%) | 3 (0.32%) |
| E.coli | 8 | 8 | 4 | 4 | 1 | 3 | 1 | 5 | 34 (0.69%) | 9 (0.95%) |
| E.faecalis | 4 | 10 | 2 | 2 | 2 | 2 | 1 | 3 | 26 (0.52%) | 6 (0.63%) |
| E.faecium | 8 | 6 | 1 | 3 | 3 | 2 | 2 | 3 | 28 (0.56%) | 3 (0.32%) |
| Group A Streptococcus | 9 | 6 | 3 | 2 | 1 | 0 | 1 | 0 | 22 (0.44%) | 5 (0.53%) |
| Group B Streptococcus | 2 | 6 | 0 | 1 | 2 | 1 | 3 | 1 | 16 (0.32%) | 3 (0.32%) |
| H.influenzae | 1 | 5 | 3 | 3 | 2 | 3 | 3 | 0 | 20 (0.4%) | 1 (0.11%) |
| K.kingae | 1 | 3 | 0 | 2 | 0 | 0 | 0 | 0 | 6 (0.12%) | 0 (0%) |
| K.pneumoniae | 9 | 9 | 2 | 6 | 4 | 9 | 3 | 8 | 50 (1.01%) | 12 (1.27%) |
| N.meningitidis | 26 | 10 | 5 | 5 | 2 | 4 | 2 | 0 | 54 (1.09%) | 7 (0.74%) |
| P.aeruginosa | 5 | 6 | 6 | 4 | 5 | 0 | 4 | 3 | 33 (0.66%) | 6 (0.63%) |
| S.aureus | 13 | 6 | 2 | 5 | 4 | 9 | 9 | 0 | 48 (0.97%) | 15 (1.58%) |
| S.marcescens | 6 | 11 | 4 | 5 | 4 | 3 | 2 | 9 | 44 (0.89%) | 6 (0.63%) |
| S.pneumoniae | 19 | 8 | 3 | 6 | 2 | 0 | 1 | 1 | 40 (0.81%) | 3 (0.32%) |
| Aspergillus Sp. | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 3 (0.06%) | 0 (0%) |
| Candida Sp. | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 3 (0.06%) | 0 (0%) |
| Total Detections | 240 | 222 | 80 | 116 | 68 | 86 | 78 | 76 | 483 (9.73%) | 91 (9.61%) |

*Table 7: Frequency of bacterial/fungal detection in blood by centralized molecular tests in relation to the clinically-assigned diagnostic category. The frequency of each pathogen's detection in relation to the patient diagnostic category is shown. The overall cases column is the sum of all detections for that pathogen, expressed as a percentage of all cases tested. The controls column shows detections for that pathogen expressed as a percentage of all controls tested.*

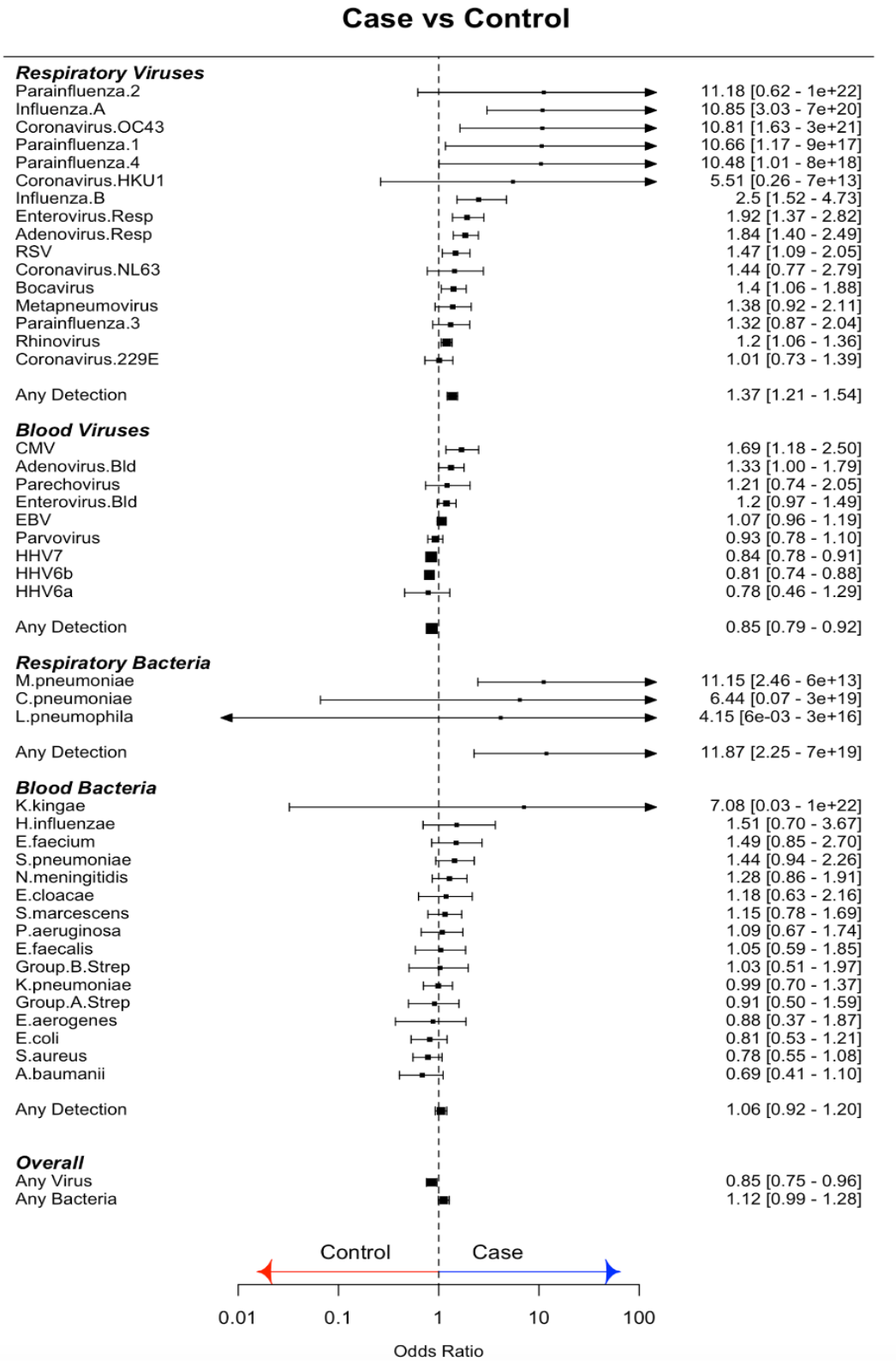



*Figure 6: Odds ratios for molecular pathogen identifications comparing febrile children and non-febrile controls. Square symbols show the mean and bars and whiskers the 95% confidence intervals. Odds Ratio's above 1 (to right of the dotted line; blue arrow) indicate increased detection of pathogen in cases. OR less than 1 (left of dotted line; red arrow) indicate increased pathogen detection in controls.*

### 3.3.4 CMT in Definite Bacterial vs Definite Viral groups

I compared pathogen detection by CMT in the patients assigned a DB or a DV phenotype, to investigate which molecular identifications were reliable predictors of a bacterial or a viral illness.

#### *3.3.4.1 Respiratory pathogens*

Influenza A and B, and RSV were more commonly detected in DV than DB, whilst *Mycoplasmoides pneumoniae* was more common in DB (Figure 7). When the Probable Bacterial (PB) and Probable Viral (PV) groups were analysed in combination with the DB and DV groups, enterovirus and coronavirus 229E were more common in the combined DV/PV group and rhinovirus was found more commonly in the DB/PB combined group (OR: 1.36, 95% CI: 1.14 – 1.61) (Figure 8).

#### *3.3.4.2 Blood viruses*

Enterovirus and EBV were more common in the DV cases, and HHV6b was more common in DB (OR 1.24, 95% CI 1.00 – 1.53) (Figure 7). When the PB and PV cases were included, only enterovirus was more common in the DV/PV combined group (OR: 0.34, 95% CI: 0.21 - 0.50) (Figure 8).

#### *3.3.4.3 Blood bacteria*

*N. meningitidis* (OR: 3.37, 95% CI: 1.92 – 5.99), *S. pneumoniae* (OR: 3.89, 95% CI: 2.07 – 7.59), Group A streptococcus (OR 2.73, 95% CI 1.13 – 6.09) and *E. coli* (OR 2.7, 95% CI 1.02 – 6.71) were more commonly detected in DB than DV cases (Figure 7). When DB and PB cases were considered together, *Enterococcus faecium* (OR: 1.95, 95% CI: 1.00 – 3.52) and *Serratia marcescens* (OR: 1.83, 95% CI: 1.01 – 3.09) were also found more commonly in the bacterial cases (Figure 8).

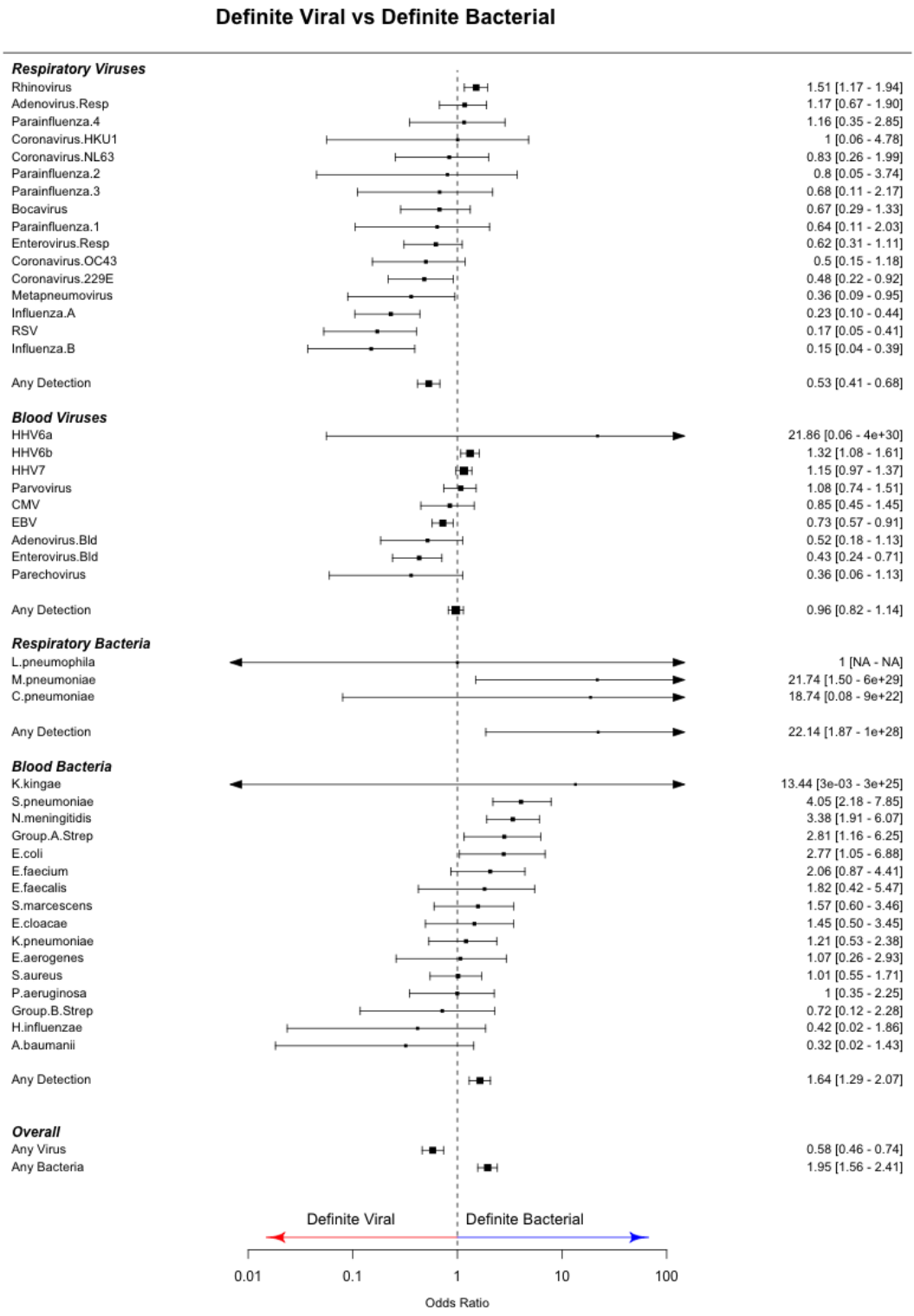



*Figure 7: Odds ratios for molecular pathogen identifications comparing definite bacterial and definite viral groups. Square symbols show the mean and bars and whiskers the 95% confidence intervals. Odds Ratio's above 1 (to right of the dotted line; blue arrow) indicate increased detection of pathogen in DB; OR less than 1 (left of dotted line; red arrow) indicate increased pathogen detection in the viral group.*

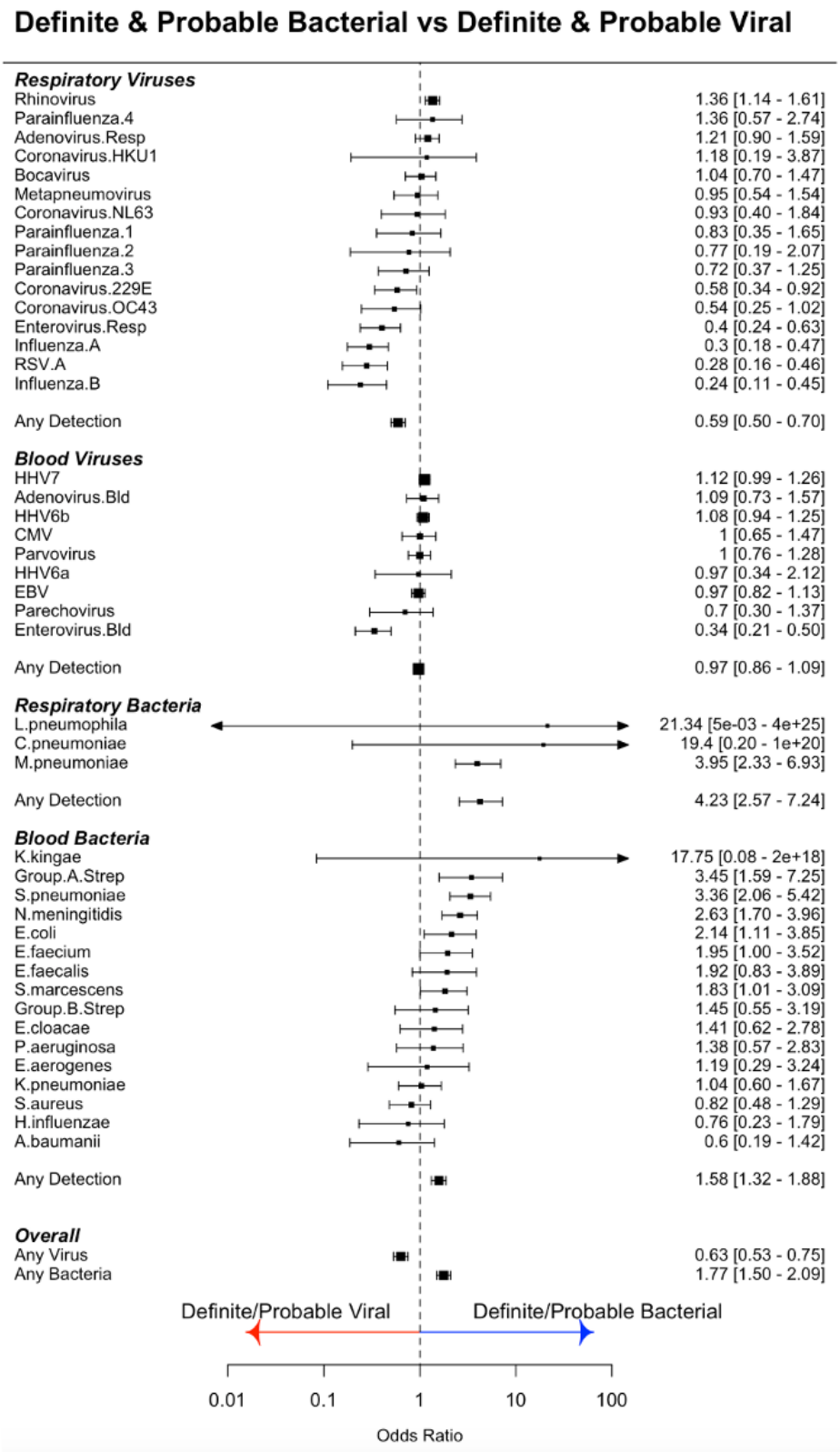



*Figure 8: Odds ratios for molecular pathogen identifications in combined Definite and Probable Bacterial groups compared to Definite and Probable Viral groups. OR above 1 indicates more frequent detection in a patient with bacterial infection and OR below 1 indicates increased detection in those with viral infection. Whiskers around the OR represent the 95% confidence interval.*

### 3.3.4 Quantitative CMT

To explore whether quantitative load of pathogen sequences improved the distinction between DB and DV groups, I established quantitative counts for the blood viruses. Enterovirus, EBV and HHV6b showed a trend towards increased frequency of detection in cases than controls at higher viral loads (Figure 9), whilst enterovirus, parvovirus and EBV showed this trend in the comparison between DV or DB cases (Figure 10). There was no relationship between Ct values of any of the bacterial assays and the odds ratio of detection in cases vs controls or DB vs DV cases.

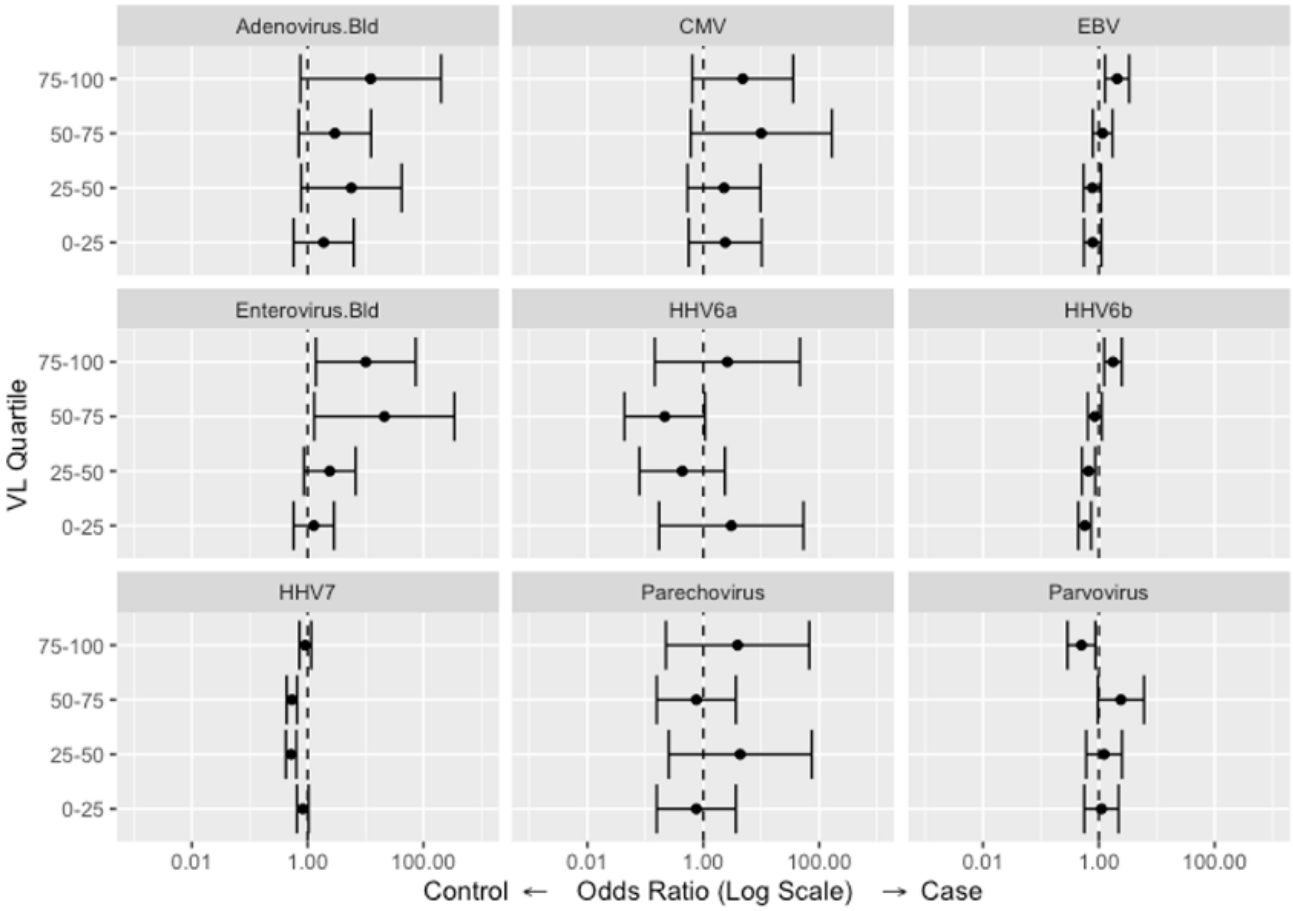



*Figure 9: Quantitative molecular detection by increasing viral load quartile – blood viruses in relation to cases vs controls. Odds ratio with 95% CIs for detection of blood viruses in cases versus controls at increasing levels of viral load.*

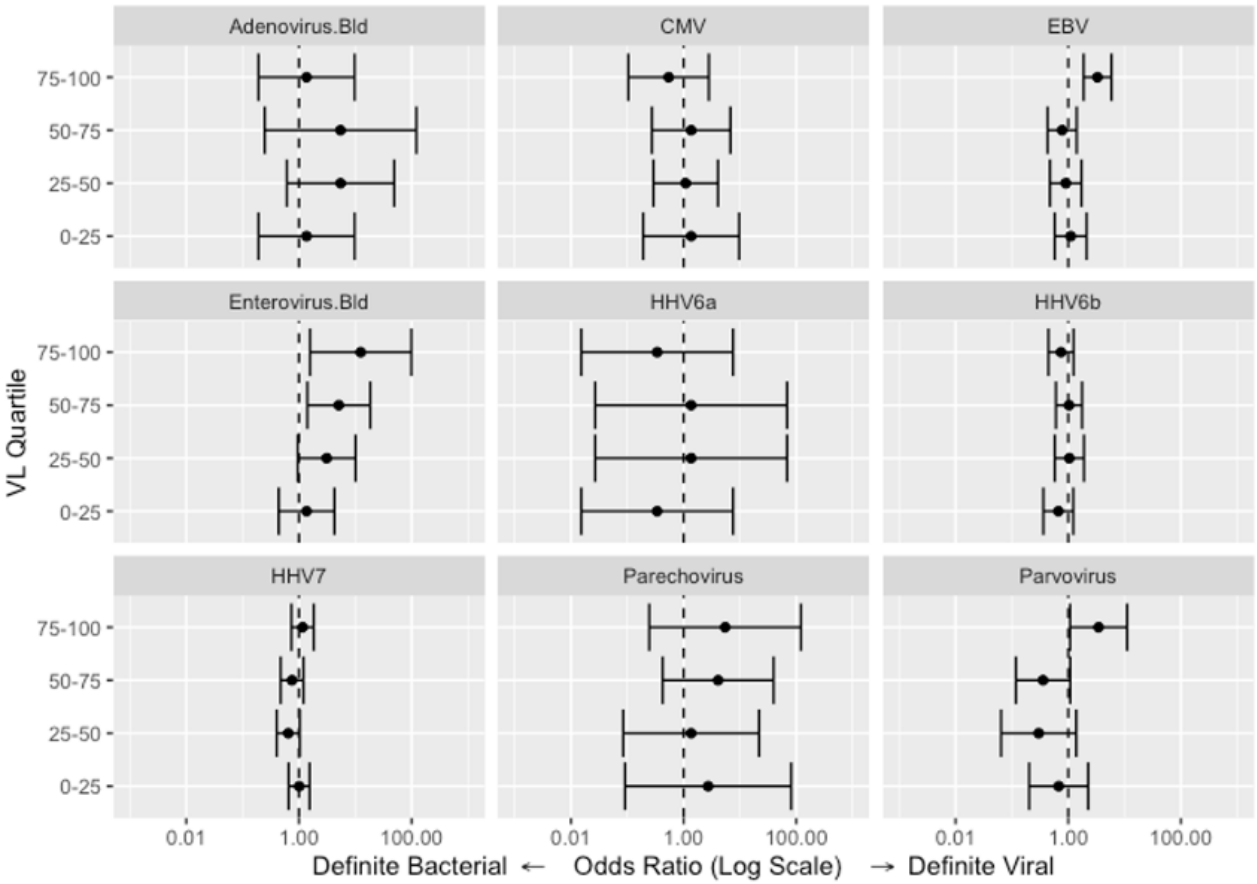



*Figure 10: Quantitative molecular detection by increasing viral load quartile – blood viruses in relation to Definite Viral vs Definite Bacterial diagnostic categories. Odds ratio for detection of blood viruses in Definite Viral versus Definite Bacterial cases, at increasing levels of viral load.*

### 3.3.5 Combination of CMT and local diagnostics

I examined the proportion of patients in each diagnostic category identified by combined local investigations and CMT (Figure 11). Overall, 25% of DB and 32% of PB had detected respiratory viruses, and 56% and 62% of DB and PB had detected blood viruses. The detection of a virus in blood or throat swab had predictive values of 0.64 and 0.68 respectively for excluding a diagnostic category DB or PB. The detection of some viruses, such as Influenza A (0.82), Influenza B (0.85), RSV (0.89) and Enterovirus (0.84) on throat

swabs, as well as Enterovirus (0.85) in blood samples showed superior predictive values in excluding a DB or PB diagnostic category.

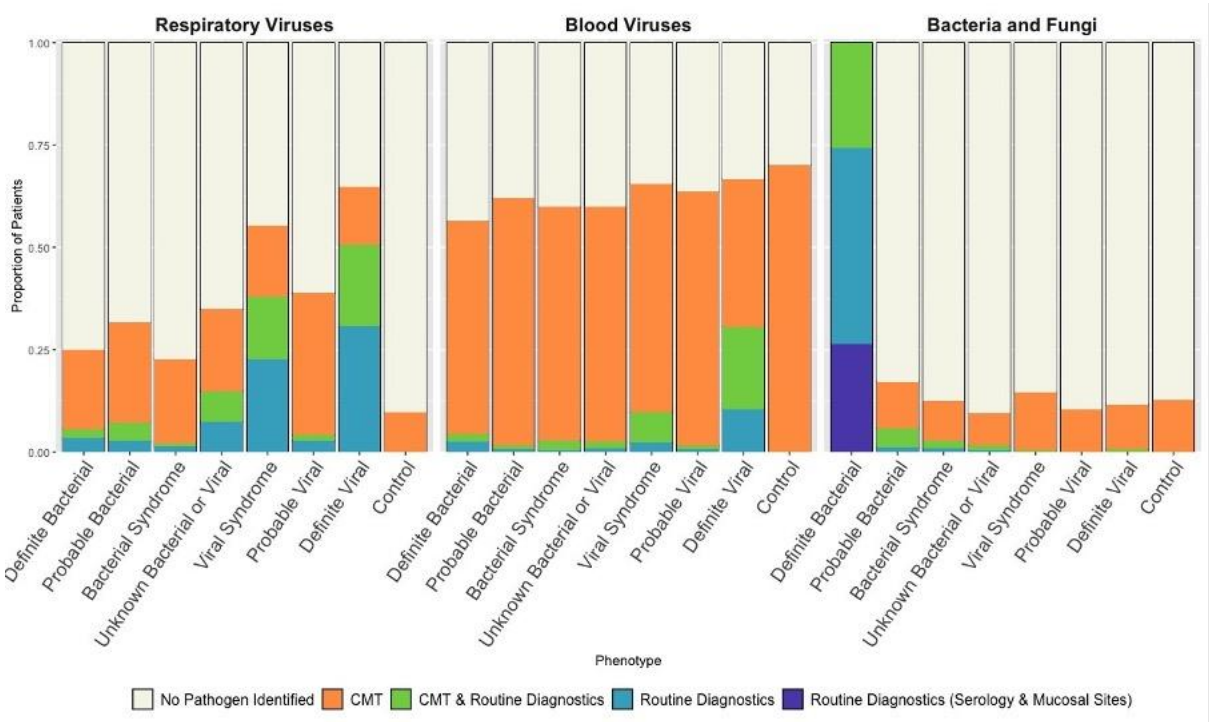



*Figure 11: Proportion of patients in each diagnostic category with pathogen detected. Blue bars show pathogens detected by local investigations; green bars show detection by both local and centralized molecular testing (CMT); orange bars show pathogen detection only by CMT. Purple bar shows patients without sterile site bacterial identification but classified as DB on the basis of mucosal detection of specific pathogens. Combined coloured bar height indicates proportion of all patients in each diagnostic category with pathogens detected. Individual panels show detection of respiratory viruses (left), blood viruses (middle) and bacterial and fungal pathogens (right). Definite viral classification included patients with pathognomic features of a viral infection even if the causative virus was not identified (e.g. typical rash in varicella zoster infection).*

In order to evaluate how the CMT results might change assignment to DB or DV or other groups, I reassigned patients from the uncertain categories (unknown bacterial or viral, PB, PV, viral syndrome or bacterial syndrome) based on the CMT results only for pathogens showing significant odds ratios for distinguishing bacterial and viral infection (*N.meningitidis*, *S.pneumoniae*, group A streptococcus, *E.coli* and *M.pneumoniae* for bacteria; influenza A and B, RSV for respiratory viruses; and EBV and enterovirus for blood viruses). Only 87 patients (3.2 %) were reassigned as DB, and 53 (2.0%) as DV (Figure 12).

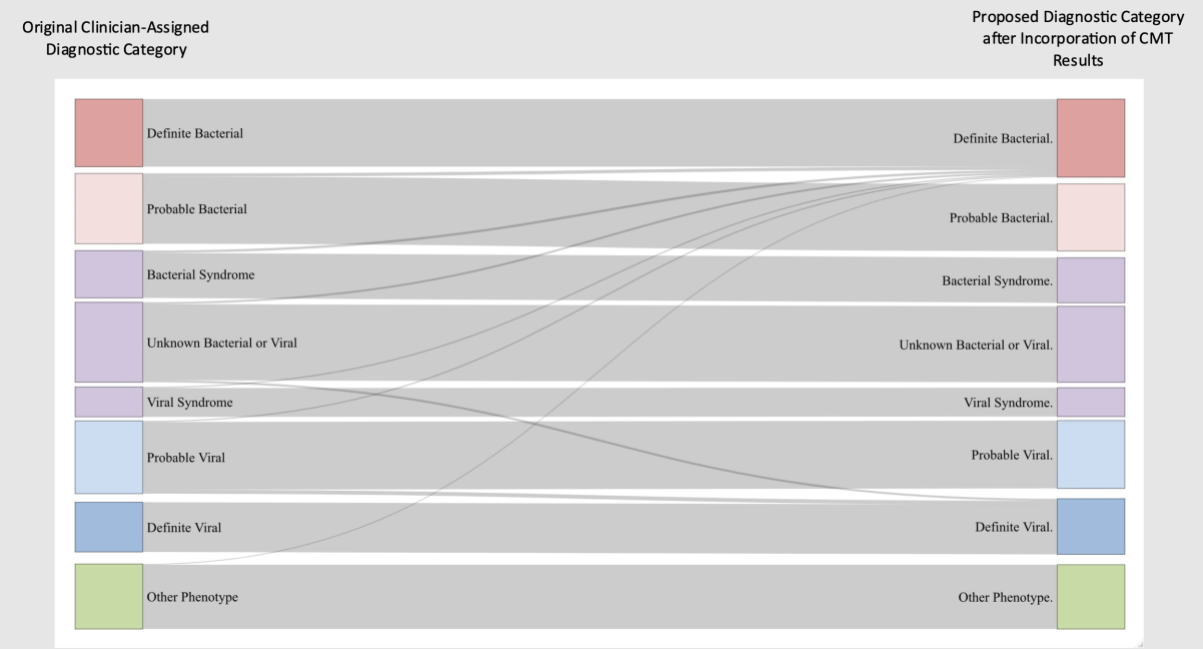



*Figure 12 Sankey diagram showing patient reassignment from final diagnostic category based on local investigations alone, to a final diagnostic group based on addition of CMT. The height of each coloured bar represents number of patients in each group, and the curves crossing groups indicate patients reassigned based on CMT results. Addition of CMT resulted in an addition of 87 (3.2%) of patients to the definite bacterial group and 53 (2.0%) of patients to the definite viral group.*

## 3.4 Discussion

The application of CMTs on blood and respiratory samples of febrile patients provided a great amount of additional pathogen data compared to conventional diagnostics alone.

Viruses were detected in respiratory samples more commonly in febrile children than in controls overall, but influenza, RSV, enterovirus, and coronavirus OC43 were the only viruses with significantly greater odds of detection in the DV than DB group. Enterovirus and EBV were the only blood viruses more common in children with DV than DB. Therefore, virus detection in blood or respiratory tract had a poor predictive value for excluding bacterial infection.

For bacterial and fungal pathogens, CMT found *N.meningitidis*, *S.pneumoniae*, Group A streptococcus and *E.coli* in blood, and *M.pneumoniae* in throat swabs more commonly in the DB than the DV group. For all other bacteria and fungi, no clear distinction was seen between the rate of detection of bacterial DNA in DB and DV or between cases and controls. Of note, CMT showed HHV6b and HHV7 were more common in controls than in cases, even though the controls had no features suggestive of infection. As herpes viruses such as EBV, CMV, HHV6 and HHV7 establish persistent infection in healthy individuals[73–75], detection of these viruses may represent reactivation and not indicate that they are involved in a current illness. I explored whether viral load differed in children in DV and DB categories. Higher viral load was associated with a higher likelihood of a viral diagnosis for EBV, and HHV6b but not HHV7 or CMV. Higher rates of detection of blood viruses in controls than cases might reflect suppression of persistent viruses by immune mediators released in response to other infections. This hypothesis will need exploration by investigating detection of these viruses in the presence of other pathogens and differences in the host response to infections at different viral loads.

Based on interpretation of available clinical and diagnostic data by local recruiting teams, only 14% had a microbiologically confirmed bacterial infection, and a further 10% had a viral infection with no evidence of bacterial infection. The remaining 3 477 (76%) had differing levels of certainty as to whether bacteria or viruses were the cause of illness. Incorporation of CMT results into the process of assigning patients to a diagnostic category increased the assignment as DB to 16%, and as DV to 12%. Although CMT greatly increased the number of patients with detected viruses, they were detected in all diagnostic categories suggesting

that a high proportion of patients with proven or probable bacterial infections are co-infected with viruses. Our inability to clearly distinguish patients with bacterial from viral infection, even after detailed investigation in a prospective research study, was reflected in the clinical management of the patients. A high proportion of all patients in the study (80%), including those with confirmed (77%) or probable (55%) viral infections, were prescribed antibiotics by the medical teams responsible for their care.

Our inability to clearly distinguish bacterial from viral infections even with extensive molecular pathogen detection approaches has several possible explanations. Many common bacterially mediated infections such as pneumonia, abdominal or bone and soft tissue infection remain localized, and are associated with bacteraemia only in severe cases. Molecular detection of bacterial DNA in blood may be no more sensitive than culture in these cases, and this may reflect the restriction for organism detection of the small (200ul) volume of blood available for molecular testing. Many of the bacterial and fungal targets detected were found equally in cases and controls, and in DB and DV, and their clinical significance is therefore unclear. In contrast to the difficulty in detecting bacteria in blood, viruses are readily detected using molecular methods on mucosal surfaces or blood. In addition to children with typical viral infections, viruses were detected in a high proportion with proven or probable bacterial infections.

A high proportion of patients with confirmed bacterial infections also had viruses identified in blood or throat swabs. Our findings challenge the paradigm that bacteria or viruses cause distinct disease processes, and that identification of viruses should be the basis for treatment decisions such as commencing or discontinuing antibiotics. Respiratory viruses

were found more commonly in febrile patients than in non-febrile controls, it is likely that viral and bacterial infection are not independent, in line with the increasing evidence that viral infections set in motion pathophysiological processes that may increase susceptibility to bacterial infection.

Respiratory viruses damage and inflame the airway mucosa, reduce mucociliary clearance, facilitate bacterial adherence and allow migration or invasion of the airway microbiome into deeper tissue[76–79]. Epidemiological studies have shown that peaks of bacterial pneumonia and invasive bacterial infection follow seasonal and pandemic upsurges of influenza[80–85], with bacterial coinfection implicated in severe influenza-associated disease and death[76,86]. Similarly, RSV seasonality has been linked to meningococcal disease[87], pneumococcal pneumonia[88–91] and bacteraemia[92]. Studies suggest pneumococcal vaccine reduces influenza-associated hospitalisations[93–97] and severity of other viral respiratory infections[98], and influenza vaccine has been reported to reduce invasive bacterial infections. *In vitro* studies show a wide range of immunological impairments following viral infection, including impaired pathogen recognition and decreased phagocytosis[99,100]. In murine models, influenza coinfection is associated with increased pneumococcal and staphylococcal bacterial loads and disease[101–104]. Thus, the finding that a large proportion of patients in our study with viruses detected also had evidence of bacterial infection supports the view that both viruses and bacteria are contributing to the disease process.

Several previous studies have attempted to determine the aetiology of fever or pneumonia in children, including studies in low and middle- income countries[105–109]. The PERCH[110], Drakenstein[106] and GABRIEL[108] pneumonia studies attempted to define the role of bacterial

and viral pathogens in childhood pneumonia, and D'Acremont *et al.* explored aetiology of outpatient febrile illness in Tanzania[105]. These studies highlighted the high proportion of febrile episodes associated with viral infection and noted that many cases had multiple pathogens detected. Our study combines a detailed molecular pathogen approach with a detailed patient-by-patient clinical categorization performed independently of the CMT data and enables comparison of the performance of current clinical and research pathogen detection approaches in diagnosing or excluding bacterial infection.

This analysis has a number of limitations. The non-febrile control patients were not matched for age, sex, time of year and country, but were an opportunistic cohort of hospitalized children admitted for conditions which did not appear to have an underlying infectious basis. In order to best match the febrile patients, I included a broad range of non-infectious hospital admissions, including orthopaedic conditions, and intensive care admissions for surgical, cardiac, or traumatic illnesses, and non-febrile children with cancer undergoing immunosuppressive treatments. They may not be truly representative of the spectrum of viruses and pathogens carried by healthy children in the community. Secondly, our approach applied all CMT assays to all patients, unlike clinical practice in which pathogen tests are selected on an individual patient basis, based on judgement of the likely cause of illness. Whilst our comprehensive approach enabled us to look at pathogen burden in patients, including an examination of pathogens not typically associated with the presenting illness, it did not measure the diagnostic utility of these assays when they are used in individual patients. This study shows that, despite the additional pathogen data these techniques provide, molecular pathogen detection alone or in addition to conventional diagnostic techniques is not adequate to resolve the clinical question of which febrile patients need

antibiotics to treat potentially life-threatening bacterial infections, and when antibiotics can be safely withheld. New approaches are needed, both to understand what may be a dynamic process of infection involving both bacteria and viruses, and to identify the population of febrile patients in whom antibiotics are beneficial, and those where antibiotics are not helpful.

This study shows that the development of infection is a more complex process than what is portrayed by conventional diagnostic techniques. It is important to fully understand the additional pathogen data that CMTs reveal and its relationship to clinical disease so that novel diagnostics can be developed and validated in models that represent the true complexity of infectious diseases in children.

# 4. Patterns of Pathogen Detection and Co-detection

## 4.1 Introduction

When analysing the CMT results of European children with a diverse range of infections, it became clear that the detection of multiple pathogens in a single patient was not an uncommon phenomenon. The pattern of pathogen detection did not always fit an expected pattern, with bacteria detected in children with presumed viral disease, viruses detected in patients with presumed bacterial detection and both bacteria and viruses detected in some controls.

### 4.1.1 Pathogen Detection in Different Clinical Syndromes

Some clinical syndromes are associated with specific pathogens or pathogen groups, such as RSV with bronchiolitis[111], *N.meningitidis, S.pneumoniae* and *H.*influenzae with meningitis[112] and *E.coli* with urinary tract infections[113]. To maximise diagnostic yield and efficiency of both time and resources, and to minimise the potential side-effects of unnecessary investigations, clinicians will often limit their investigations to look for positive confirmation of pathogens associated with the patient's presenting syndrome.

While this is likely the best approach for patients and health services, when challenging the current model of infection using the additional pathogen data from CMT results, these assumed relationships between pathogens and clinical syndrome should also be challenged.

### 4.1.2 Patterns of Pathogen Co-Detection and Their Impact on Infection

A number of reports have implicated interactions between specific combinations of pathogens in the aetiology of specific clinical syndromes, such as asthma exacerbations in

patients with co-detection of rhinovirus and *S.pneumoniae*[114]. In some cases, including bacterial superinfection following influenza infection[115], a direct link between the presence of one pathogen and concurrent or subsequent infection with a second pathogen has been described. In other cases, the chronology or mechanism of co-infection is unclear, but a clear association between the two pathogens exists.

### 4.1.3 Aims

Here I examine how the presence of individual pathogens or groups of pathogens is related to the clinical syndrome and other features of a child's illness. To do this I will:

1. Examine whether there are any common patterns of pathogen co-detection.
2. Examine whether the detection of individual pathogens is related to specific clinical syndromes.
3. Examine whether the co-detection of any combinations of pathogens is related to any differences in clinical syndrome, compared to if the pathogens were detected in isolation.

## 4.2 Methods

To assess whether the occurrence of co-detections was uniform across febrile children and controls and in each clinical syndrome, a two-sided two-proportions Z-test was applied to assess whether the proportion of children with co-detections was significantly higher or lower than the background proportion of children with pathogen detections in controls, cases and in DB and DV subgroups. The odds of co-detection of two pathogens compared to individual detection of each pathogen was calculated using a multiple logistic regression model with iteratively re-weighted least squares using the R package glmnet[72].

## 4.3 Results

The number of children recruited with each clinical syndrome is shown in Table 4.

The most common syndromes assigned were in the upper and lower respiratory tract and gastrointestinal infection groups.

### 4.3.1 Pathogen Detection in Clinical Syndromes

#### *4.3.1.1 Upper Respiratory Tract Infections*

In the 1032 patients with upper respiratory tract infections (URTIs) and non-specific URTIs, HHV7 (n=375/1032, 36%), HHV6b (n=314/1032, 30%), EBV (n=184/1032, 18%) and Rhinovirus (n=122/1032, 12%) were the most commonly detected pathogens in patients with upper respiratory tract infections (URTIs) and non-specific URTIs (n=467/1032, 45%), tonsilitis (n=303/1032, 29%) and pharyngitis (n=195/1032, 19%) were the most commonly recorded syndromes (Figure 13). Patients with upper respiratory tract infections were

significantly more likely to have multiple pathogens detected than a single pathogen (OR: 1.44, 95% CI: 1.24-1.68).

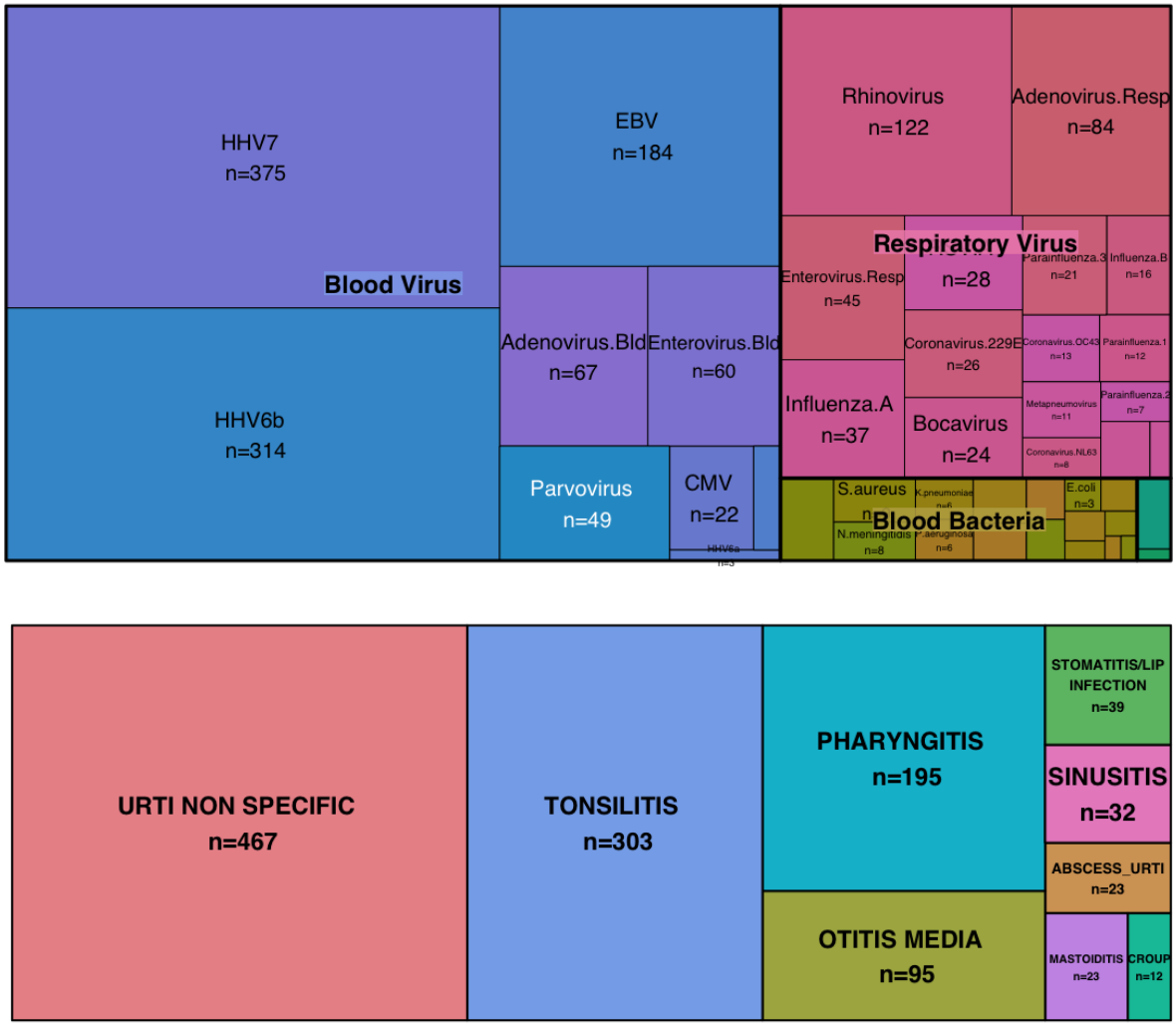



*Figure 13: Treemaps showing pathogen detections and clinical syndromes in patients with upper respiratory tract infections. Size of each rectangle indicates the number of detections of each pathogen or the number of patients with each syndrome in patients with URTIs.*

Patients had significantly greater odds of having tonsilitis if adenovirus (OR 5.85, 95% CI: 4.03-8.50) or EBV (OR: 1.96, 95% CI: 1.49-2.58) was detected in respiratory samples or adenovirus (OR: 5.67, 95% CI: 3.63-8.87) was detected in blood (Figure 14). Detection of enterovirus (OR: 6.43, 95% CI: 2.47-16.7) in respiratory samples was associated with greater odds of stomatitis and HHV6b (OR: 5.15, 95% CI: 2.51-10.6) was associated with greater

odds of sinusitis. Pharyngitis was significantly more likely in patients where parainfluenza 2 (OR: 8.80 95% CI: 2.84-27.2), coronavirus 229E (OR: 8.07, 95% CI: 4.48-14.5), adenovirus (OR: 4.65, 95% CI: 2.93-7.40) and enterovirus (OR: 3.85, 2.2-6.73) were detected in respiratory samples and adenovirus (OR: 5.86, 95% CI: 3.51-9.80) was detected in blood. The detection of HHV6b (OR: 2.58, 95% CI: 1.71-3.88) in blood was associated with greater odds of otitis media.

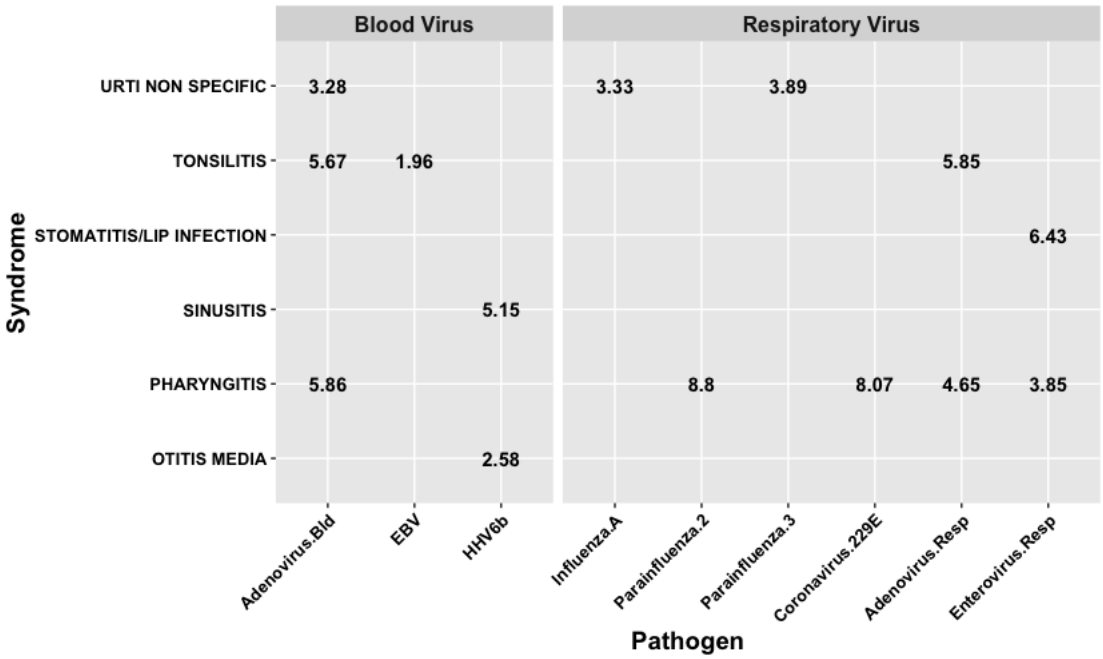



*Figure 14: Odds ratios of pathogen detection in patients with upper respiratory tract infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed.*

When adenovirus and HHV6b were co-detected, there were greater odds of a patient having otitis media (OR: 5.14, 95% CI: 2.31-11.4) and pharyngitis (OR: 6.88, 95% CI: 3.98-11.9). Although the odds ratios for these syndromes in the presence of both adenovirus and HH6b appeared to be higher than those for each pathogen in isolation, this finding was not statistically significant.

Overall, there were greater odds of a patient having an URTI when parainfluenza 3 (OR: 3.19, 95% CI: 1.82-5.60), coronavirus 229E (OR: 2.64, 95% CI: 1.62-4.30), adenovirus (OR: 4.38,

95% CI: 3.23-5.94), influenza A (OR:2.26, 95% CI: 1.52-3.38) or enterovirus (OR: 2.41, 95% CI: 1.67-3.49) were detected in respiratory samples and adenovirus (OR: 6.25, 95% CI: 4.31-9.05), enterovirus (OR: 1.87, 95% CI: 1.38-2.55) or HHV6b (OR: 1.43, 95% CI: 1.24-1.67) were detected in blood (Figure 15).

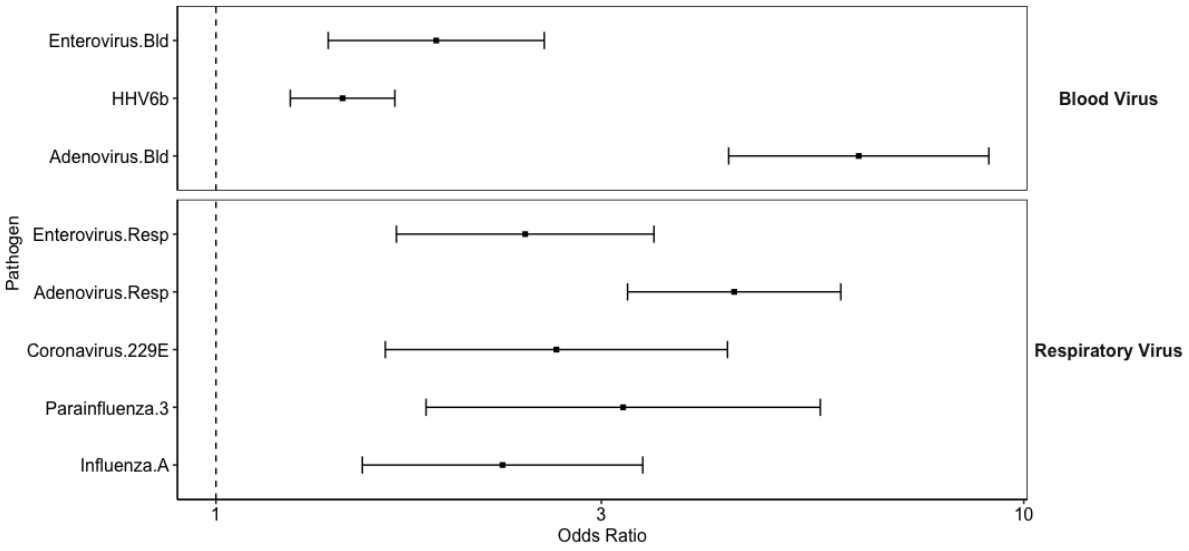



*Figure 15: Odds ratios for pathogen detection in patients with upper respiratory tract infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed.*

#### *4.3.1.2 Lower Respiratory Tract Infections*

In patients with lower respiratory tract infections (LRTIs), HHV7 (n=292/903, 32%), HHV6b (n=212/903, 24%), rhinovirus (n=138/903, 15%) and EBV (n=133/903, 15%) were the most common pathogens detected and pneumonia (n=485/903, 54%), undefined LRTI (n=233/903, 26%) and bronchiolitis (n=120/903, 13%) were the most commonly recorded syndromes (Figure 16). Multiple pathogen detections were significantly more likely in patients with LRTIs (OR: 1.47, 95% CI: 1.25-1.74).

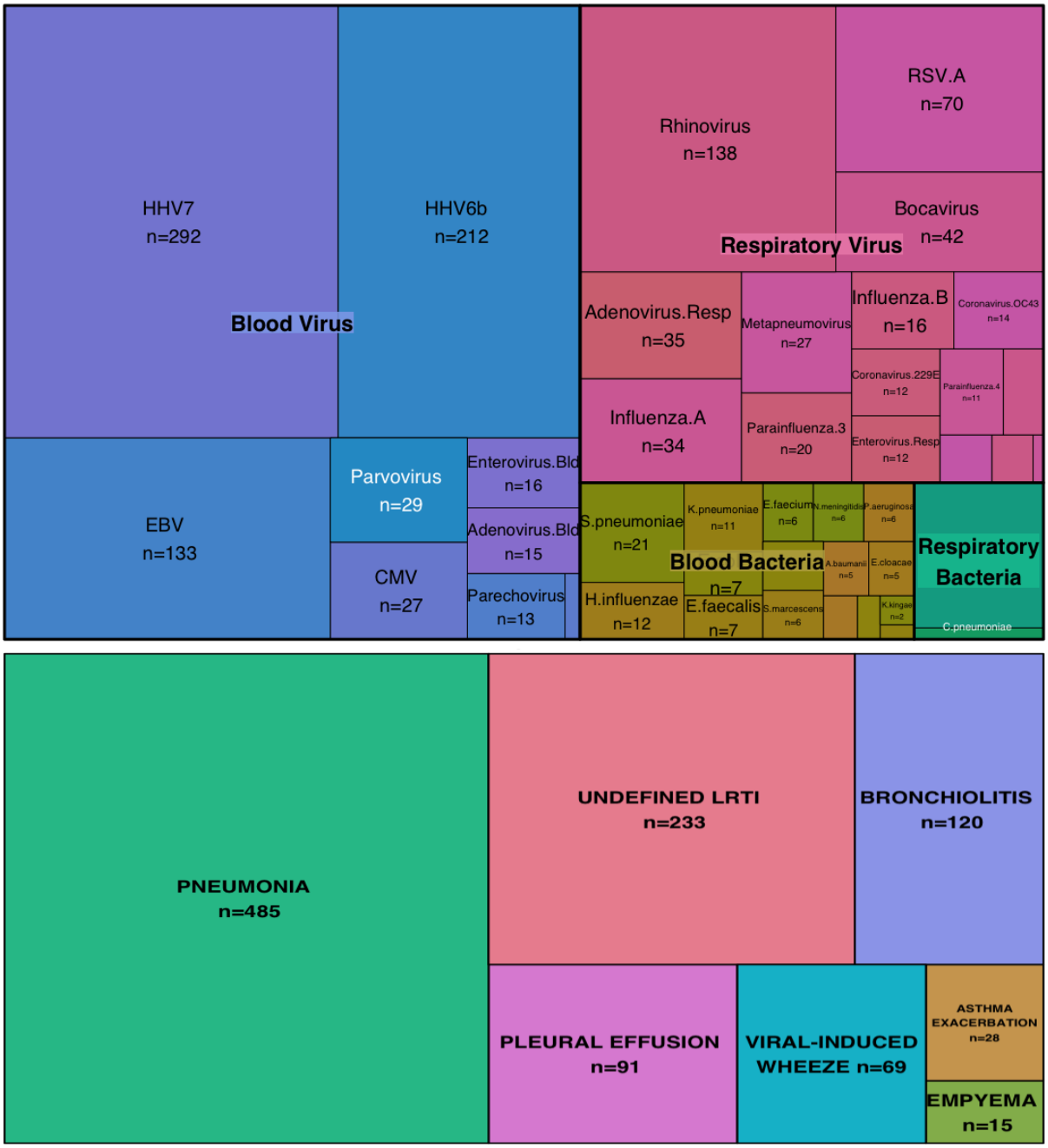



*Figure 16: Treemaps showing pathogen detections and clinical syndromes in patients with lower respiratory tract infections. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen.*

Bocavirus was associated with greater odds of a patient having viral-induced wheeze (OR: 5.39, 95% CI: 2.41-12.0) and pneumonia (OR: 3.32, 95% CI: 2.15-5.12) (Figure 17). RSV was associated with greater odds of pneumonia (OR: 3.52, 95% CI: 2.30-5.40), bronchiolitis (OR: 16.9, 95% CI: 10.5-27.1) and undefined LRTI (OR: 3.75, 95% CI: 2.18-6.45). *S.pneumoniae* and

*M.pneumoniae* were associated with greater odds of pneumonia (OR: 5.67, 95% CI: 3.14-10.3; OR: 40.3, 95% CI: 19.8-81.8 respectively) and pleural effusion (OR: 10.7, 95% CI: 4.69-24.5; OR: 19.6, 95% CI: 9.37-40.8 respectively). *S.pneumoniae* was also associated with greater odds of empyema (OR: 44.4, 95% CI: 13.6-144). Metapneumovirus was associated with greater odds of pneumonia (OR: 3.82, 95% CI: 2.0-7.08) and bronchiolitis (OR: 7.26, 95% CI: 3.21-16.4). Bronchiolitis was also more likely when *H.influenzae* was detected (OR: 18.5, 95% CI: 6.5-52.8), and less likely when HHV7 was detected (OR: 0.13, 95% CI: 0.07-0.26).

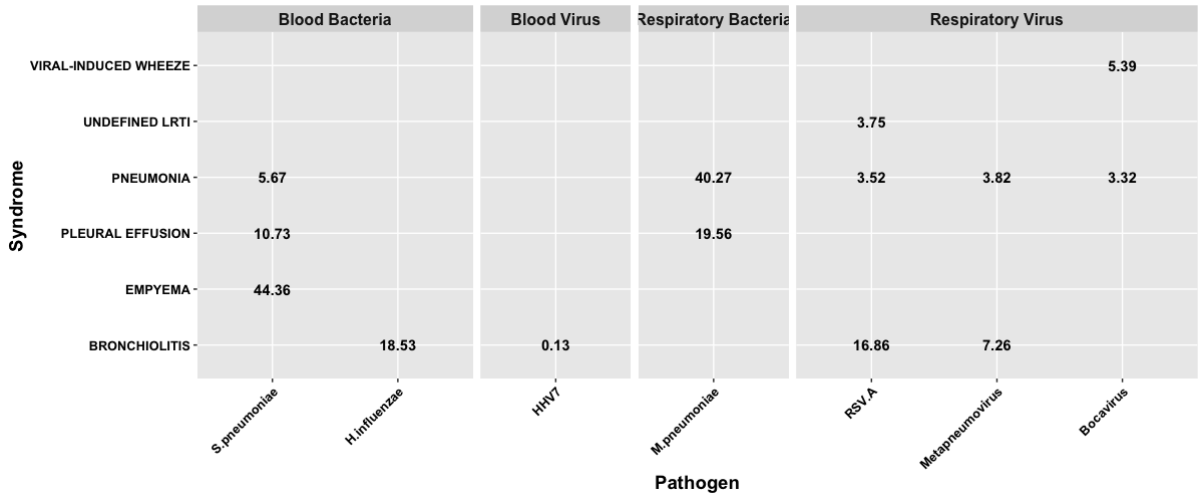



*Figure 17: Odds ratios of pathogen detection in patients with lower respiratory tract infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed.*

When both RSV and rhinovirus were detected, patients were more likely to have bronchiolitis (OR: 24.7, 95% CI: 10.4-58.9) than other febrile children. When bocavirus and influenza A were co-detected, there were greater odds of a patient having pneumonia (OR: 10.8, 95% CI: 3.11-37.4). Once again, although there appeared to be greater odds of these syndromes in the presence of these pathogens together, as opposed to individually, this finding was not statistically significant.

Overall, lower respiratory tract infections were more likely when *H.influenzae* (OR: 10.7, 95% CI: 4.00-28.6) or *S.pneumoniae* (OR: 3.89, 95% CI: 2.21-6.86) were detected in blood and *M.pneumoniae* (OR: 29.9, 95% CI: 13.3-67.1), rhinovirus (OR: 1.91, 95% CI: 1.55-2.35), bocavirus (OR: 2.86, 95% CI: 1.95-4.18), metapneumovirus (OR: 5.41, 95% CI: 3.16-9.27), parainfluenza 3 and 4 (OR : 3.46, 95% CI: 1.96-6.10; OR: 4.51, 95% CI: 2.01-10.1 respectively), RSV (OR: 7.77, 95% CI: 5.38-11.2) and Influenza A (OR: 2.35, 95% CI: 1.56-3.54) were detected in respiratory samples (Figure 18).

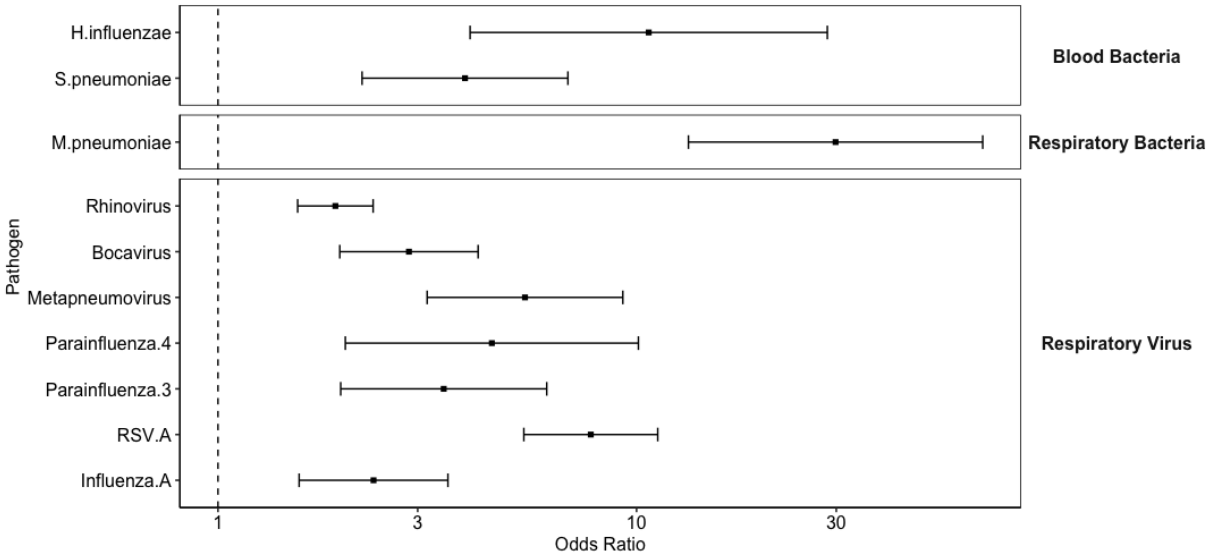



*Figure 18: Odds ratios for pathogen detection in patients with lower respiratory tract infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed.*

#### *4.3.1.3 Gastrointestinal Infections*

In patients with a gastrointestinal infection, moderate/severe diarrhoea (n=285/425, 67%) and appendicitis (n=92/425, 22%) were the most common syndromes recorded (Figure 19). There was no significant difference between multiple and single pathogen detection in patients with gastrointestinal infections.

No organisms showed significant association with gastrointestinal syndromes, or gastrointestinal infections overall.

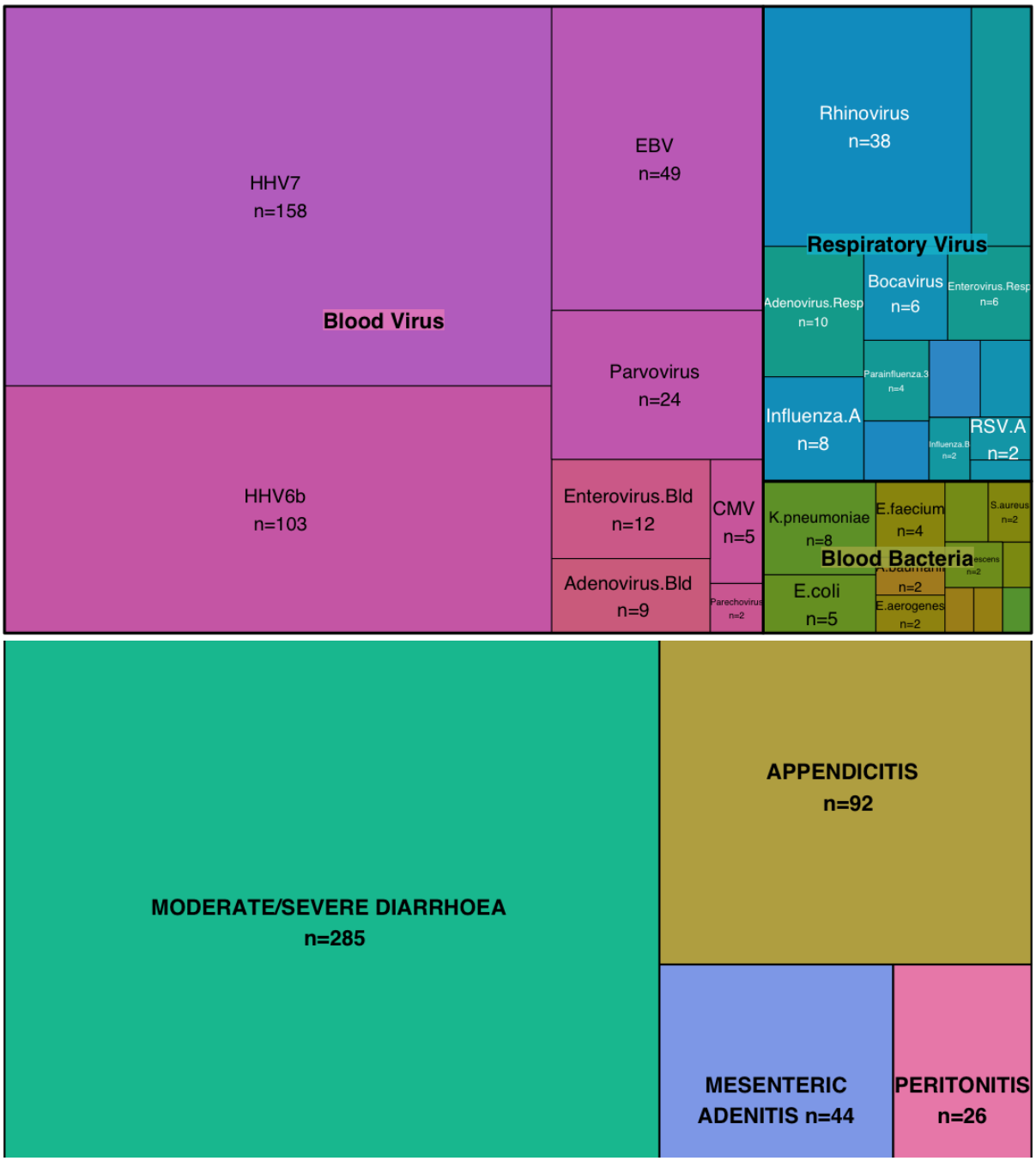



*Figure 19: Treemaps showing pathogen detections and clinical syndromes in patients with gastrointenstinal infection.*

*Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen.*

#### *4.3.1.4 Central Nervous System Infections*

In patients with central nervous system (CNS) infections, HHV7 (n=36/172, 21%), HHV6b (n=21/172, 18%), enterovirus in both blood (n=31/172, 18%) and respiratory samples (n=21/172, 12%) and EBV (n=21/172, 12%) were the most common pathogens detected and meningitis (n=131/172, 76%) was the most common syndrome recorded (Figure 20). Patients with CNS infections were not more likely to have multiple pathogens detected.

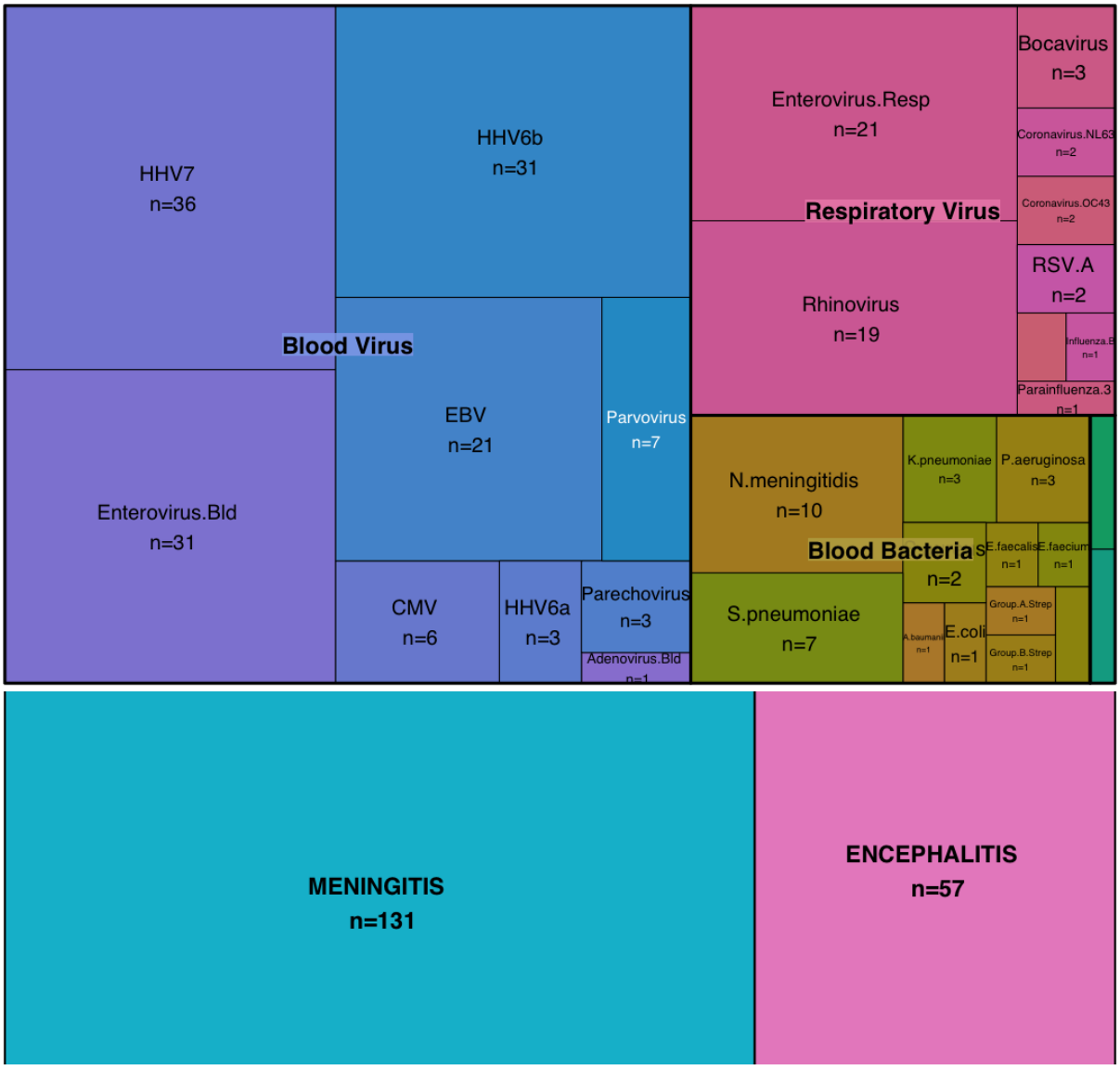



*Figure 20: Treemaps showing pathogen detections and clinical syndromes in patients with central nervous system infections. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen.*

There were greater odds of meningitis in patients with enterovirus detected in respiratory samples (OR: 9.43, 95% CI: 5.70-15.6) and enterovirus (OR: 8.95, 95% CI: 5.80-13.8), *N.meningitidis* (OR: 8.81, 95% CI: 4.22-18.4) and *S.pneumoniae* (OR: 2.51, 95% CI: 2.51-14.3) (Figure 21). When HHV7 was detected, there was significantly lower odds of a patient having meningitis (OR: 0.38, 95% CI: 0.25-0.60).

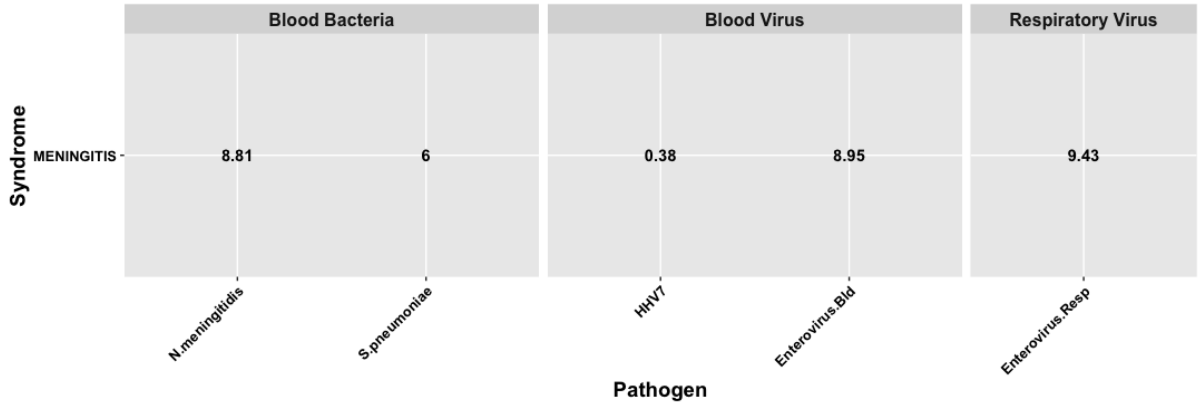



*Figure 21: Odds ratios of pathogen detection in patients with central nervous system infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed.*

CNS infection was more likely when *N.meningitidis* (OR: 7.48, 95% CI: 3.71-15.1), enterovirus (OR: 6.61, 95% CI: 4.16-11.2) or *S.pneumoniae* (OR: 5.38, 95% CI: 2.39-12.1) were detected in blood or enterovirus (OR: 6.81, 95% CI: 4.16-11.2) was detected in respiratory samples (Figure 22). There were significantly lower odds of CNS infection in patients with HHV7 (OR: 0.43, 95% CI: 0.30-0.63).

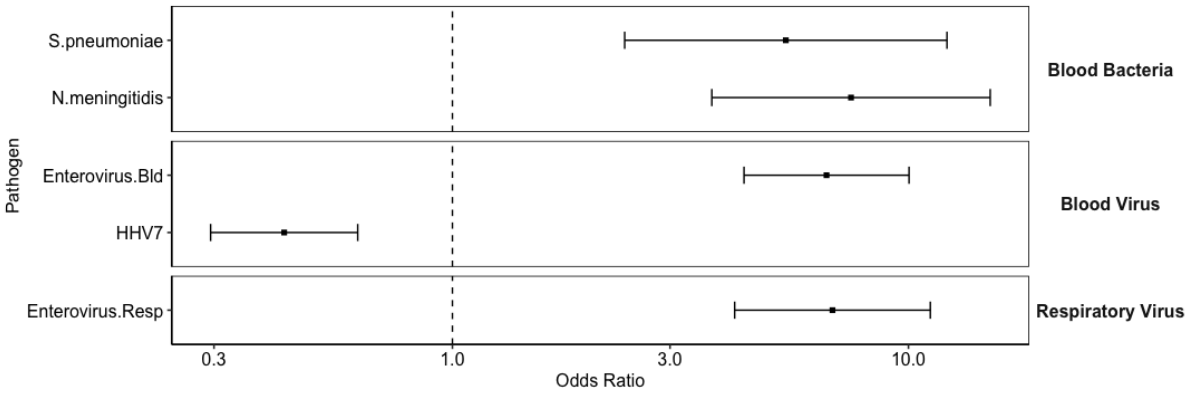



*Figure 22: Odds ratios for pathogen detection in patients with central nervous system infections. Only odds ratios that are statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed.*

### *4.3.1.5 Urinary Tract Infections*

The most common pathogens detected in patients with urinary tract infections (UTIs) were HHV7 (n=92/287, 32%), HHV6b (n=81/287, 28%), EBV (n=43/287, 15%) and rhinovirus (n=41/287, 14%). 159/287 (55%) patients had lower UTIs and 140/287 (49%) had pyelonephritis. 12/287 (4%) had both lower UTI and pyelonephritis (Figure 23).

Patients with UTIs were not more likely to have multiple pathogens detected and there were no significant associations between the detection of any pathogen and UTIs.

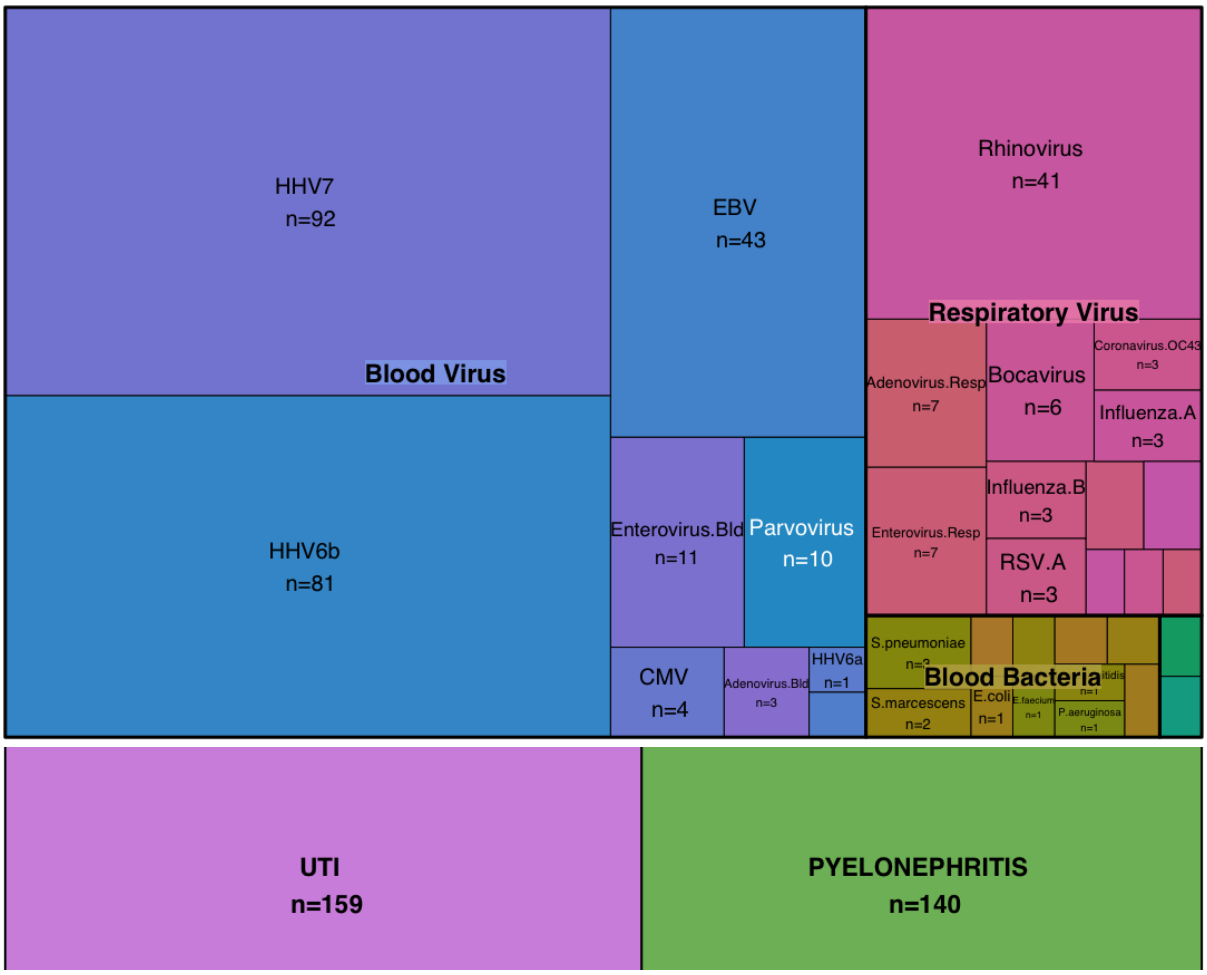



*Figure 23: Treemaps showing pathogen detections and clinical syndromes in patients with urinary tract infections.*

*Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen.*

#### *4.3.1.6 Musculoskeletal Infections*

In patients with musculoskeletal infections, osteomyelitis (n=42/125, 34%), reactive arthritis (n=29/125, 23%) and septic arthritis (n=26/125, 20%) were the most commonly recorded syndromes (Figure 24). Patients with musculoskeletal infections were not more likely to have more than one pathogen detected.

Myositis was significantly more likely in patients where influenza B was detected (OR: 29.1, 95% CI: 10.3-82.0).

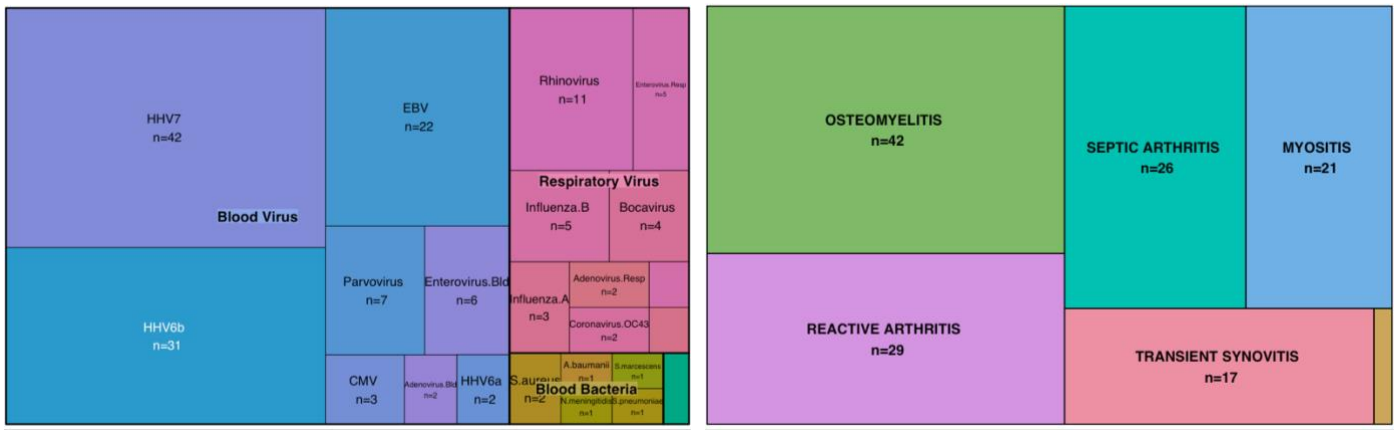



*Figure 24: Treemaps showing pathogen detections and clinical syndromes in patients with musculoskeletal infections. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen.*

#### *4.3.1.7 Skin and Soft Tissue Infections*

HHV7 (n=174/383, 45%), HHV6b (n=108/383, 28%), EBV (n=71/383, 19%) and rhinovirus (n=41/383, 11%) were the most commonly detected pathogens and peri-orbital cellulitis (n=86/383, 23%), lymphadenitis (n=65/383, 17%), soft tissue abscess (n=63/383, 16%) and wound infection (n=45/383, 12%) were the most recorded syndromes in patients with skin and soft tissue infections (Figure 25).

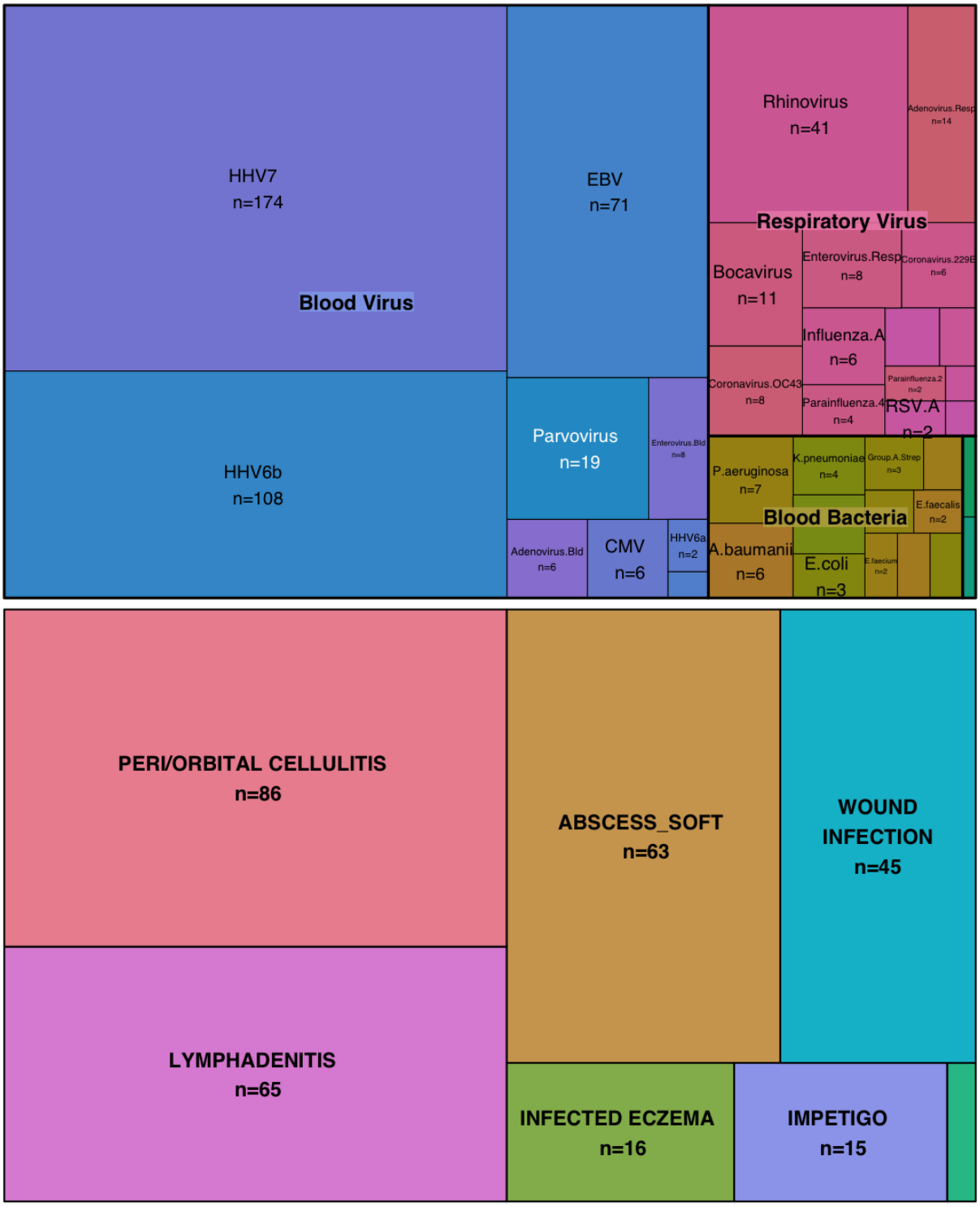



*Figure 25: Treemaps showing pathogen detections and clinical syndromes in patients with skin and soft tissue. Rectangle size represents the number of patients with each clinical syndrome or detection of each pathogen.*

There was no significant difference between the odds of one or more than one pathogen being detected and there were no significant associations between detection of any pathogens and skin and soft tissue infections.

#### *4.3.1.8 Conclusions*

Upper and lower respiratory tract infections and gastrointestinal infections were the most common syndromes in our cohort. Some syndrome groups, such as upper and lower respiratory tract infections were associated with greater odds of multiple pathogen detection, but this was not the case in all syndrome groups.

The detection of pathogens was often associated with greater odds of a clinical syndrome known to be related to that pathogen, such was the case for upper respiratory tract infections in patients with respiratory viruses. However some known pathogen-syndrome relationships, such as *E.coli* and UTI[113] were not demonstrated. This may be due to a lack of power to determine pathogen-syndrome relationships in less frequently detected pathogens, or due to infections at local sites that do not result in sufficient circulating pathogen DNA to result in a positive detection.

There were also some unexpected pathogen-syndrome relationships. For example there were greater odds of sinusitis when HHV6b is detected and bronchiolitis was more common when *H.influenzae* was detected and less likely when HHV7 was detected. These may represent direct or indirect relationships between pathogens and syndromes. The relationship between HHV7 and bronchiolitis may be explained by the theory postulated in chapter 3, that the host anti-viral to another viral pathogen may result in reduced circulating

HHV7, making syndromes typically caused by viruses, such as bronchiolitis, less common when HHV7 is detected.

### 4.3.2 Co-detection of Pathogens

Of 5672 children with CMT results available, no pathogens were detected in 1580/5672 (28%) children (1305/4611 cases and 275/1061 controls), one pathogen was detected in 2008/5672 (35%) children (1570/4611 cases and 438/1061 controls) and more than one pathogen was detected in 2084/5672 (37%) children (1736/4611 cases and 348/1061 controls) (Figure 26). Amongst the cases, a significantly higher proportion had multiple pathogen detections than a single detection (p=0.0003); and cases were significantly more likely to have multiple pathogens detected than controls (OR 1.39, 95% CI: 1.19-1.63).

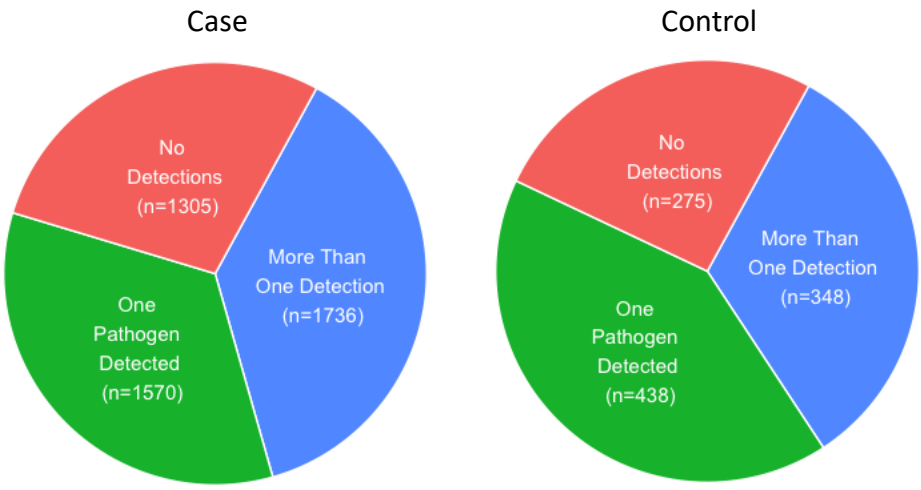



*Figure 26: Proportion of febrile children and controls with no pathogens detected, a single pathogen detection and more than one pathogen detected.*

Figure 27 shows the pathogens that had significantly greater odds of co-detection than either pathogen being detected in isolation.

#### *4.3.2.1 Virus-virus co-detections*

Adenovirus was more likely to be detected in respiratory samples alongside bocavirus (OR: 6.37, 95% CI: 4.69-11.3) or rhinovirus (OR: 2.95, 95% CI: 2.32-4.58) in the same respiratory sample and HHV6b (OR: 2.07, 95% CI: 1.47-2.69) in blood. Adenovirus detection in blood was also associated with detection of bocavirus (OR: 3.48, 95% CI: 1.82-7.20) in respiratory samples and HHV6b (OR: 2.06, 95% CI: 1.40-2.94) in blood. Bocavirus detection was more likely in the presence of a number of respiratory viruses, including influenza A and B (OR: 4.0, 95% CI: 2.25-8.33; OR: 4.02, 95% CI: 1.81-9.64 respectively), parainfluenza 3 (OR: 5.35, 95% CI: 3.02-14.2), coronavirus OC43 (OR: 5.03, 95% CI: 1.98-12.4) and rhinovirus (OR: 2.95, 95% CI: 2.12-4.81). Rhinovirus was also more likely to be detected when coronavirus OC43 (OR: 3.35, 95% CI: 1.64-5.23) and enterovirus (OR: 3.28, 95% CI: 2.22-4.42) were detected. There were greater odds of influenza B and coronavirus 229E (OR: 5.5, 95% CI: 2.37-15.2) to be co-detected than detected in isolation.

In blood, CMV was more likely to detected when parechovirus was detected (OR: 6.4, 95% CI: 2.03-15.9) and EBV was more likely to be detected in patients with parvovirus (OR: 1.65, 95% CI: 1.24-2.14), HHV6b (OR: 1.61, 95% CI: 1.37-1.84) and HHV7 (OR: 2.54, 95% CI: 2.00, 95% CI: 2.64). There were greater odds of HHV7 detection in patients with HHV6b (OR: 1.45, 95% CI: 1.25-1.51) and parvovirus (OR: 1.79, 95% CI: 1.38-1.98) detection.

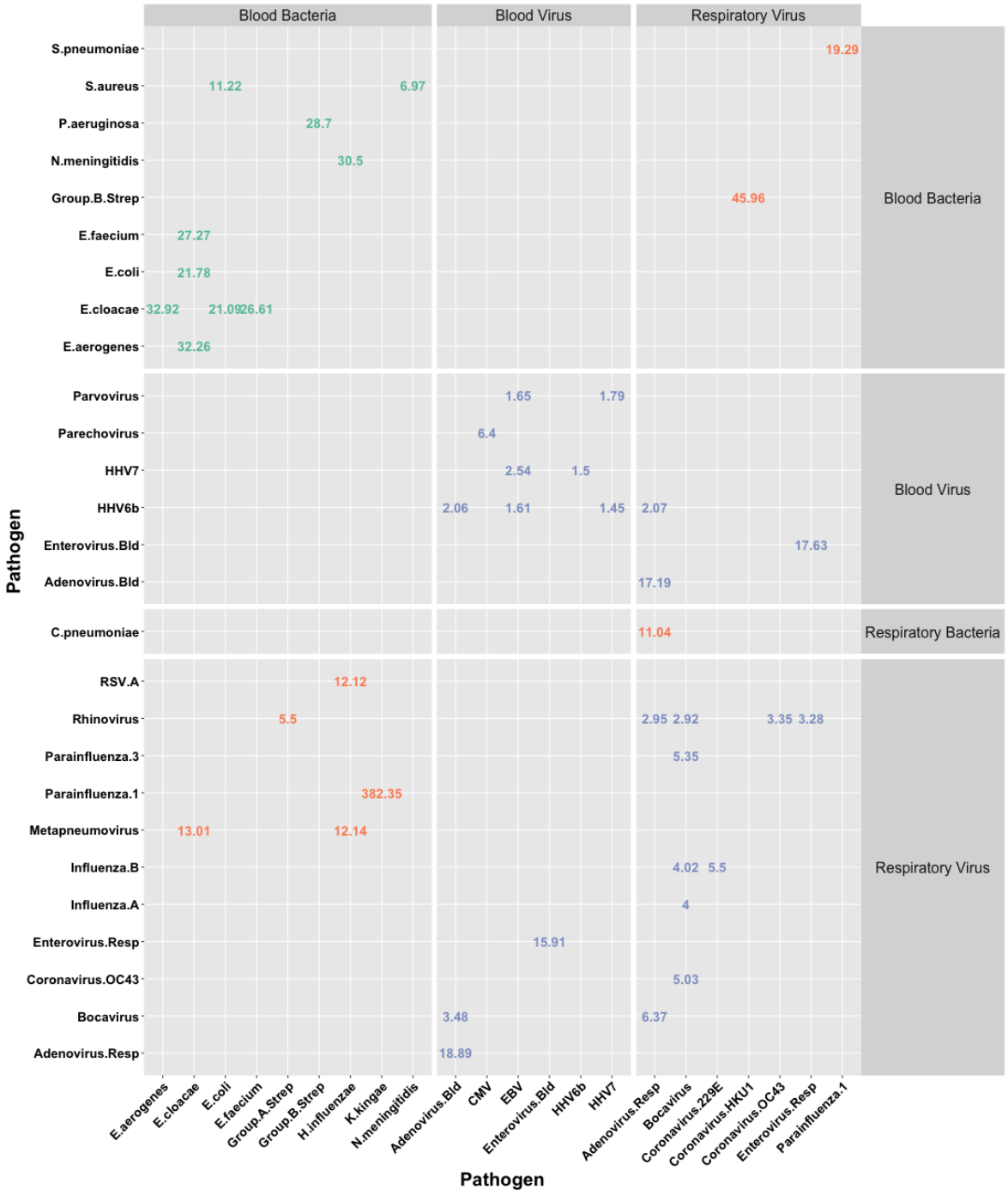



*Figure 27: Odds ratios for co-detection. The odds ratios for viral-viral (blue), bacterial-bacterial (green) and bacterial-viral co-detections in patients, compared to detection of each pathogen in isolation. Only odds ratios that were statistically significant after Benjamini-Hochberg correction for multiple comparisons are displayed. As each pathogen combination could exist in 2 places on this plot, data positions have been arranged to maximise clarity.*

#### 4.3.2.2 Bacterial-Bacterial Co-detections

*N.meningitidis* was more likely to be detected in patients where *H.influenzae* (OR: 30.5, 95% CI: 12.7-83.7) and *S.aureus* (OR: 6.97, 95% CI: 2.28-18.5) were also detected (Figure 27). *P.aeruginosa* was more likely to be detected in patients with group B streptococci (OR: 28.7, 95% CI: 3.59-72.8). *S.aureus* detection was more likely in patients with *E.coli* (OR: 11.2, 95% CI: 3.21-26.8), *E.coli* was also more likely to be co-detected with *E.cloacae* (OR: 21.78, 95% CI: 2.68-53.8) than detected alone and there were greater odds of detection of *E.cloacae* when E.*faecium* was detected (OR: 27.3, 95% CI: 3.89-85.7).

#### 4.3.2.3 Bacterial-Viral Co-detections

In patients with parainfluenza 1, there are greater odds of detection of *S.pneumoniae* (OR: 19.3, 95% CI: 6.42-52.5) and *K.kingae* (OR: 382.35, 95% CI: 45.1-10133) (Figure 27). When metapneumovirus was detected, *H.influenzae* (OR: 12.1, 95% CI: 2.09-46.5) and *E.cloacae* (OR: 13.0, 95% CI: 1.97-43.4) were more likely to be detected. *H.influenzae* was also more likely to be detected in patients with RSV (OR: 12.1, 95% CI: 3.77-37.1) and group B streptococci were more likely to be detected in patients coronavirus HKU1 (OR: 46.0, 95% CI: 2.20-196).

#### 4.3.2.4 Conclusions

When CMTs were used to investigate the pathogens detected in febrile children, multiple pathogen detection was more common than detection of a single pathogen. Multiple pathogen detection was also more common in febrile children compared to controls.

In bacterial co-detections, respiratory and gut commensals were often more likely to be detected alongside commensals from the same site. Viruses were commonly co-detected alongside other viruses.

## 4.4 Discussion

I aimed to investigate whether detection of certain pathogens was associated with specific clinical syndromes and whether the detection of multiple pathogens was random or influenced by the detection of other pathogens. I found that the detection of multiple pathogens was not random, with certain combinations of pathogens more likely to be detected together.

The types of pathogens that were more likely to be co-detected may explain the mechanisms behind their presence. All bacterial-viral pathogen combinations with increased odds of detection together than in isolation involved co-detection of respiratory viruses with bacteria that are known to be upper respiratory tract or gastrointestinal commensals[116,117]. Furthermore, upper and lower respiratory tract infections were associated with greater odds of a patient having multiple pathogen detections. Certain pathogens resulted in greater odds of specific syndromes. Some of these were expected and previously described, while others, such as the relationship between HHV6b and sinusitis and between HHV7 and bronchiolitis were unexpected.

A number of studies offer potential explanations for the observed patterns of pathogen co-detection and pathogen-syndrome relationships. Firstly, physical disruption, for example due to cytopathy secondary to some infections or the host response to them, allows migration of

commensal bacteria into the bloodstream and sites where they can subsequently cause infection, such as migration of an upper respiratory tract commensal bacteria to the lower respiratory tract resulting in pneumonia.[115,118] This hypothesis is supported by the higher proportion of patients with upper and lower respiratory tract infections with multiple pathogen detections than single detections. Co-detection of viruses may be explained by immune dysregulation secondary to infection, resulting in a state, for example a state of low antiviral activity, that is permissive to infection with a subsequent pathogen[119]. A third hypothesis, which may explain the prevalence of multiple pathogen detections in cases, is that the presence of multiple pathogens increases illness severity and therefore increases the likelihood of patient enrolment into the study[33,98].

Since the emergence of molecular diagnostic techniques, many new viruses have been discovered and the role of these viruses in infection is still being understood. It has been postulated that some viruses, such as bocavirus, are bystander pathogens that are often co-detected with other pathogens, but do not have a known pathogenic role on their own[120,121]. However, there is an increasing body of evidence suggesting that bocavirus may have an important role, both when detected alone or with other pathogens, in acute respiratory infections and long-term airway disease following infection[122,123].

It is clear that our current understanding of infectious diseases and the relationship between pathogens is incomplete. This study shows that molecular techniques provide an insight into the complexity of infectious diseases, but they are not able to shed light on the mechanisms that determine the clinical presentations or clinical course of infections. For that we must investigate the host response to infections at different sites and with different pathogens.

# 5. Transcriptomic Similarities between IgA Vasculitis and Kawasaki Disease

## 5.1 Introduction

IgA vasculitis (IgAV), also known as Henoch-Schoenlein Purpura, and Kawasaki Disease (KD) are the two most common vasculitides in children, with estimated annual incidences of 23 per 100,000 children under 17 and 20 per 100,000 children under 5 years of age in the U.S.A. respectively[124]. Both diseases affect boys more than girls and predominantly affect children younger than 5 years of age[125,126]. Despite extensive research, the aetiology of both IgAV and KD remains unknown, though both are suspected to be triggered by microbial agents: IgAV is seasonal, and commonly preceded by an upper respiratory tract infection[127], and KD shows epidemic and seasonal fluctuations in incidence, and has similarities with toxin-mediated diseases[128,129]. However, no responsible triggers have yet been identified.

IgAV is a small vessel leucocytoclastic vasculitis, primarily affecting the skin (100%), joints (75%), intestines (50-75%) and kidneys (40-50%)[130]. The clinical symptoms and signs associated with IgAV result from deposition of IgA1 in vessel walls[131]. By comparison, KD affects medium-sized arteries, including the coronary arteries, and histopathological studies on inflamed arteries demonstrate early neutrophilic infiltration, followed by later T cell infiltration, with CD8 T cells, macrophages and type 1 interferon responses implicated[132,133]. An IgA immune response is also believed to play a role in some patients with KD[134,135], and case reports identify patients with an overlap of symptoms[136]. Both IgAV and KD are usually self-limiting illnesses but can result in sequelae with significant morbidity. IgAV can cause nephritis in approximately 15-20% of patients, with 30-50% of this subset having persistent

renal abnormalities lasting for months or years, and end-stage renal disease in 1-3% of these[130,137,138]. In KD, the most important sequelae is the development of coronary artery aneurysms (CAAs) and it is the most common cause of acquired heart disease in children[139].

A number of studies have demonstrated that in whole blood, specific patterns of gene expression can be associated with particular infectious and inflammatory conditions[55,140,141], including KD[54,142–144]. Whole blood gene expression analysis has led to enhanced understanding of pathogenesis and aetiology of a number of diseases and disorders, including TB[51], cancers[145,146] and inflammatory disorders[147,148]. Here, I have used whole blood transcriptomics to show the close similarity in the host response in children with KD and IgAV, and highlight the biological relevance of the differences in gene expression between the diseases.

## 5.2 Methods

### 5.2.1 Patient groups

This analysis utilises data from patients recruited to studies other than the PERFORM study. Blood samples were collected from 18 children with IgAV, 78 with KD, 61 healthy controls (HC) and 61 febrile controls (FC) with self-limiting viral infections (adenovirus, RSV & influenza) from the United States and the U.K. as part of the UK-based Immunopathology or Respiratory, Inflammatory and Infectious Disease Study (IRIS)[55] and the US-based Kawasaki Disease Research Centre Programme (Figure 28). The studies recruited patients between the 1st March 2009 and 14th November 2013.

IgAV was diagnosed in children presenting with palpable purpura, typically over the buttocks and extensor surfaces, associated with abdominal pain, arthralgia or renal abnormalities. KD was diagnosed on the basis of the American Heart Association criteria for either complete or incomplete KD.[126] Whole blood samples were collected at the time of recruitment and pre-treatment into blood RNA tubes (PAXgene; PreAnalytiX). The samples were then frozen, extracted and analysed using Illumina BeadChip microarrays (HumanHT-12 version 4.0). These data are publicly available on the Gene Expression Omnibus database (Accession number: GSE73464).

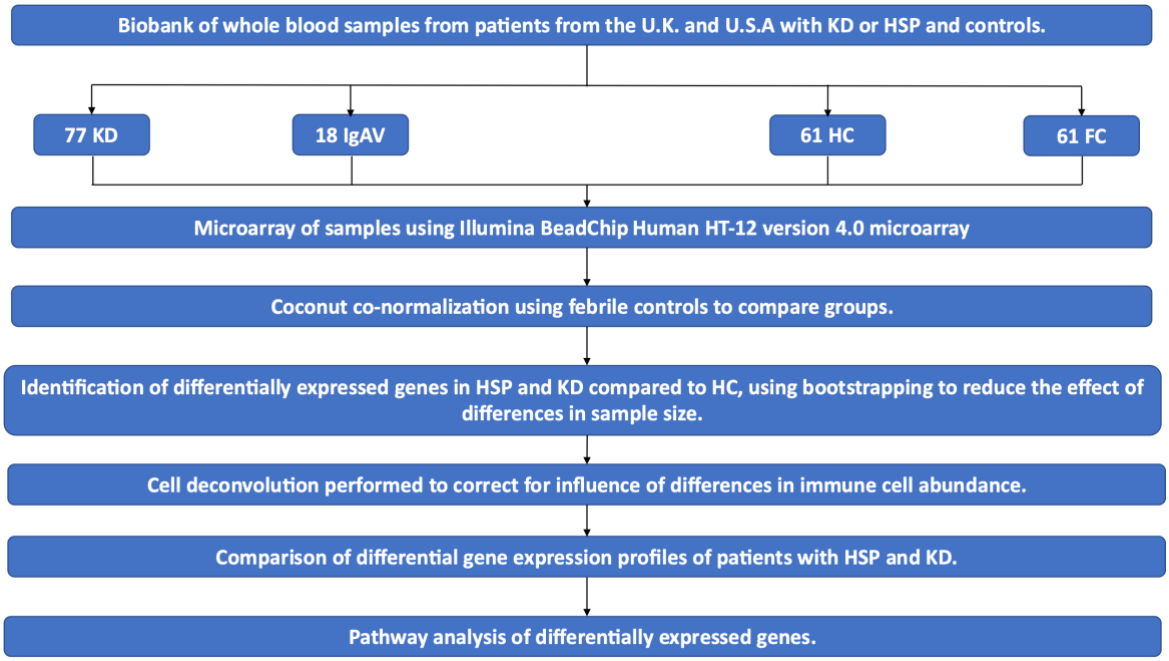



*Figure 28: Breakdown of patient samples and data analysis plan including the number of patients recruited with Kawasaki disease (KD), IgA Vasculitis (IgAV), healthy controls (HC) and febrile controls (FC). There were 93 patients recruited from the U.K. (1 IgAV, 61 HC and 31 FC) and 124 from the U.S.A. (77KD, 17 IgAV and 30 FC).*

### 5.2.2 Ethical Approval

Children were recruited, with written parental informed consent and patient assent as appropriate, under approvals by the research ethics committees of the United Kingdom (St

Mary's Hospital 09/H0712/58, 13/LO/0026) and the University of California San Diego (Human Research Protection Program 140220).

### 5.2.3 Statistical Analysis

The R package lumi[149] was used to apply background subtraction and robust spline normalisation to the raw expression data.

R Coconut co-normalisation method[150] was used to correct for batch effects relating to differences in geographic (U.K. vs U.S.A.) recruitment of patients. Co-normalisation was anchored on transcriptome data from febrile patients with the same community acquired viral infections ('febrile controls' - 30 from the U.S.A and 31 from the U.K.) who were recruited and processed together with the KD and IgAV patients (Figure 29).

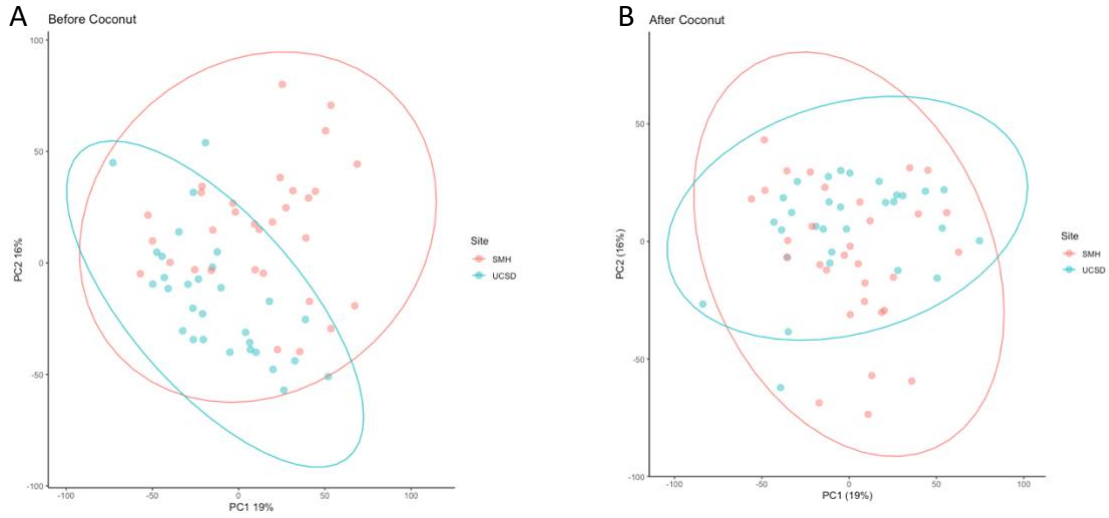



*Figure 29: PCA plot of patients with RSV, Influenza or adenovirus, grouped by Site. A: Before and B: After COCONUT co-normalisation.*

To mitigate for differences in sample size between the comparator groups, when performing differential expression analysis, (77 in KD, 18 in IgAV and 61 in HC), I used a bootstrapping approach by repeatedly randomly selecting 18 samples from each group. The R package

limma[151] was used to identify significantly differentially expressed genes (DEGs) between the groups. This process of random sampling and identification of significantly differentially expressed genes was performed 5,000 times. A p-value of 0.01 after Bonferroni correction was used as the threshold for significance.

Cell deconvolution was performed to correct for the influence of differences in the abundance of neutrophils and lymphocytes on the gene expression profiles of samples. The CIBERSORT analytical tool[152] was used to estimate impact of the abundance of different immune cell types on gene expression and correct for this (Figure 30). Concordant and discordant gene expression patterns between IgAV and KD samples were identified and the discordance/concordance score was calculated for each gene using the R package disco.[153] Gene expression libraries[154,155] and pathway analysis packages[156,157] were used to identify the biological relevance of the DEGs.

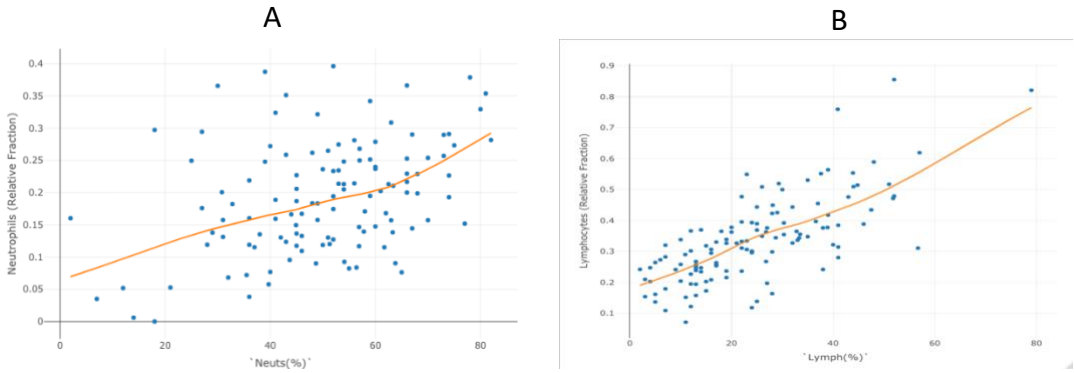



*Figure 30: Comparison of cell abundance estimation using CIBERSORT and clinical haematology results for A: Neutrophils and B: Lymphocytes.*

## 5.3 Results

A total of 218 patients were recruited: 78 with KD, 18 with IgAV, 61 HC and 61 febrile controls with community-acquired viral infections. There was no significant difference in age or gender between the groups (Table 8).

| | | Kawasaki Disease | IgA Vasculitis | Healthy Controls | Febrile Controls |
|---|---|---|---|---|---|
| **No. of patients** | | 78 | 18 | 61 | 61 |
| **Age, median (IQR), months** | | 27 (16 to 45) | 55 (42 to 80) | 33 (18 to 72) | 21 (3 to 54) |
| **Male sex, No. (%)** | | 43 (55.1) | 9 (50) | 32 (52.5) | 41 (67.2) |
| **Recruited in the U.K. (%)** | | 0% | 5.6% | 100% | 50.8% |
| **Illness day at sample collection, median (IQR)** | | 5 (4 - 6) | 4 (2-6) | NA | 5 (4-6) |
| **Laboratory Values, median (IQR)** | | | | | |
| | C-reactive protein, mg/L | 111 (46.5 – 192) | <5 (<5 – 17.8) | NA | 24 (6.3 – 43.5) |
| | While blood cell count, /μL | 14,200 (10,400 – 18,300) | 11,700 (9,700 – 14,100) | 7,300 (6,400 – 10,900) | 7,600 (5,700 – 10,700) |
| | Neutrophil count, /μL | 9,500 (7,150 to 12,300) | 6,500 (4,200 – 8,600) | 2,900 (2,500 – 4,500) | 3,600 (2,580 – 5,600) |
| **Ethnicity, No. (%)** | | | | | |
| | Not Stated | 0 (0) | 1 (5.6) | 15 (24.6) | 4 (6.6) |
| | African | 3 (3.9) | 0 (0) | 6 (9.8) | 6 (9.8) |
| | Asian, including Indian subcontinent and Far East | 12 (15.6) | 0 (0) | 3 (4.9) | 2 (3.3) |
| | European | 20 (26.0) | 4 (22.2) | 10 (16.4) | 27 (44.3) |
| | Hispanic | 25 (32.5) | 4 (22.2) | 8 (13.1) | 14 (22.9) |
| | Mixed | 14 (18.2) | 8 (44.4) | 4 (6.6) | 5 (8.2) |
| | Other | 3 (3.9) | 1 (5.6) | 15 (24.6) | 3 (4.9) |
| **Coronary artery status, No. (%)** | | | | | |
| | Normal | 44 (57.1) | NA | NA | NA |
| | Dilated | 25 (32.5) | NA | NA | NA |
| | Aneurysm | 8 (10.4) | NA | NA | NA |

*Table 8: Clinical characteristics and initial laboratory findings for patients with Kawasaki disease, IgA Vasculitis, febrile controls and healthy controls. Abbreviations: IQR = interquartile range. Aneurysmal coronary arteries were defined as those with z-score of ≥5.*

Dimensionality reduction using principal component analysis (PCA) revealed one clear outlier patient who was removed from further analysis (Figure 31), therefore 77 patients with KD, 18 with IgAV and 61 HC were taken forward for analysis. Coconut co-normalisation was performed using samples from patients with the same seasonal viral infections in the U.K. and U.S.A. to correct for site specific effect (Figure 29).

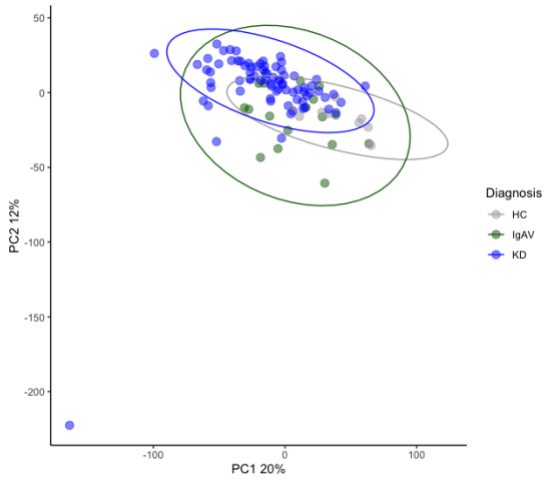



*Figure 31: Primary component analysis identified one clear outlier which was removed from the analysis.*

To compensate for the neutrophilia associated with KD[158], as seen in Table 8, and reveal the transcriptomic similarities or differences that relate to the underlying biological processes associated with KD and IgAV, cell deconvolution was performed. Before the applying deconvolution, comparison of the gene expression profiles of KD and IgAV patients to healthy controls found 3,638 DEGs and 686 DEGs respectively, in 80% of the bootstrapping cycles. Following cell deconvolution and bootstrapping there were 5,027 DEGs between KD and healthy controls, and 4,164 DEGs between IgAV and controls in more than 80% of cycles during bootstrapping. Taken together, a total of 5,084 genes were significantly differentially expressed either between IgAV and HC, and/or KD and HC, and this set was taken forward for further analysis.

### 5.3.1 Transcriptome of IgAV patients

Of the 4,164 DEGs between IgAV and healthy controls, 1,407 (34%) had increased expression in IgAV and 2,752 (66%) had reduced expression. IgAV patients clustered together based on the expression profile of the top 200 DEGs between IgAV and controls (Fig 31A). The DEGs with the highest fold change (LFC <-2.5 or >2.5) and lowest p-value (adjusted p value $<10^{-11}$) (Figure 32B and Table 9) are discussed below.

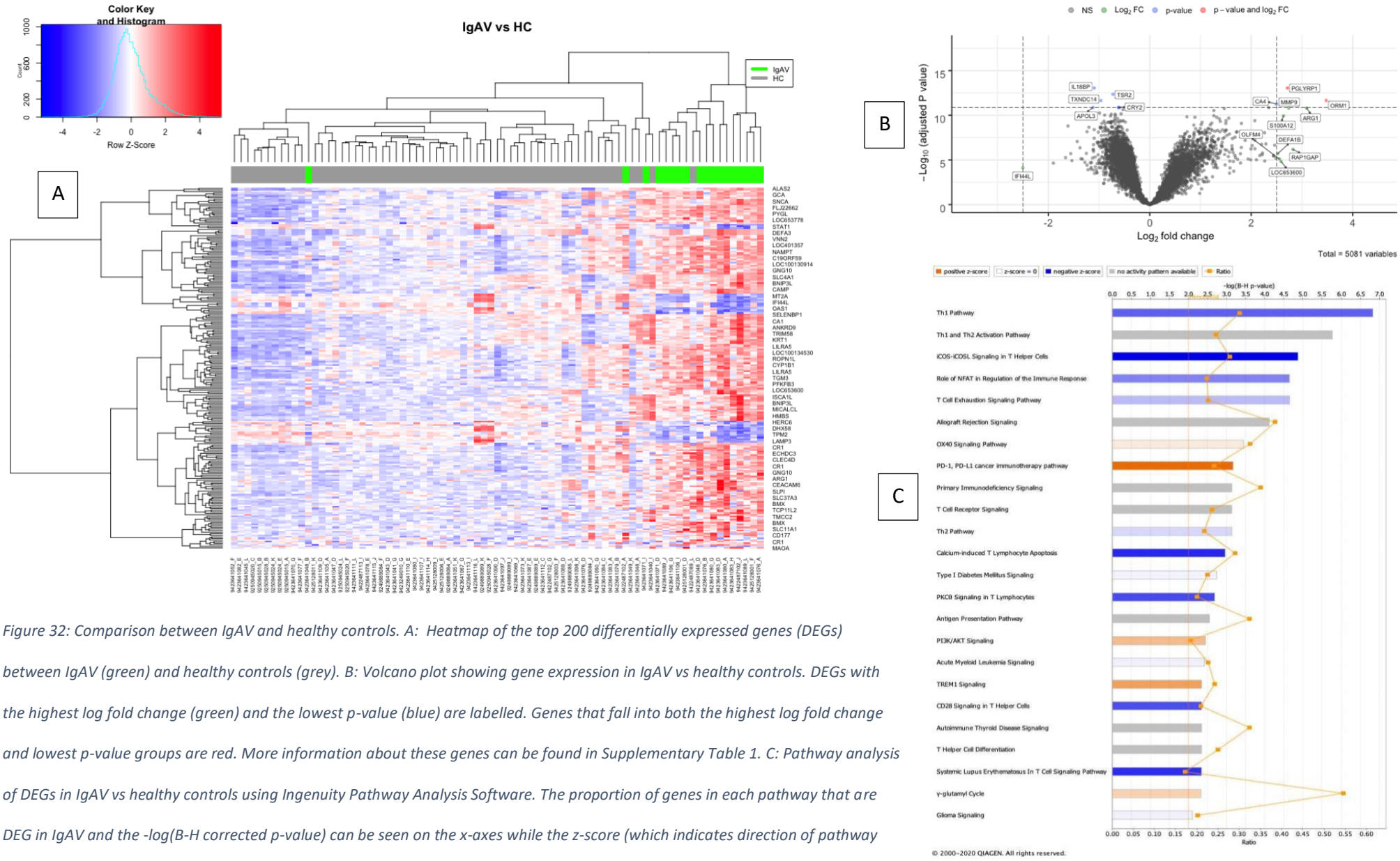



*Figure 32: Comparison between IgAV and healthy controls. A: Heatmap of the top 200 differentially expressed genes (DEGs) between IgAV (green) and healthy controls (grey). B: Volcano plot showing gene expression in IgAV vs healthy controls. DEGs with the highest log fold change (green) and the lowest p-value (blue) are labelled. Genes that fall into both the highest log fold change and lowest p-value groups are red. More information about these genes can be found in Supplementary Table 1. C: Pathway analysis of DEGs in IgAV vs healthy controls using Ingenuity Pathway Analysis Software. The proportion of genes in each pathway that are DEG in IgAV and the -log(B-H corrected p-value) can be seen on the x-axes while the z-score (which indicates direction of pathway regulation) is represented by the colour of each bar (red=high; blue=low).*

| Symbol | Name | Log Fold Change | B-H Adjusted p-value | Function |
|---|---|---|---|---|
| ***ORM1*** | Orosomucoid 1 | 3.47 | 2.2 $x10^{-12}$ | Encodes a key acute phase protein with a role in thrombin generation[159], which has been shown to be raised in IgAV, possibly in response to vascular inflammation[160,161]. |
| ***ARG1*** | Arginase 1 | 3.09 | 1.5 $x10^{-11}$ | Increased expression has been noted in the acute phase of KD[162]. Also associated with atopy to aeroallergens[163]. |
| ***RAP1GAP*** | RAP1 GTPase Activating Protein | 2.82 | 6.7 $x10^{-07}$ | Encodes a protein that downregulates RAP1 which is associated with IgA nephropathy[164] and increased expression has been reported in patients with non-diabetic glomerulosclerosis[165]. |
| ***MMP9*** | Matrix Metallopeptidase 9 | 2.74 | 1.6 $x10^{-11}$ | Increased expression has been linked with a number of vasculitides, with a suggestion that it may enable infiltration of immune cells into the vessel wall[166–168]. Increased MMP9 levels and activity are seen following allergen exposure[169]. |
| ***PGLYRP1*** | Peptidoglycan Recognition Protein 1 | 2.71 | 9.0 $x10^{-14}$ | Shown to be significantly differentially expressed in KD[170]. Increased expression has also been shown to enhance allergic response to environmental allergens[171]. |
| ***S100A12*** | S100 Calcium Binding Protein 12 | 2.64 | 1.2 $x10^{-10}$ | Overexpressed in various inflammatory conditions and increased concentration has been shown to correlate with disease activity of Kawasaki disease as well as inflammatory bowel disease, juvenile idiopathic arthritis, glomerulonephritis, psoriasis and asthma[172]. |
| ***OLFM4*** | Olfactomedin 4 | 2.55 | 7.8 $x10^{-06}$ | Animal models suggest expression of OLFM4 may reduce the immune response to pathogens such as *S.aureus*[173] and *H.pylori* and result in an altered microbiome[174]. |
| ***DEFA1B*** | Defensin Alpha 1B | 2.52 | 2.0 $x10^{-06}$ | Encodes a protein that is part of the innate immune response and expression is increased in response to certain viral infection[175,176]. Plays a role in mediating susceptibility of IgA nephropathy[177] and has been shown to play a role in allergy[178]. |
| ***CA4*** | Carbonic Anhydrase 4 | 2.49 | 5.4 $x10^{-12}$ | Expressed by eosinophils in response to IL-5 or allergen exposure[179]. |
| ***CRY2*** | Cryptochrome Circadian Regulator 2 | -0.62 | 1.4 $x10^{-11}$ | A cryptochrome that has been shown to be underexpressed in some inflammatory conditions[180,181], with a suggestion that it suppresses inflammatory pathways. Underexpression has also been associated with allergy and atopy[182]. |
| ***TSR2*** | TSR2 Ribosome Maturation Factor | -0.73 | 4.4 $x10^{-13}$ | Inhibits the pro-inflammatory NF-kB signalling pathway[183]. |
| ***TXNDC14 / TMX2*** | Thioredoxin Related Transmembrane Protein 2 | -0.96 | 2.2 $x10^{-12}$ | May play a role in reducing oxidative tissue damage associated with inflammatory conditions[184]. |
| ***IL18BP*** | Interleukin 18 Binding Protein | -1.09 | 9.0 $x10^{-14}$ | An antagonist of IL-18. Increased IL18BP expression reduces the T-helper type 1 immune response[185]. |
| ***APOL3*** | Apolipoprotein L3 | -1.11 | 1.4 $x10^{-11}$ | Inhibition of APOL3 has been associated with kidney disease[186,187] and underexpression in IgAV may provide an explanation for the renal sequelae of the disease. |
| ***IFI44L*** | Interferon Induced Protein 44 Like | -2.50 | 7.7 $x10^{-05}$ | Associated with viral infection[55]. It has been reported that downregulation of IFI44L may allow some viruses to evade the cellular innate immune response and promote their replication[188]. |

*Table 9: Top DEGs (by log fold change and B-H adjusted p-value) between IgAV and healthy controls.*

#### *5.3.1.1 IgAV Genes Associated with Inflammation and Infection*

There was increased expression of pro-inflammatory genes, and reduced expression of anti-inflammatory genes. *ARG1* (LFC=3.09, p=1.5x$10^{-11}$), *MMP9* (LFC=2.74, p=1.6 x$10^{-11}$), *PGLYRP1* (LFC=2.71, p9 x$10^{-14}$), *S100A12* (LFC=2.64, p=1.2 x$10^{-10}$) have been shown to be upregulated in KD and other inflammatory conditions, and overexpression of *MMP9* has been linked with infiltration of immune cells into vessel walls in some vasculitides[166–168]. *CRY2* (LFC=-0.62,

p=1.4 x$10^{-11}$) and *TSR2* (LFC=-0.73, p=4.4 x$10^{-13}$) are associated with suppression of inflammatory pathways[180,181,183] and *TXNDC14*/*TMX2* (LFC=-0.9, p=2.2 x$10^{-12}$) is thought to play a role in reducing oxidative tissue damage associated with inflammatory conditions[184]. In addition, *IL18BP* (LFC=-1.09, p=7.7 x$10^{-5}$) is an inhibitor of IL-18, which is part of the IL-1 pathway, and reduces the T-helper type 1 immune response, which is responsible for perpetuating autoimmune responses[185].

Genes linked to protection from infection showed both increased and reduced expression. *DEFA1B* (LFC=2.52, p=2 x$10^{-6}$) encodes Defensin alpha 1 protein, a neutrophil granule component that plays a role in the phagocyte-mediated host response[175,176]. *IFI44L* (LFC=-2.50, p=7.7 x$10^{-5}$) has antiviral properties, with increased expression restricting RSV infection[189], and downregulation associated with viral evasion of the cellular innate immune response, leading to increased viral replication[188]. In animal models, *OLFM4* (LFC=2.55, p=7.8 x$10^{-6}$) is thought to reduce the immune response to pathogens such as *S.aureus* and *H.pylori* and result in an altered microbiome[173,174].

### *5.3.1.2 IgAV Genes Associated with Allergen Exposure*

Among the top DEGs in IgAV vs HC, *ARG1* (LFC=3.09, p=1.5 x$10^{-11}$), *MMP9* (LFC=2.74, p=1.6 x$10^{-11}$), *PGLYRP1* (LFC=2.71, p9 x$10^{-14}$), *DEFA1B* (LFC=2.52, p=2 x$10^{-6}$), *CA4* (LFC=2.49, p=5.4 x$10^{-12}$) and *CRY2* (LFC=-0.62, p=1.4 x$10^{-11}$) have all been associated with allergen exposure[163,169,171,178,182], with *ARG1* and *PGLYRP1* specifically associated with exposure to environmental aeroallergens.

#### *5.3.1.3 Genes Associated with Renal Disease*

Among the 15 top DEGs in IgAV, 3 that have been previously linked to nephropathy (Table 9). There was increased expression of *DEFA1B* (LFC=2.52, p=2 x$10^{-6}$) and *RAP1GAP* (LFC=2.82, p=6 x$10^{-7}$), which have been associated with IgA nephropathy[164,177], and reduced expression of *APOL3* (LFC=-1.11, p=1.4 x$10^{-11}$), null variants of which are associated with nephropathy[187]. In addition to the above, *ORM1* (LFC=3.47, p=22 x$10^{-12}$), which showed the greatest log-fold change in IgAV compared to HC, encodes an acute phase protein with a role in thrombin generation[159], which has been shown to be raised in patients with IgAV[160,161].

#### *5.3.1.4 Pathway analysis in IgAV*

In order to explore the biological role of the genes that had perturbed expression, Ingenuity Pathway Analysis (QIAGEN Inc.)[157] software was used to identify significantly enriched biological pathways . For the comparison between IgAV and healthy controls, 24 pathways were significantly enriched ($p < 0.01$) (Figure 32C). Of these, seven pathways were associated with T cells and T helper cells and had a negative z-score, implying reduced activity. In addition the PD-1 PD-L1 pathway cancer immunotherapy pathway, which can cause inhibition of the T cell response[190], was found to have a positive z-score (p=6.7 x$10^{-4}$). The *TREM1* signalling pathway had a positive z-score; the *TREM1* receptor signalling can amplify an inflammatory response[191].

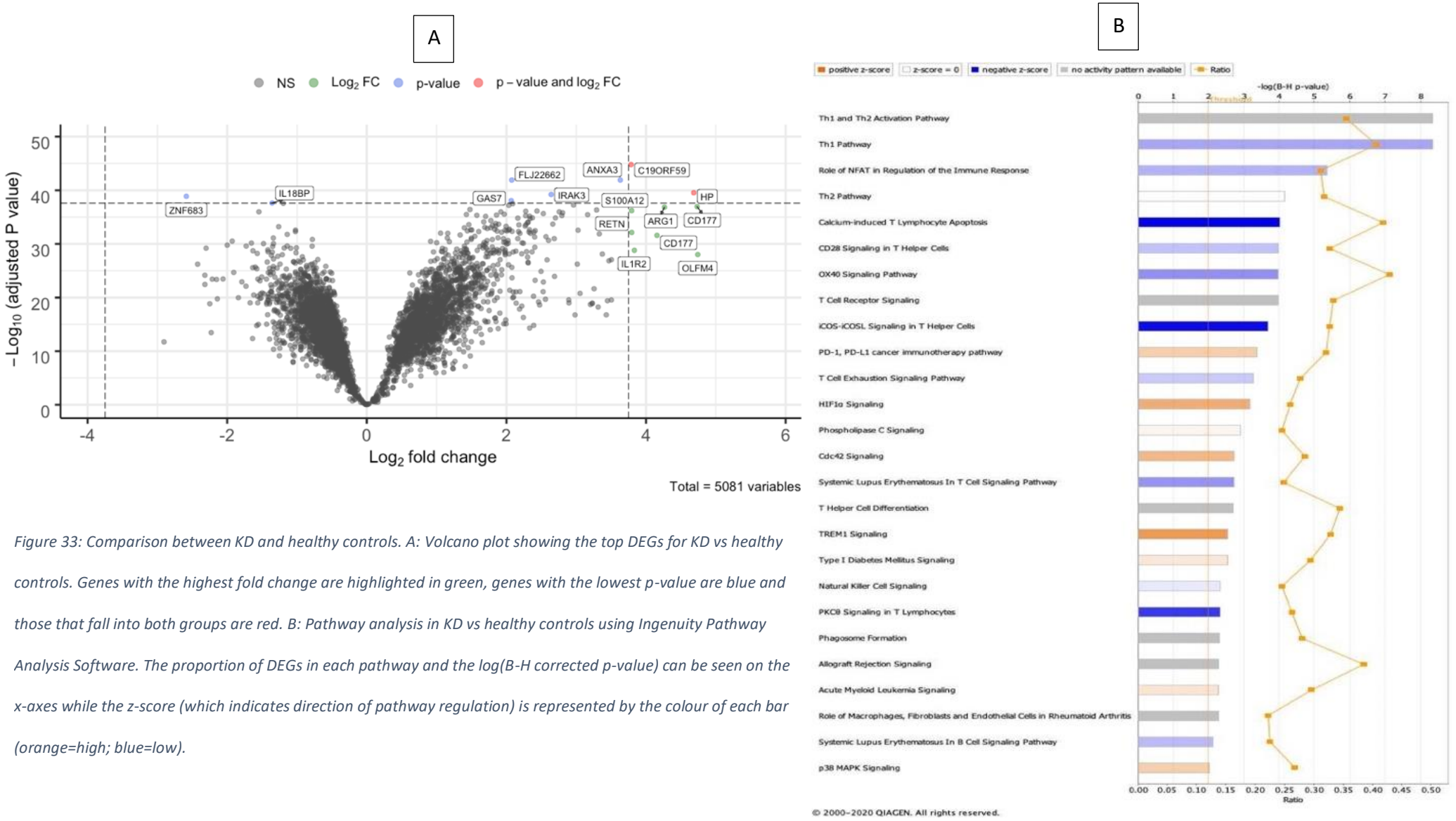



*Figure 33: Comparison between KD and healthy controls. A: Volcano plot showing the top DEGs for KD vs healthy controls. Genes with the highest fold change are highlighted in green, genes with the lowest p-value are blue and those that fall into both groups are red. B: Pathway analysis in KD vs healthy controls using Ingenuity Pathway Analysis Software. The proportion of DEGs in each pathway and the log(B-H corrected p-value) can be seen on the x-axes while the z-score (which indicates direction of pathway regulation) is represented by the colour of each bar (orange=high; blue=low).*

### 5.3.2 Transcriptome of KD patients

Of the 5,027 DEGs between KD and healthy controls, there were 2,060 (41%) overexpressed genes in KD and 2,967 (59%) underexpressed. The overall pattern of expression was similar to previously published data[142]. The top DEGs are involved in the innate immune response (*C19ORF59, ZNF683, IL18BP, IRAK3, S100A12, RETN, OLFM1),* inflammation (*IRAK3, S100A12*), toll-like receptor signalling (*S100A12*), prostaglandin regulation (*ANXA3*), adaptive T cell response (*ZNF683*), antiviral activity (*ZNF682, IL18BP*), antibacterial activity (*S100A12, CD177*) and cell surface interactions at the vascular wall (*CD177*). These genes are highlighted in the volcano plot in Figure 33A. Pathway analysis found significant negative z-scores in 9 pathways related to T cell and T helper cell function, and 1 pathway related to B cell function. There were positive z-scores in pathways related to hypoxic stress, the innate immune response and the inflammatory response (Figure 33B).

### 5.3.3 Comparative Analysis of IgAV and KD

Of the 5,084 genes that were significantly differentially expressed in either KD vs HC or IgAV vs HC, 4,930 (97%) genes demonstrated concordant expression in KD and IgAV as compared to healthy controls, with a greater fold change in the KD vs HC comparison for 3901/4930 (79%). 154 (3%) genes showed a discordant direction of regulation, of which only 7 were significantly differentially expressed in both IgAV vs HC and KD vs HC (Figure 34 and Table 10). All 7 genes showed overexpression in KD and underexpression in IgAV compared to healthy controls.

| Symbol | Name | Function | Expression |
|---|---|---|---|
| ***KLHL6*** | Kelch-Like 6 | Involved in B-lymphocyte receptor signalling and germinal-centre B-cell maturation. Upregulation is associated with immunoglobulin gene hypermutation.[192] | Spleen, lung, liver and kidney with weak expression in the heart[193] |
| ***DCUN1D3*** | Defective in Cullin Neddylation 1 Domain Containing 3 | Promotes cell differentiation and prevents cell growth, maturation and transformation.[194] | Adipose tissue and in the coronary and tibial arteries.[154] |
| ***TMEM110*** | Transmembrane Protein 110 | Essential role in calcium signalling at the endoplasmic reticulum - plasma membrane junction.[195] | Coronary, tibial arteries, the aorta and the liver, with lowest expression in the heart.[154] |
| ***C1QB*** | Complement C1q B Chain | Stress response, T-cell mediated immune[196] response and antibody-mediated complement activation[197]. | Spleen, lung, adipose tissue, aorta.[154] |
| ***C2ORF67*** | Chromosome 2 Open Reading Frame 67 | Transcription regulator. | Most abundantly expressed in the tibial artery and nerve. |
| ***PLAC8*** | Placenta-specific Gene 8 | Associated with the innate immune response. Diseases associated with *PLAC8* include Retinitis Pigmentosa[198]. | Blood leucocytes, spleen, lymph nodes and bone marrow. |
| ***PLIN2*** | Perilipin 2 | Involved in development and maintenance of adipose tissue. Expression is important in diseases such as Leprosy, where pathogens survive and replicate in lipid droplets within macrophages. | Widely expressed, including adipose tissue, endothelium, epithelium, liver, skeletal muscle.[154] |

*Table 10: Documented functions and tissue expression for DEGs with discordant, direction of expression between IgAV and KD. All these genes are significantly overexpressed in KD compared to controls and significantly underexpressed in IgAV compared to controls.*

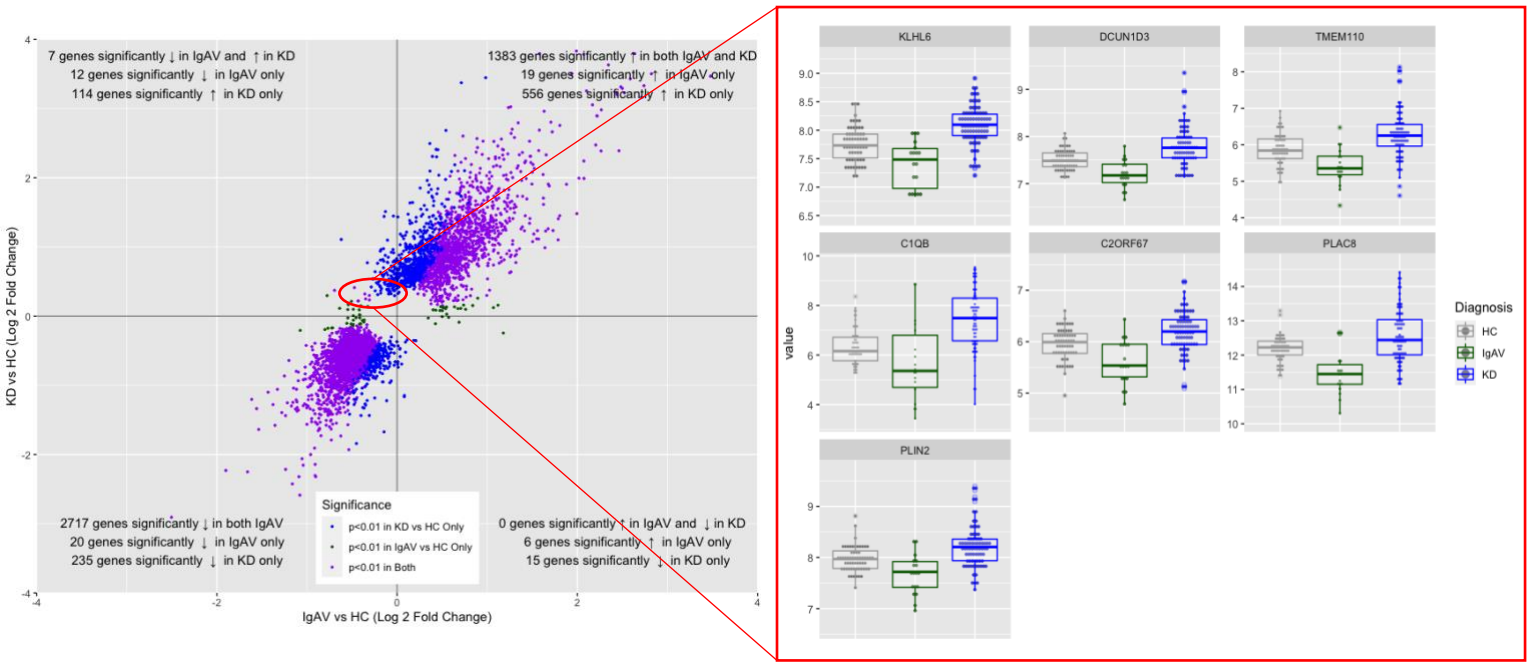



*Figure 34: Concordant/Discordant Genes and boxplots highlighting difference in expression levels of DEGs. The axes of the concordant/discordant plot represent the $log_2$ fold change and the colours represent genes that are significant in KD vs HC only (blue), IgAV vs HC only (green) and both (purple). 4,930 (97%) genes showed concordant expression and 3,901 (79%) of these had a greater fold change in KD vs HC compared to IgAV vs HC. 154 (3%) genes showed opposite direction of change in expression, of which 129 (83%) were significantly differentially expressed in KD only, 18 (13%) were significantly differentially expressed in IgAV only and 7 (4%) were significantly differentially expressed in both. Boxplots show the gene expression profiles of these 7 genes, with expression in healthy controls (HC) shown in grey, IgAV shown in green and KD shown in blue. Median expression in each group is shown, with box margins representing interquartile range.*

The IPA platform returned/calculated z-scores for 23 of 32 pathways that were significantly enriched in either IgAV or KD in comparison to healthy (Figure 35A). Of those, 17 (74%) pathways had concordant direction of z-score. Of note T cell and T helper cell, B cell and NK cell pathways had negative z-scores in both IgAV and KD, and there were positive z-scores for pathways that inhibit T cell activity. Other pathways with positive z-scores in both conditions included the p38 MAPK signalling pathway, which is targeted by the *P13K/AKT* signalling pathway and is important in the recruitment of leucocytes to sites of inflammation, and the Cdc42 signalling pathway, which is associated with endothelial senescence – a feature of inflammatory atherosclerosis.

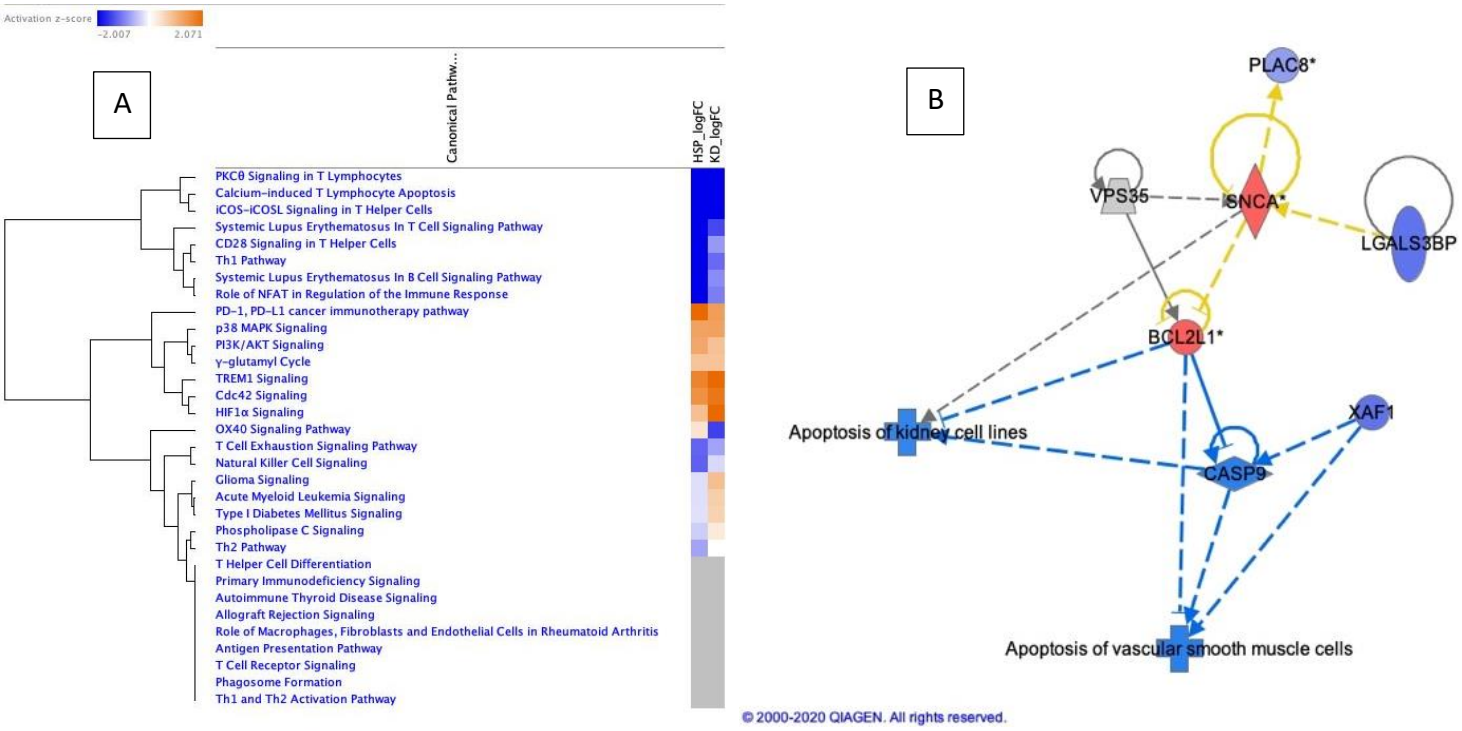



*Figure 35: Pathway comparison of KD and IgAV. A: Comparison of significantly differentially expressed pathways in IgAV and KD, compared to healthy controls and organised using hierarchical clustering. Pathways with a positive z-score are represented in orange and pathways with a negative z-score are blue. Pathways for which a z-score could not be calculated are grey. B: Pathway analysis of the 64 DEGs which were either significantly differentially expressed in IgAV vs HC but not KD vs HC or were significantly differentially expressed in both IgAV and KD vs HC with opposite direction of change. Upregulated genes and pathways in IgAV are highlighted in red and downregulated gene and pathways in IgAV in blue. Yellow links between genes indicate that the state of the downstream molecule is inconsistent with the state of the upstream regulator. In IgAV pathways of apoptosis of kidney cells and vascular smooth muscle cells are inhibited. The gene PLAC8 is one of the 7 genes that show significant orthogonal expression in IgAV and KD (Fig 33).*

### 5.3.4 Differences between IgAV and KD

57 DEGs in the IgAV vs HC comparison did not have significant differential expression in KD vs. HC. Seven genes were significantly differentially expressed in both IgAV and KD vs HC but had opposite direction of expression. These genes all had lower expression in IgAV in comparison to KD. Heatmap analysis of the combined 64 genes shows that most of these genes were overexpressed in KD and underexpressed in IgAV, however Euclidian clustering based on similarities in the transcriptome did not clearly separate patients with KD and IgAV. Network analysis highlighted one network that consists of nine genes, with underexpression

of *XAF1* and *CASP9* and overexpression of *BCL2L1* in patients with IgAV. This network is associated with inhibition of apoptosis of both vascular smooth muscle ($p=2 \times 10^{-6}$) and kidney cell lines ($p=5 \times 10^{-5}$) (Figure 35B).

## 5.4 Discussion

I performed transcriptomic analysis on patients with IgAV and KD to shed light on the differences in the vasculitic processes that lead to the distinct clinical syndromes and sequelae attributable to these two conditions. To our knowledge this is the first analysis of the whole blood gene expression profiles of patients with IgAV.

Limitations included the small number of patients in each of the comparator group, although this is not uncommon in transcriptomic studies. Both IgAV and KD are heterogenous diseases that are diagnosed based on a constellation of clinical signs and symptoms, and it is possible that some patients may have been misdiagnosed. In addition, the study was not powered to compare the transcriptomic patterns associated with severe sequelae, including renal impairment or aneurysm formation.

In IgAV, there was an expected overexpression of genes and pathways involved in the immune and inflammatory response. Among the top DEGs in IgAV, there were 6 genes associated with the allergen exposure, including 2 associated with exposure to environmental aeroallergens. As IgAV is an Ig-A mediated disease, an allergic aetiology is feasible and would be consistent with studies demonstrating an increased incidence of IgAV in children with atopic disease[199] and case reports of IgAV following exposure to allergens[200,201]. The top DEGs also contained 3 genes associated with nephropathy, 2 of

which are associated with Ig-A mediated nephropathy, potentially providing a lead point for further exploration of the cause of renal sequelae in IgAV.

Amongst the top DEGs in KD, when corrected for the abundance of neutrophils and lymphocytes, there was overexpression of genes associated with the innate immune response, the systemic inflammatory response, and the antibacterial response and underexpression of genes associated with the antiviral response and T-cell activation. These findings are consistent with the published literature on KD[125,133,202,203]. *CD177* showed increased expression as has previously been found in patients with KD, including those with atypical presentations and IVIG-resistance[204–207].

There was a remarkable similarity in gene expression profiles of IgAV and KD. As demonstrated in the Disco plot (Figure 34), 97% of genes showed concordant gene expression profiles for IgAV and KD, most of which (79%) had a greater change in expression in KD than IgAV. Similarly, pathway analysis found that most differentially regulated pathways in IgAV were common to KD, albeit with a larger effect size in KD, with common downregulation of T cell and T-helper cell signalling and function pathways. The close similarity in the gene expression profiles of KD and IgAV, with similar patterns of inflammatory change, is at odds with their distinctive clinical presentations, end-organ disease and sequelae. In contrast, clinically similar infections caused by different pathogens tend to have a diverse transcriptomic response, as seen for instance in the transcriptomes of children with respiratory infection caused by RSV or influenza[52]. Whilst the aetiology of both conditions remains obscure, the transcriptomic similarities between KD and IgAV may relate to similarities in aetiology or pathophysiology and warrants further exploration.

The small number of discordant genes included several expressed in key target tissues of IgAV (kidney) and KD (coronary arteries). Network analysis indicated that the overexpression of *BCL2L1* and under-expression of *CASP9* and *XAF1* in patients with IgAV would inhibit pathways associated with apoptosis of vascular smooth muscle cells, a pro-aneurysmal process that has been shown to play an important role in the formation of arterial aneurysms[208,209]. These genes were not significantly differentially expressed in KD. All 7 DEGs with discordant expression between IgAV and KD were upregulated in KD and downregulated in IgAV. Three of these genes (*DCUN1D3*, *TMEM110* and *C2ORF67*) were highly expressed in coronary and tibial arteries, and another 3 (*KLHL6*, *C1QB*, *PLAC8*) are involved in the immune response.

This study represented an opportunity to study distinct diseases with unclear aetiology and develop the skills necessary to analyse gene expression data. These skills would be necessary for my overall project which involved combined analysis of clinical, pathogen and gene expression data.

# 6. Gene Expression Analysis of Febrile Children

## 6.1 Introduction

### 6.1.1 Host Response to Infection with Bacteria and Viruses

Viruses require host cell factors for nearly every stage in their life cycle, and therefore many host defences against viral infections are aimed at inhibiting viral entry into cells or host factors required for viral replication, usually via interferon-related pathways[20,210–212]. As bacteria are more autonomous than viruses and are less dependent on host factors, the host response to infection with bacteria is more heterogenous, involving many different cell types and immune pathways depending on invading pathogen and site of infection[40,210,213]. This diversity in host responses to bacterial infection may explain why the 12-gene signature includes more genes aimed to identify bacterial infection than viral infection.

### 6.1.2 The 12-Gene Signature for Distinguishing Bacterial from Viral Disease

By examining the performance of two previously reported gene expression signatures for infection, consisting of 35 and 2 genes respectively[55,214], and a new in-house analysis of data from the EUCLIDS study[33], a new 12-gene signature was discovered by the Department of Paediatric Infectious Diseases, Imperial College London (Personal Communication with Dr M Kaforou, unpublished). This gene expression signature was developed to distinguish bacterial and viral infection and consists of 7 genes used to identify patients with bacterial disease and 5 genes to identify those with viral disease. The known function and expression patterns of these 12 genes are described below.

#### *6.1.2.1 Bacterial Genes*

##### ADGRE1

Adhesion G protein-coupled receptor E1 (ADGRE1), is highly expressed in alveolar and intestinal macrophages and encodes the F4/80 antigen. It is a marker for the differentiation of CD16++ monocytes and is thought to play a role in the host defence to infection[215].

##### EBI3

Epstein-Barr virus induced gene 3 forms part of IL-27, a member of the IL-6 family of cytokines. This interleukin has both pro and anti-inflammatory actions and can regulate T-helper cell development, suppress T-cell proliferation and stimulate cytotoxic T-cell activity[216,217].

Increased expression of EBI3 has been associated with bacterial sepsis[218] and bacterial infections, including *C.difficile*, *S.aureus* and *M.tuberculosis*[219].

##### FAM89A

The function of FAM89A remains poorly understood, and there are no detailed reports of its biological function. It has, however been included in a 2-gene signature designed to distinguish bacterial from viral infections[56].

##### IL1R2

IL1R2 is thought to act as a decoy receptor for IL-1, which captures circulating IL-1 and reduces its bioavailability[220]. Increased expression of IL1R2 is thought to lead to negative regulation of inflammation due to the activation of anti-inflammatory pathways by glucocorticoid hormones or in response to anti-inflammatory agents[221]. It has been shown to have increased expression in critically ill patients, and those with gram negative bacterial infections compared to gram positive infections[221].

#### OLAH

Oleoyl-ACP hydrolase (OLAH) is involved in the medium-chain fatty acid biosynthesis process. Although there are no reports of functional links to infection or the immune response, it has been reported to have increased expression in patients with pneumonia and infections of the central nervous system[222].

#### PI3

Proteinase inhibitor 3 has antibacterial activity against both gram-positive and gram-negative bacteria at epithelial sites and plays a major role as an anti-inflammatory mediator at mucosal surfaces[223].

#### S100P

S100 calcium-binding protein P belongs to a superfamily of calcium-binding proteins and is thought to be involved in cytoplasmic calcium control[224]. Although it has be suggested that S100P plays an important role in many cancers, with increased expression in patients with breast, colon, pancreatic and lung cancer, its cellular function is poorly understood[225]. However, despite no clear functional links to infectious diseases, S100P has been included in a 3-gene transcriptomic signature that was able to distinguish bacterial and viral infections in pigs[226].

### *6.1.2.2 Viral Genes*

#### IFI44L

Interferon-induced protein 44-like (IFI44L) is a type I interferon-stimulated gene[227] that is induced by many viruses including influenza A, coronavirus 229E and RSV[228]. IFI44L appears to act as a feedback regulator of interferon responses[173], as well as reduce viral

replication[189]. IFI44L was used with FAM89A in a 2-gene transcriptomic signature aimed at differentiating patients with bacterial infection from those with viral infection[56].

### IFIT1

Interferon-induced protein with tetraticopeptide repeats 1 (IFIT1) is one of the most strongly and rapidly induced interferon-stimulated genes, with increased expression in response to viral infection. IFIT1 has been shown to inhibit viral RNA translation and therefore inhibit viral replication[229].

### IFI27

Interferon-alpha-inducible protein 27 is another interferon-stimulated gene. Although the function of IFI27 is not fully understood, it has been shown to have antiviral activity and inhibit viral replication[212,230].

### OAS1

2’ – 5’ – Oligoadenylate Synthetase 1 (OAS1) is an interferon-induced gene and critical component of the innate immune response to viruses that degrades viral and cellular RNAs, leading to impairment of viral replication[231]. Increased circulating OAS1 was shown to be protective against infection, hospitalisation and disease severity in patients with COVID-19[232].

### RSAD2

Radical S-adenosyl methionine domain-containing protein 2 (RSAD2) is an interferon-induced gene associated with inhibition of replication of a number of DNA and RNA viruses, including herpes viruses, measles, dengue virus, influenza, rabies virus, and HIV-1[233–235].

### 6.1.2 Using the 12-gene signature to investigate infectious diseases

The 12-gene signature was designed to be used in a combined statistical model to predict bacterial or viral disease. Gene expression data was available for 656 PERFORM patient samples (369 DB and 287 DV). The statistical model with an 80:20 training:test split using 25-fold cross validation, repeated 10 times. On an independent dataset of PERFORM samples, the resulting model was shown to accurately distinguish between DB and DV patients (Figure 36).

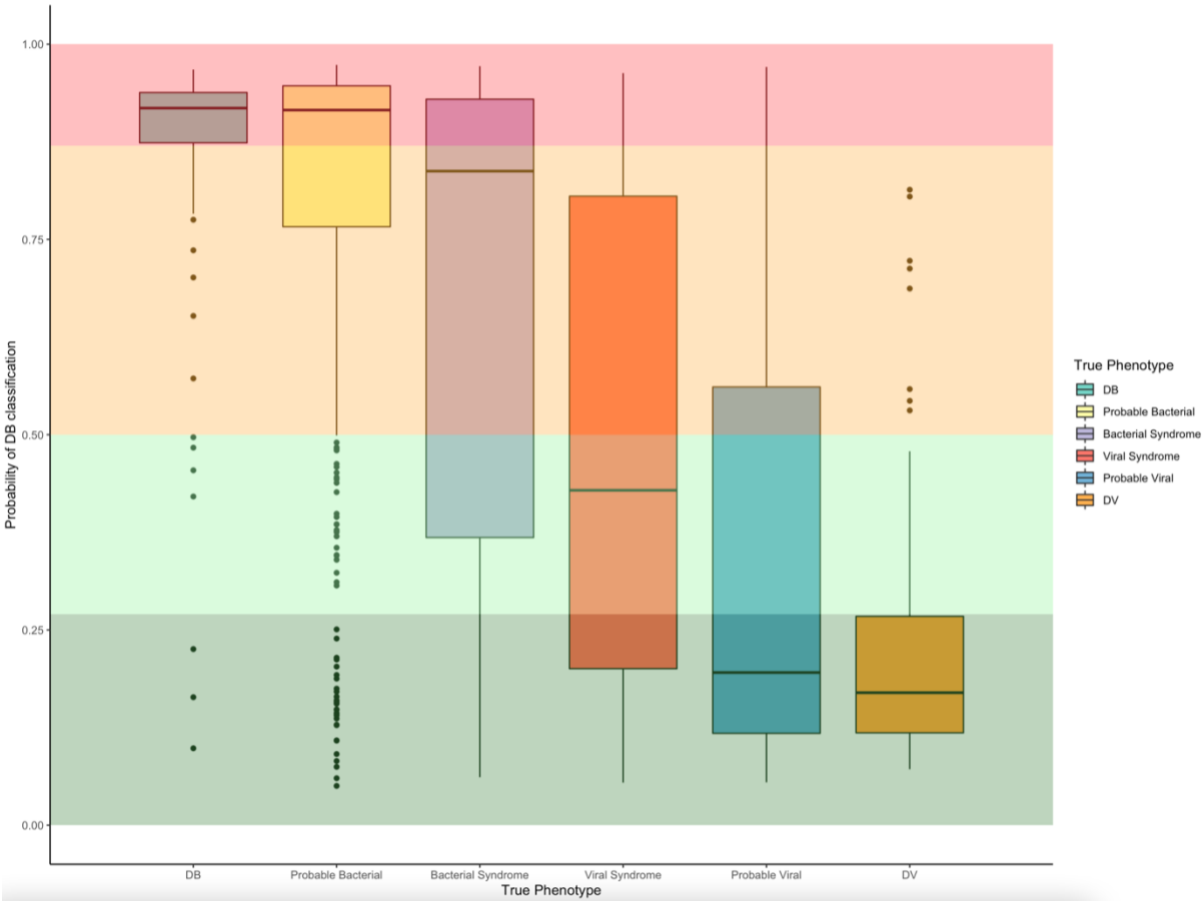



*Figure 36: Validation performance of the 12-gene signature and accompanying model. The statistical model transformed gene expression of the 12 genes into a single numerical output that represents the predicted probability of bacterial infection. This plot shows the median and IQR predicted probability of bacterial infection for patients in each phenotype. Figure created by R Womersley (unpublished).*

### 6.1.3 Aims

In this study, I investigate whether there are differences in the host response to different pathogens, pathogen co-detections or clinical syndromes, by observing for differences in expression of the 12-genes. I aimed to:

- Examine whether the gene expression signature, which has been validated in a larger group of PERFORM patients, can be applied to this smaller group of patients with both CMT and gene expression data.
- Examine the effect of clinical syndromes or pathogen detection on the output of this gene expression signature and the associated statistical model.
- Examine the effect of pathogen co-detection on the host response.
- Examine whether using pathogen load cut-off limits of detection for blood bacterial and viral assays lead to differences in gene expression.

## 6.2 Methods

RNA extracted from whole blood PAXgene samples from patients recruited to the PERFORM study were sent for analysis at bioMérieux, France and processed by their in-house team. Expression of the genes in the 12-gene signature, as well as 3 housekeeping genes were analysed using the bioFire® filmArray® system – a multiplex PCR system that integrates sample preparation, amplification, detection and analysis. The filmArray® system reported gene expression data for each gene in the 12-gene signature with in-device normalisation using 3 housekeeping genes (PPIB, HPRT1 and HP).

Gene expression was investigated in cases and controls, including comparing expression of each gene in controls with and without pathogen detections. This 12-gene signature was

designed to be used with a support vector machine model, developed by my colleague Rebecca Womersley (unpublished) that takes the fold change of each gene in relation to housekeeping genes as an input and outputs a single numerical result which represents the predicted probability of bacterial infection (PPBI). This model has been shown to distinguish DB and DV patients (Figure 36). I compared the PPBIs of DB and DV patients in our cohort to assess whether it was comparable to that of the validation of the signature model. I then explored the PPBI of patients with each clinical syndrome and the detection of each pathogen. To improve clarity of the results and only include results with sufficient accuracy, only pathogens detected in more than 20 febrile patients were included in this analysis.

As shown in chapter 2, certain pathogens were more likely to be detected in blood of cases vs controls and DB vs DV patients at higher pathogen loads. Pathogen load data was available for blood virus assays, apart from enterovirus and parechovirus. For enterovirus, parechovirus and blood bacterial assays, the inverse of the Ct value was used as a surrogate marker for pathogen load. I investigated whether there was a relationship between pathogen load and gene expression for any pathogen. I used the R package cutpointR[236] to identify the optimal pathogen load cut-offs for each pathogen to maximise the odds ratio for detection in cases vs controls. I then assessed whether only including detections with pathogen loads above these cut-offs had a greater effect on gene expression compared to any detection of that pathogen and compared to other febrile children.

To investigate the impact of co-detection on the host response, PPBI was compared in viral patients with and without bacterial co-detection and in bacterial patients with and without viral co-detections.

PPBIs in febrile children were not normally distributed, and I confirmed this using the Shapiro-Wilk test for normality ($p<0.0001$). Figure 40 demonstrates a bimodal distribution of PPBIs in febrile children. Significant differences in PPBI between groups were therefore assessed using the Mann-Whitney-U test using the R package stats[71] and significant similarities in PPBI between groups was assessed using the non-parametric Two One-Sample T-test for Equivalence[237], using the R package TOST[238].

### 6.2.1 Results

Of the 4611 febrile children and 1061 controls with CMT results and clinical data available, gene expression data was available for 2266 febrile children and 206 healthy controls (Figure 37). Compared to the larger group with CMT results available, in the patients with both CMT and gene expression data available, demographic, syndrome and phenotypes were generally similar, but there was a greater proportion with upper and lower respiratory tract infections and assigned DB, PV and DV phenotypes and a smaller proportion of controls (Table 11).

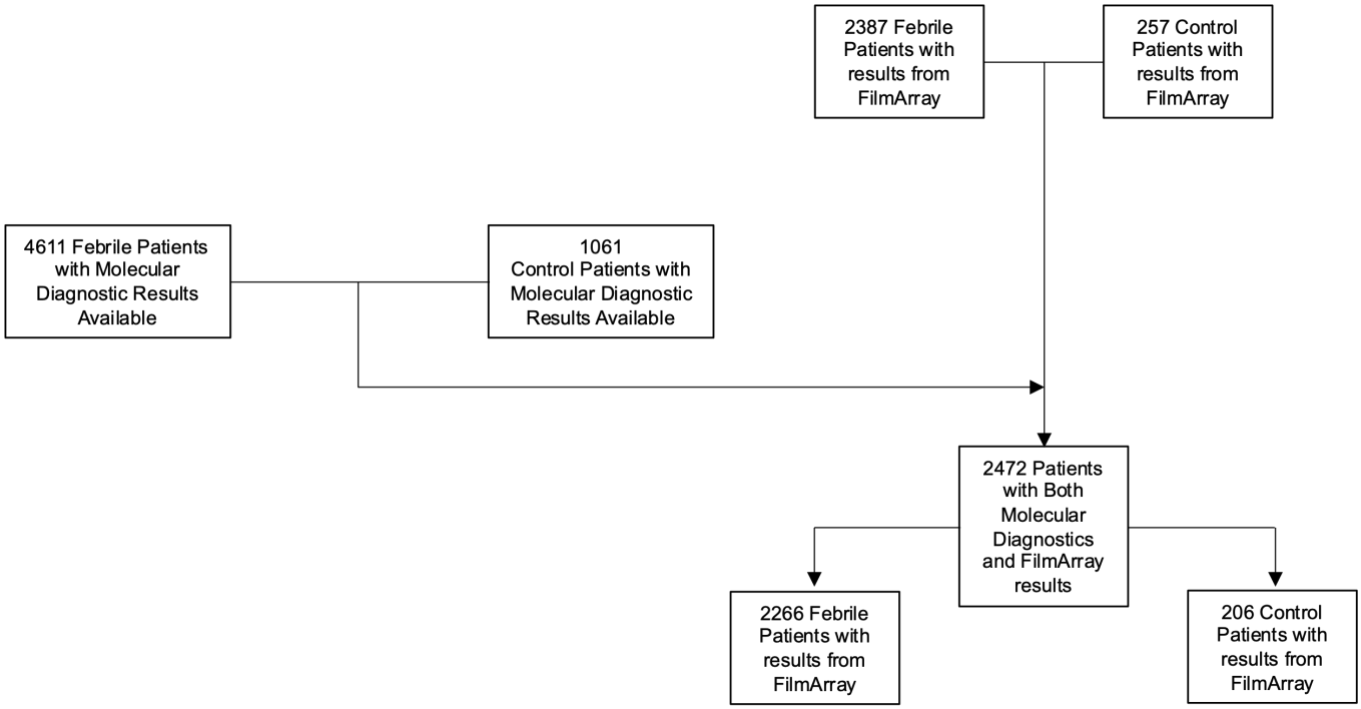



*Figure 37: Consort diagram showing the number of febrile children and controls with both CMT and gene expression data available.*

| | **Patients with CMT and Gene Expression Results (n=2472)** | **Patients with CMT Results (n=5672)** |
|---|---|---|
| Gender (M) | 1303 (52.7%) | 3159 (55.9%) |
| Age in months (median [IQR]) | 61.8 [34.5 – 123] | 64.5 [22.7 – 134] |
| **Syndrome Group** | | |
| Upper Respiratory Tract Infection | 516 (20.9%) | 1032 (18.2%) |
| Lower Respiratory Tract Infection | 438 (17.7%) | 903 (15.9%) |
| Gastrointestinal Infection | 221 (8.9%) | 425 (7.5%) |
| Central Nervous System Infection | 77 (3.1%) | 172 (3%) |
| Urinary Tract Infection | 154 (6.2%) | 287 (5.1%) |
| Musculoskeletal Infection | 47 (1.9%) | 125 (2.2%) |
| Skin & Soft Tissue Infection | 195 (7.9%) | 383 (6.8%) |
| **Clinical Phenotype** | | |
| Definite Bacterial | 324 (13.1%) | 643 (11.3%) |
| Probable Bacterial | 320 (12.9%) | 677 (11.9%) |

*Table 11: Table comparing demographic, syndrome and phenotype data between patients with both CMT and gene expression data and the larger group of patients with CMT data available.*

| | | |
|---|---|---|
| Bacterial Syndrome | 213 (8.6%) | 456 (8.0%) |
| Unknown Bacterial or Viral | 322 (13.0%) | 791 (13.9%) |
| Viral Syndrome | 136 (5.5%) | 299 (5.3%) |
| Probable Viral | 327 (13.2%) | 635 (11.2%) |
| Definite Viral | 280 (11.3%) | 491 (8.7%) |
| Other Phenotype | 224 (9.1%) | 619 (10.9%) |
| Control | 206 (8.3%) | 1061 (18.7%) |

### 6.2.2 Gene expression in controls

In chapter 2 I showed that pathogens were detected in controls as well as cases. To ascertain the clinical significance of these pathogen detections in controls, I assessed whether there was any measurable difference in the host response of controls in whom pathogens were detected by CMT (n = 169/206, 82%) and those in whom no pathogen was detected (n = 37/206, 18%), compared to housekeeping genes. On comparing these subgroups of healthy controls, there was no significant difference in expression of any of the 12 genes (Figure 38).

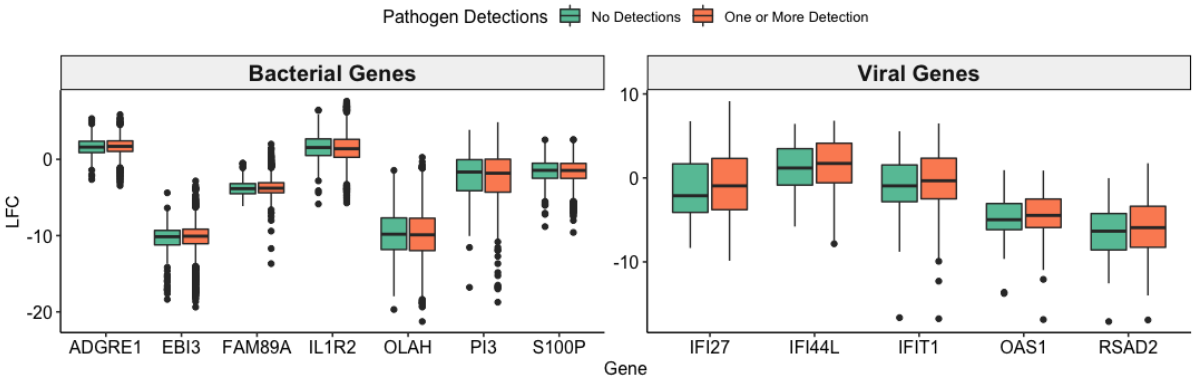



*Figure 38: Expression of the 7 bacterial genes and 5 viral genes in controls with no pathogens detected (green) and one or more pathogens detected (orange) by CMT. Expression is expressed as Log2 fold change (LFC) in relation to housekeeping genes.*

### 6.2.3 Expression in cases, DB and DV

To examine whether differences in gene expression between children with and without infections can be identified by the 12-gene signature, I compared expression of the 12 genes

in cases and controls (Figure 39).There was higher expression in cases vs controls of ADGRE1 (LFC: 0.25, p=0.004), FAM89A (LFC 0.48, p<0.0001), OLAH (LFC 1.48, p<0.0001), IFI44L (LFC: 1.81, p<0.0001), IFIT1 (LFC: 1.47, p<0.0001), IFI27 (LFC: 2.37, p<0.0001), OAS1 (LFC: 1.28, p<0.0001) and RSAD2 (LFC: 2.11, p<0.0001). Cases had significantly lower expression of PI3 compared to controls (LFC: -1.2, p<0.0001).

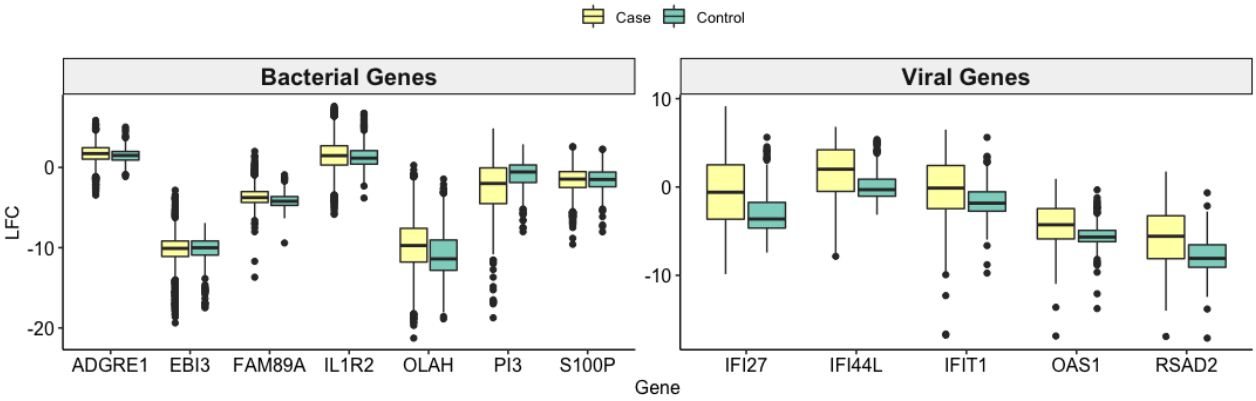



*Figure 39: Comparison of gene expression of the 7 bacterial genes and 5 viral genes between cases (yellow) and controls (green). In this analysis all cases are grouped together, irrespective of their cause of illness. Expression is expressed as Log2*

The SVM model developed by Rebecca Womersley was designed to use gene expression data from the 12-gene signature to compute a predicted probability of a febrile child having a bacterial infection (PPBI). In cases, the model outputs a bimodal distribution, and the majority of febrile children have a very high or a very low PPBI (Figure 40).

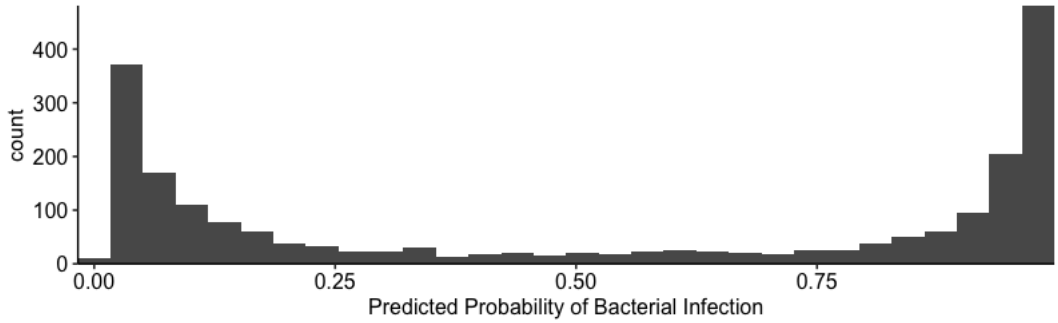



*Figure 40: Density plot showing the distribution of predicted probabilities of bacterial infection in febrile children.*

The model was built on a larger dataset of febrile children recruited to the PERFORM study. Using a binomial logistic regression, I assessed whether the correlation between PPBI and odds of being assigned to the DB phenotype still exists in this smaller subset of patients with both gene expression and CMT data available. I found that for every 0.1 increase in PPBI, the log odds of the patient having a DB phenotype increased by 0.21 ($p<0.0001$). Figure 41A shows that this model clearly separates patients with DB and DV ($p<0.0001$), and controls from DV ($p<0.0001$), but not controls from DB. Figure 38 shows that the detection of pathogens in controls does not result in altered expression of any of the 12 genes that form the input of this model and we can therefore infer that the similarity in predicted probability of bacterial infection in DB patients and controls is not due to the presence of pathogens in controls. This is clearly demonstrated in Figure 41B. Although this model does not separate DB patients from controls, in febrile patients it does appear to separate DB and DV patients well. This may be because the model is identifying the differences of viral and non-viral host responses. I have therefore limited further analysis of gene expression to febrile patients to

examine whether syndromes or pathogen detection can be related to a viral-like or bacterial-like (non-viral) host response.

### 6.2.4 Gene expression in different clinical syndromes

Figure 42 show the predicted probabilities of bacterial illness in each clinical syndrome, overlayed on the confidence intervals of PPBI in DB and DV patients to show which syndromes were associated with a bacterial-like and viral-like host response. To improve clarity and ensure accuracy of results, only clinical syndromes detected in more than 20 patients were included in this analysis.

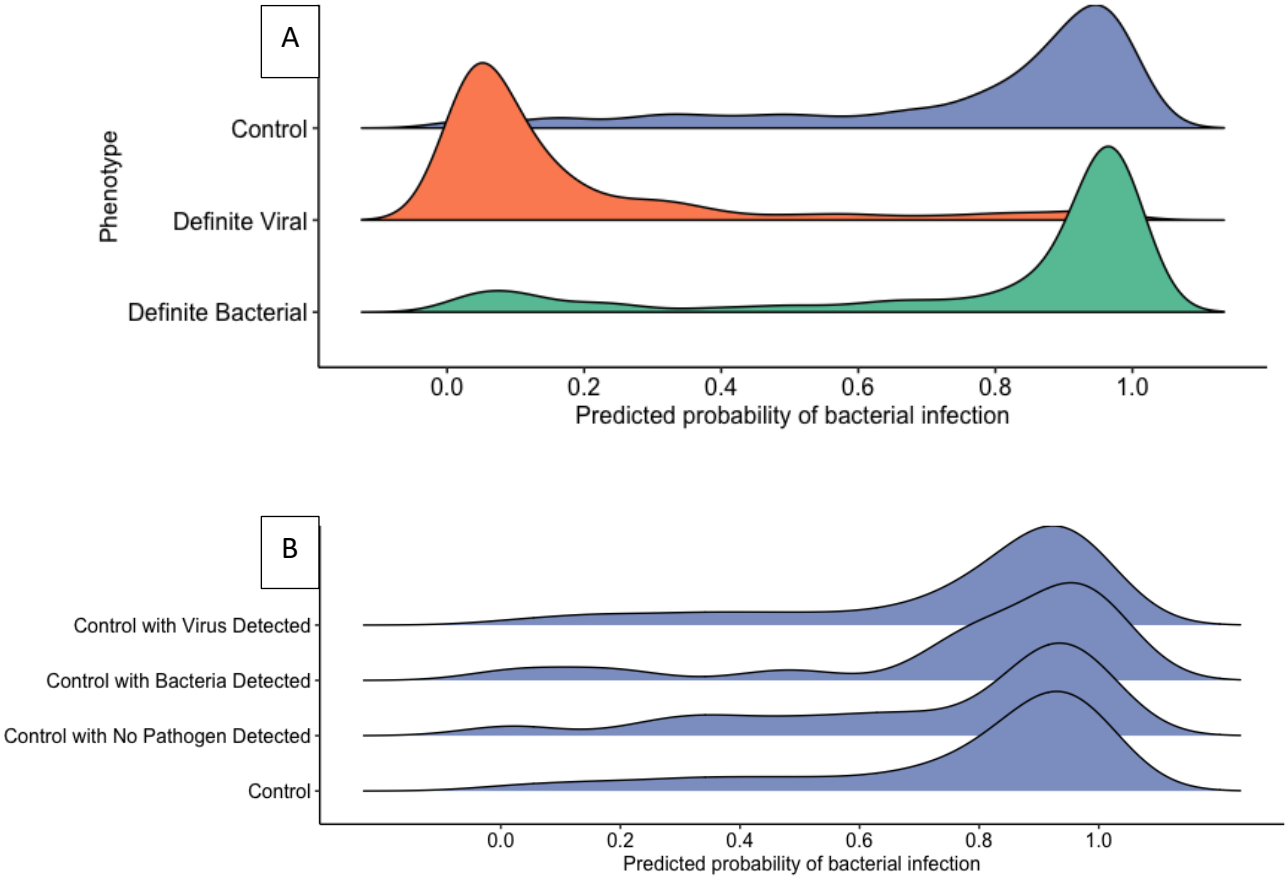



*Figure 41: Predicted probability of bacterial infection (PPBI) in febrile children assigned to definite bacterial (green) and definite viral (orange) phenotypes and controls (blue) calculated using a support vector machines model developed by Rebecca Womersley (unpublished). A: Density plot showing distribution PPBIs in DB and DV patients and controls. B: Distribution of PPBIs in all controls, controls with bacteria and viruses detected and controls with no pathogen detections.*

#### *6.2.4.1 Upper respiratory tract infections*

Patients with non-specific URTI were associated with PPBIs that were significantly lower than the DB group (p=0.001) and statistically equivalent to DV patients (p=0.03). PPBIs in patients with otitis media, pharyngitis and tonsilitis were significantly higher than those with a DV phenotype (p = 0.03, 0.0004, <0.0001 respectively) and lower than those with a DB phenotype (p = <0.0001 for each).

#### *6.2.4.2 Lower respiratory tract infections*

All lower respiratory tract syndromes were associated with PPBIs that fell between those of DV and DB patients apart from pleural effusions which were associated with significantly higher PPBIs than both DB (0.001) and DV (p<0.0001) groups. No lower respiratory tract syndromes were associated with PPBIs equivalent to DB or DV groups.

#### *6.2.4.3 Gastrointestinal Infections*

PPBIs in patients with appendicitis and mesenteric adenitis were similar to those of DB patients and significantly higher than DV patients (p<0.0001 for both). PPBI in patients with diarrhoeal illness was between that of DV and DB patients. No gastrointestinal syndromes were associated with PPBIs statistically equivalent to those of DV or DB patients.

#### *6.3.4.4 Central Nervous System Infections*

PPBIs in patients with meningitis and encephalitis were between those of DB and DV patients and not equivalent to either.

#### *6.3.4.5 Urinary Tract Infections*

PPBIs of patients with lower UTI and pyelonephritis was statistically equivalent to that of DB patients (p=0.007, 0.006 respectively) and significantly higher than DV patients (p<0.0001 for both).

#### *6.3.4.6 Skin and Soft Tissue Infections*

PPBIs in patients with clinical syndromes of abscess, lymphadenitis and wound infection were significantly higher than that of DV patients (p<0.0001 for all) and similar to DB patients. No skin or soft tissue syndromes were associated with PPBIs that were equivalent to that of DV or DB patients.

#### *6.3.4.7 Pathogen Related Syndromes*

Bacteraemia was associated with PPBIs equivalent to those of DB patients (p<0.0001). Syndromes associated with viruses such as flu-like illness, varicella and glandular fever had PPBIs that were significantly lower than those of DB patients (p = 0.0003, <0.0001, <0.0001 respectively). PPBIs in patients with flu-like illness were statistically equivalent to those of DV patients (p=0.03).

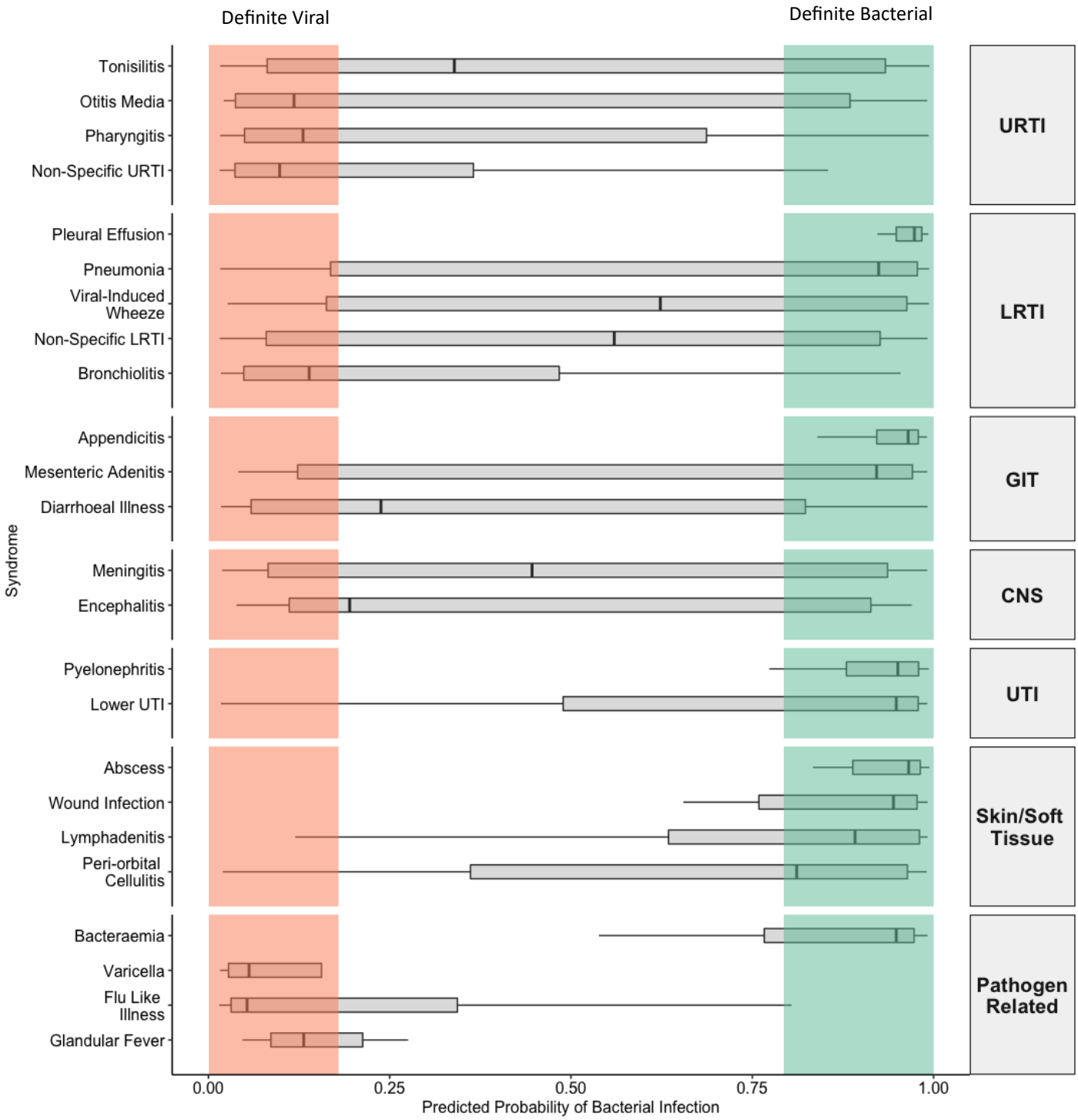



*Figure 42: Boxplots showing predicted probability of bacterial infection (PPBI) in patients with each clinical syndrome, organised by syndrome group. The black vertical line, grey box and grey horizontal lines represent the median, IQR and 1.5*IQR PPBIs for each syndrome. Outliers outside the horizontal line representing 1.5*IQR are omitted for clarity. The green and orange bars show the confidence interval of PPBI for DB and DV patients respectively when the model was applied to them. For clarity, only syndromes detected in more than 20 patients are displayed. It was possible for a patient to be assigned more than one syndrome and in these cases patients may be represented more than once in this plot.*

### 6.2.5 Expression (in response) to pathogen detection

#### *6.2.5.1 Expression in patients with viruses detected by CMT*

To assess the impact of the detection of viruses on the output of the 12-gene signature, I first examined PPBI in patients in whom only viruses were detected in the CMT data. As shown in Figure 43, in these patients PPBIs fell between that of DV and DB patients and was not statistically equivalent to either.

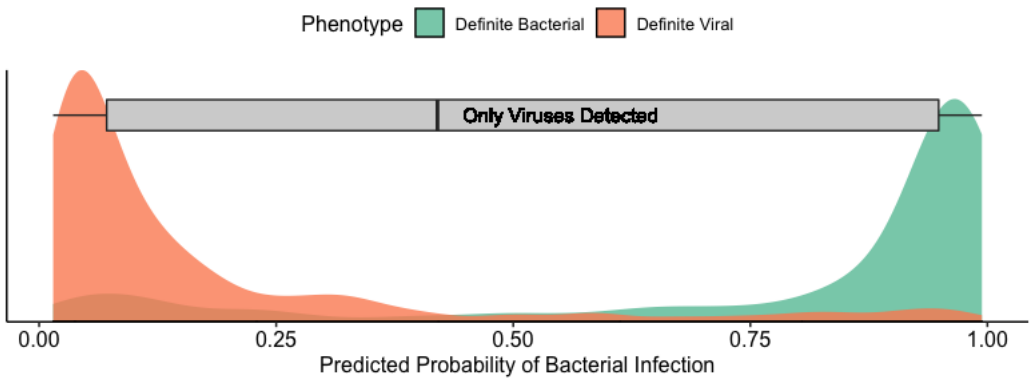



*Figure 43: Boxplot showing median and IQR predicted probability of bacterial infection (PPBI) in febrile children with detection of viruses and no bacterial co-detections. This boxplot has been superimposed on a density plot showing distribution of PPBI in DB (green) and DV (orange) patients.*

I performed a subgroup analysis by each virus tested for in the CMT data, limiting this analysis to viruses detected in at least 20 patients. In respiratory viruses (Figure 44), median PPBIs in 8 of the 10 respiratory viruses tended towards a viral-like response. When assessing whether PPBIs appeared to arise from a population with the same host response as DV patients, only PPBIs for patients with Influenza A and B were statistically equivalent to DV patients. No respiratory viruses were associated with PPBIs similar to DB patients.

In blood viruses (Figure 45), PPBIs of patients with enterovirus was statistically equivalent to DV patients. No other pathogens were associated with PPBIs similar to DB or DV patients.

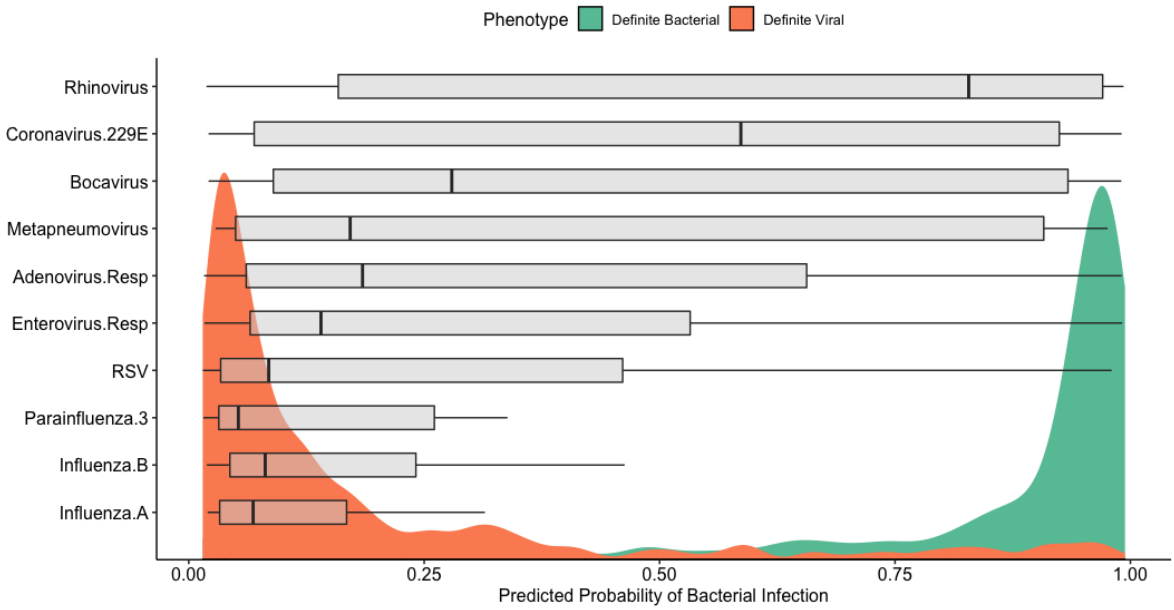



*Figure 44: Boxplots showing median and IQR predicted probabilities of bacterial infection for respiratory viruses. Only respiratory viruses detected in more than 20 patients with both CMT and gene expression results available are displayed. Boxplots are overlayed on a density plot showing PPBI in DV (orange) and DB (green) patients.*

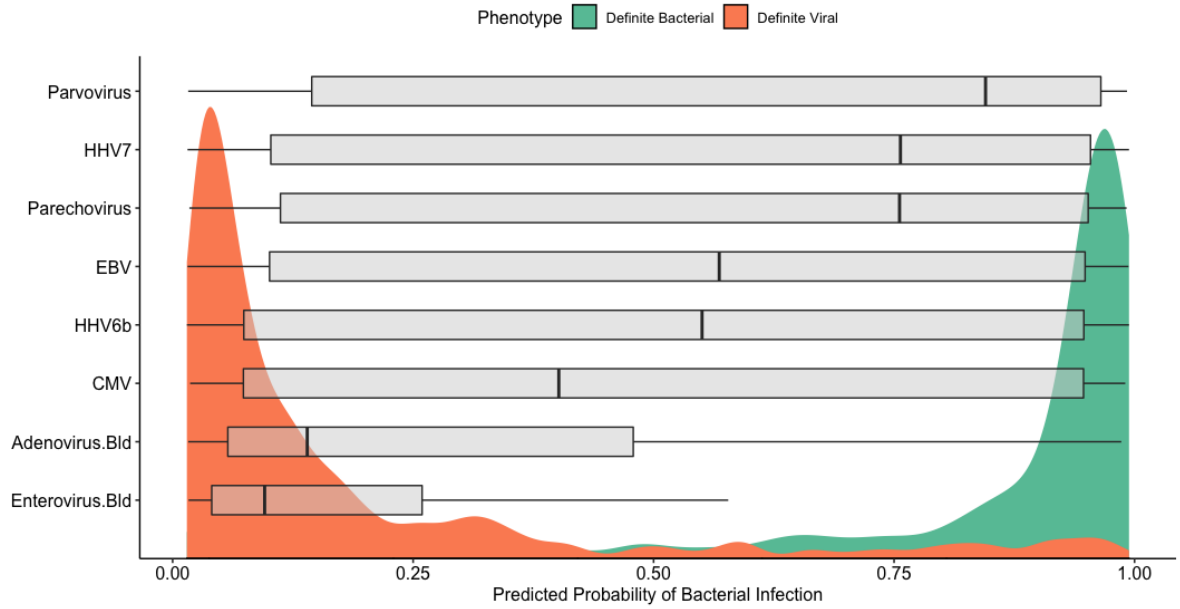



*Figure 45: Boxplots showing median and IQR predicted probability of bacterial infection (PPBI) when each blood virus is detected. Boxplots are superimposed on a density plot showing the distribution of PPBIs in DV (orange) and DB (green) patients. Only blood viruses that were detected more than 20 times in patients with both CMT and gene expression data are displayed.*

In chapter 3 I showed that some pathogens were associated with greater odds of detection in cases vs controls and DV vs DB at higher pathogen loads. With the addition of gene expression data, I investigated whether PPBI was related to pathogen load (Figure 46). The herpes viruses HHV6b, HHV7, CMV and EBV appeared to have a larger concentration of detections at lower viral loads without clear correlation to PPBI. Detections at higher viral loads appeared to have a clearer relationship with PPBI. To test this further, I calculated the optimal pathogen load cut-offs which, when values above the cut-off are considered positive detections and values below considered negative, maximised the odds ratio for detection in cases vs controls. These cut-offs for each virus are shown as red dashed lines in Figure 46.

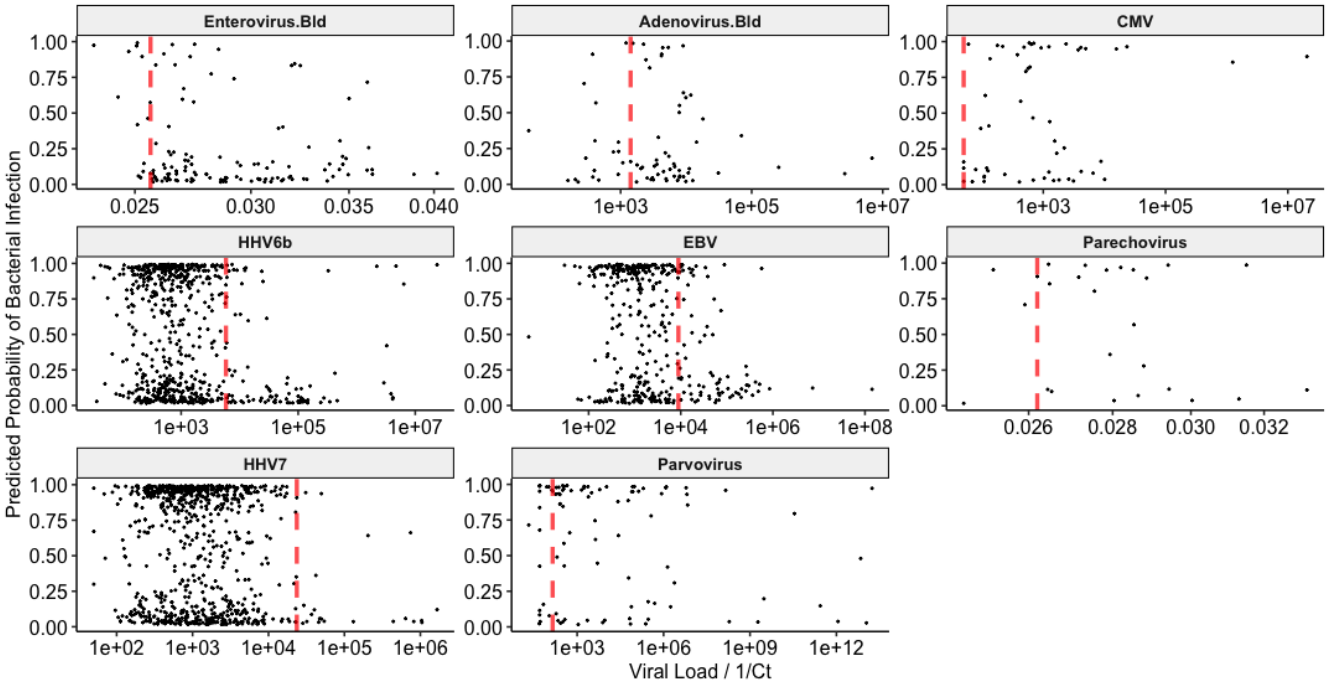



*Figure 46: Scatterplots showing predicted probability of bacterial infection at increasing viral load for viruses detected in blood. For enterovirus and parechovirus, the inverse of the Ct value is used as a surrogate marker for pathogen load. The optimal viral load cut-off limit to maximise odds of detection in cases vs controls for each pathogen in shown by the red dashed line.*

When only detections with pathogen load above these cut-offs were considered as positive detections, the detection of EBV, HHV6b and HHV7 were associated with lower PPBIs (p = 0.004, <0.0001, 0.0001 respectively). There was no significant change in the PPBI in other viruses.

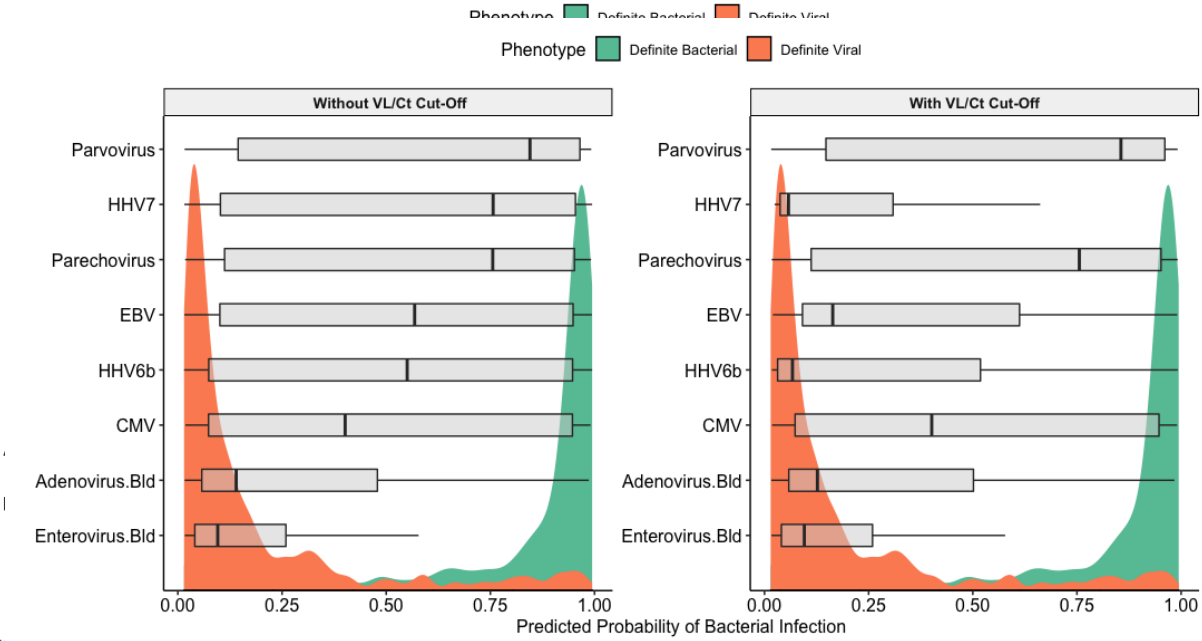



*Figure 47: Boxplots showing median and IQR predicted probability of bacterial infection before and after the application of viral load cut-off limits.*

### *6.2.5.2 Expression in patients with bacteria detected in blood*

In patients in whom only bacteria were detected, PPBIs were between those of DB and DV patients and not statistically equivalent to either (Figure 48).

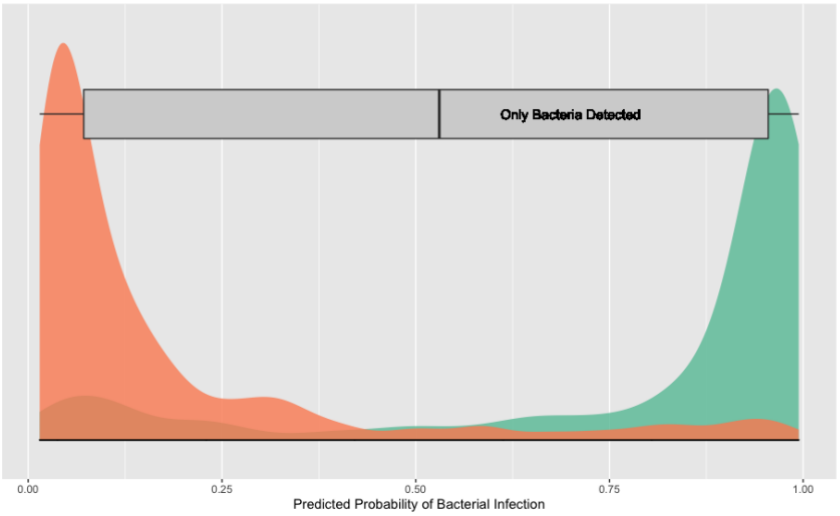



*Figure 48: Boxplot showing median and IQR predicted probability of bacterial infection in febrile children with detection of bacteria and no viral co-detections. This boxplot has been superimposed on a density plot showing distribution of probabilities in DB (green) and DV (orange) patients.*

There were 5 bacteria with more than 20 detections in febrile children with both CMT and gene expression results available and the PPBIs associated with the detection of each of these is show in Figure 49. Of these, only the detection of *N.meningitidis* and *S.pneumoniae* were associated with PPBIs that were statistically equivalent to DB patients (p = 0.036, 0.023 respectively).

As described in chapter 3, there was no clear relationship between Ct values of bacterial assays and the odds ratio of detection in cases vs controls or DB vs DV patients. Furthermore, gene expression analysis shows there is no relationship between Ct value of bacterial assays and the PPBI.

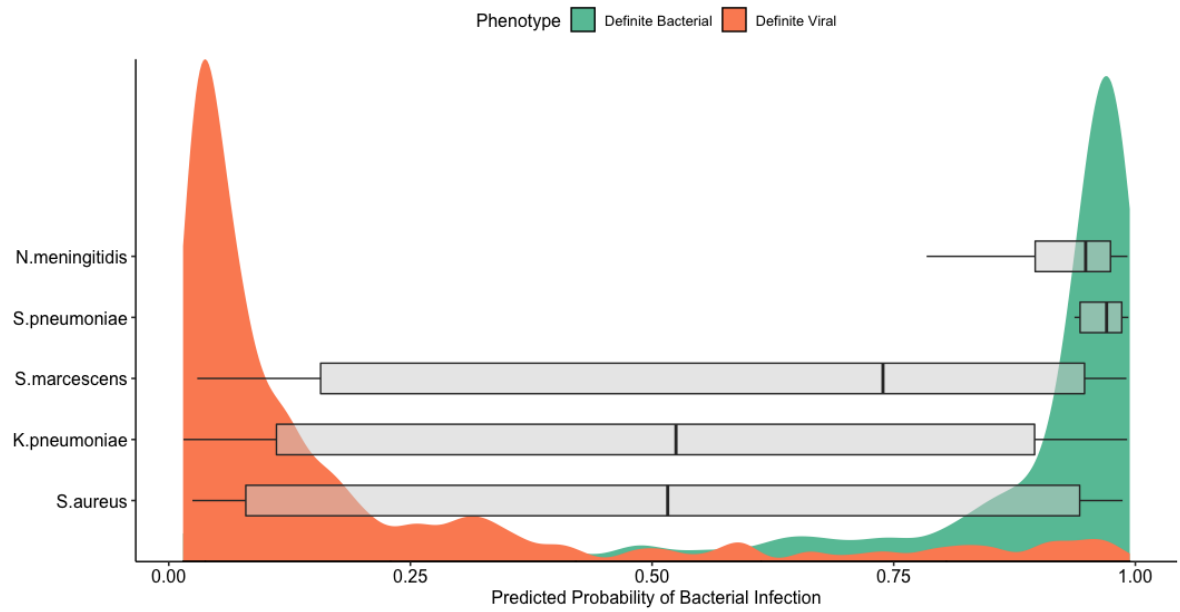



*Figure 49: Boxplots showing median and IQR predicted probability of bacterial infection (PPBI) when each bacteria is detected. Boxplots are superimposed on a density plot showing the distribution of PPBIs in DV (orange) and DB (green) patients. Only bacteria that were detected more than 20 times in patients with both CMT and gene expression data are displayed.*

### *6.2.5.3 Expression in patients with co-detections*

In chapter 4 I showed that there were common patterns of pathogen co-detection. There were insufficient numbers of patients with both CMT and gene expression results available with detection of each combination of pathogens for robust statistical analyses. Instead, I compared PPBIs from patients with the most frequently detected viruses, with and without co-detection of bacteria, and when each of the most frequently detected bacteria were detected with and without the co-detection of viruses. This analysis was limited to viruses that were detected in at least 20 patients with and without co-detection of any bacteria, and bacteria detected in at least 20 patients with and without co-detection of any virus.

#### Respiratory Viruses

When influenza B, enterovirus and rhinovirus were co-detected with bacteria the PPBIs were significantly higher than when no bacteria were detected (Figure 50). PPBIs of patients with influenza A detection remained low whether co-detected with bacteria or detected in isolation.

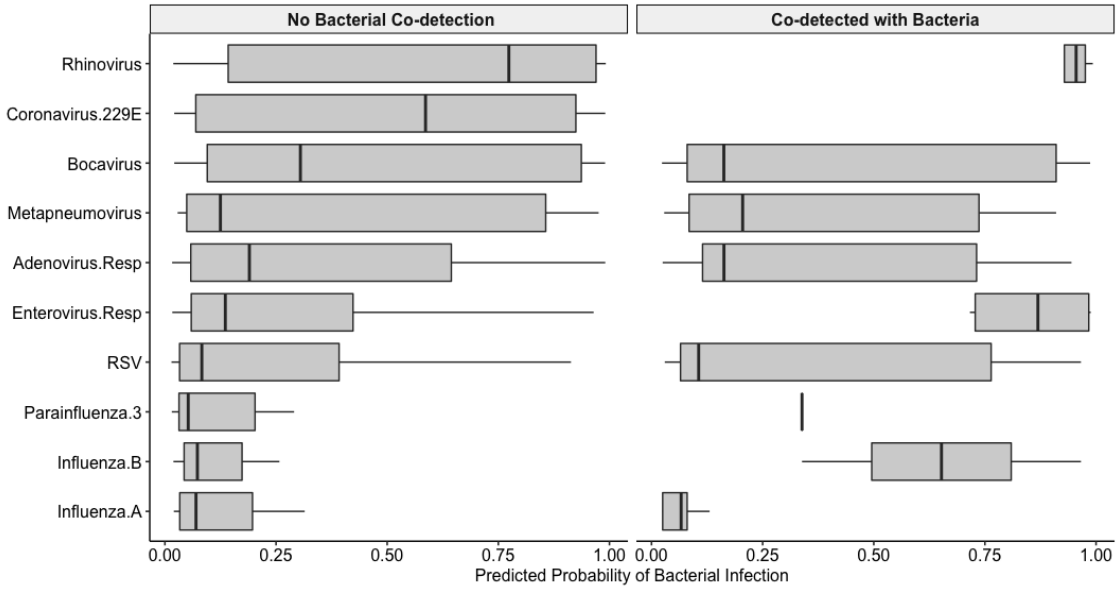



*Figure 50: Boxplots showing median and IQR of the PPBI score for patients with respiratory viruses when detected with and without bacterial co-detections. Only viruses detected in more than 20 patients are displayed.*

## Blood Viruses

When parvovirus was detected in blood, PPBI was significantly higher when bacteria were co-detected compared to parvovirus in the absence of bacteria (Figure 51). PPBIs were low when enterovirus and adenovirus were detected, whether when co-detected with bacteria or detected in isolation.

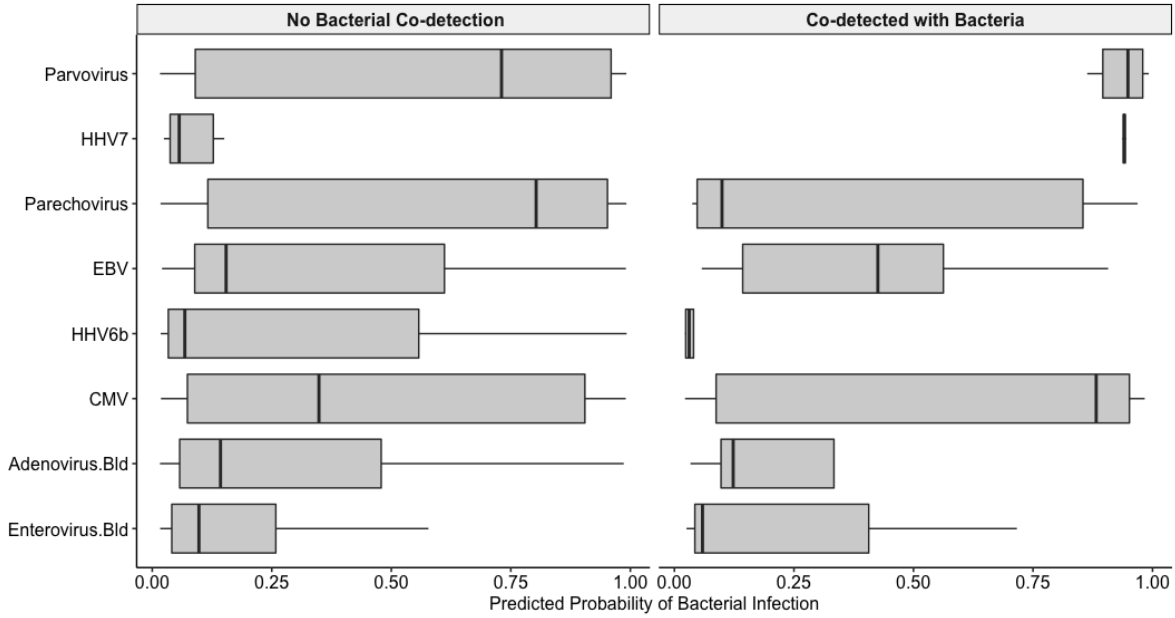



*Figure 51: Boxplots showing median and IQR of blood viruses when detected with and without bacterial co-detections. Only viruses detected in more than 20 patients are displayed.*

## Blood Bacteria

The detection of *S.pneumoniae* was associated with a significantly higher PPBI when co-detected with viruses than when detected in the absence of viruses (Figure 52). The PPBI associated with the detection of other bacteria was unaffected by co-detection with viruses.

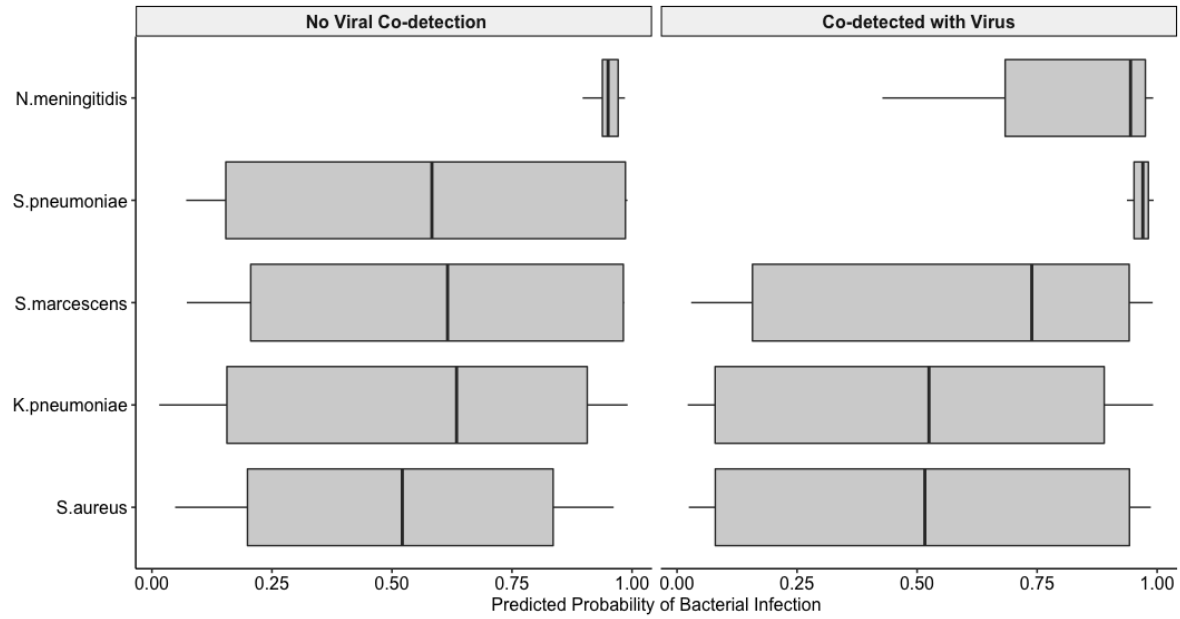



*Figure 52: Boxplots showing median and IQR of bacteria when detected with and without viral co-detections. Only bacteria detected in more than 20 patients are displayed.*

## 6.3 Discussion

I aimed to investigate the host response to the presence of different pathogens and different clinical syndromes to better understand febrile illnesses in children.

As reported in chapter 3, pathogens were detected in both cases and controls. However, there was no significant difference in expression of any of the genes in the 12-gene signature when comparing controls with pathogen detections and those without, suggesting that the pathogens detected in controls are not eliciting a host response that is detectable by this gene expression signature. However, when I compared the host response in cases vs controls, there was a measurable and significant difference in expression of genes in the 12-gene signature.

This gene expression signature was developed to distinguish patients with bacterial and viral infection, using a support vector machine model to analyse expression of the 12 genes that make up the signature to output a predicted probability of bacterial infection. As discussed in chapters 3 and 4, the complex nature of infections makes it difficult to categorise into those caused by bacteria and those caused by viruses. In febrile children, DB patients were associated with very high PPBIs and DV patients had very low PPBIs. Controls were associated with PPBIs similar to DB patients, despite the absence of any difference in expression of any of the 12 genes in controls with and without pathogen detection. This is likely because this model was developed on a binary model of infection, with the intention of distinguishing bacterial and viral disease in patients with known infection. As the viral genes have greater change in expression values in cases vs controls than bacterial genes, this likely forces the model to interpret controls as having a bacterial-like PPBI. This made it difficult to infer causality from pathogen detection based on the PPBI but this can be avoided by limiting the analysis to febrile patients, making it possible to use PPBI in DB and DV patients to assess whether there was a viral-like or bacterial-like host response.
Some syndromes, such as non-specific URTIs and flu-like illness were associated with a viral-like host response whereas others, such as UTI and pyelonephritis were associated with a bacterial-like host response. Other syndromes were associated with PPBIs that were significantly higher than DV patients, lower than DB patients or both. The detection of influenza A and B in respiratory samples and enterovirus in blood was associated with a viral-type host response and the detection of *N.meningitidis* and *S.pneumoniae* in blood was associated with a bacterial-type host response. This may be because patients used in the discovery of this signature had bacterial or viral aetiology that could be confirmed by conventional diagnostic techniques and tended to have more severe disease. The

heterogenicity of PPBIs in patients with other syndromes and with the detection of other pathogens supports the findings in chapter 3 and 4 that infectious processes are more complex than a simple dichotomy of bacterial and viral infection. In most cases a specific clinical syndrome or the detection of a specific pathogen does not lead to a host response that is clearly bacterial or viral. It may be that this 12-gene signature is not able to pick up the subtle differences in the host response to infection by different pathogens or at different sites. Alternatively, it may be because there are situations in which viruses can result in a more bacterial-like response or bacteria result in a more viral-like response[239].

When viral load lower cut-off limits that maximised odds ratios for detection of viruses in cases vs controls were applied, the PPBI of patients with EBV, HHB6b and HHV7 was significantly lower. This suggests that these herpes viruses elicit a viral-type host response at higher viral loads and detections at lower viral loads represent detection of circulating virus not causing active disease, as has previously been postulated[240,241].

Co-detection also appeared to alter the host response to infection. The co-detection of bacteria in patients with rhinovirus, enterovirus in respiratory samples and parvovirus in blood resulted in a more bacterial-like response, whereas the co-detection of bacteria with influenza A in respiratory samples and enterovirus and adenovirus in blood did not lead to an increase in PPBI. Most notably, the co-detection of viruses with *S.pneumoniae* resulted in significantly higher PPBIs than when *S.pneumoniae* was detected without viruses. This may support two proposed theories of co-detection: that viruses impair host defences and allow known commensal bacteria to invade into sites where they subsequently cause infection[82,115,242] or that some bacterial-viral combinations have a synergistic effect, leading to a more bacterial-like host response. Another potential explanation is that bacterial

detection may represent either the detection of non-disease causing bacterial DNA or active bacterial disease, which is more likely when bacteria are also detected.

One of the major limitations of this study is that, while informative, this 12-gene signature only reveals a small part of the host response to infections. As it was developed with the aim of differentiating bacterial from viral infections and discovered using a cohort of patients with bacteria and viruses that could be confidently related to aetiology, the gene signature may not capture some of the more subtle aspects of the immune response that may be specific to a pathogen or group of pathogens or to a site of infection. To ensure that the pathogen tests performed on each patient, and therefore the pathogen data was standardised, and as pathogen detections in the patient's clinical course largely influenced their phenotype, only pathogen data from CMTs was included in this analysis. This may omit a small number of pathogens detected by conventional diagnostics and not CMT. In addition, although the number of recruited patients was very large and more representative of children presenting to emergency departments than most paediatric studies of this kind, the number of positive detections of some pathogens or combinations of pathogens, or the number of patients with rarer clinical syndromes was still quite small.

Despite this, this study has shown that host transcriptomics can be used to better understand febrile illnesses. The addition of chronology – with molecular diagnostic results and gene expression data that spans the course of a patient's illness will add a further dimension to future analyses, allowing the processes by which co-detections occur to be studied. With increased numbers of patients with each pathogen or type of infection, expression data on a wider set of genes and at multiple time points during the illness, it may

be possible to investigate how pathogen detection relates to pathogenicity and how to interpret the huge amount of pathogen data revealed by molecular diagnostics.

# 7. Overall Conclusions and Discussion

Deciphering the aetiology of a child presenting with a febrile illness and decided the optimal management plan remains a common problem facing clinicians in primary and secondary healthcare settings. The uncertainty surrounding aetiology means that clinicians must strike a balance between the risk of misdiagnosing a child with serious infection requiring antimicrobial therapy and the risks associated with overuse of antibiotics.

This has directed much research into the development of novel diagnostic approaches to the febrile child. The arrival of highly sensitive molecular diagnostic techniques revealed additional pathogen data, highlighting the complexity of childhood infections and leading to the discovery of a number of new pathogens. However, it remains unclear how to infer aetiology, and therefore the most appropriate management, from pathogen detection.

Transcriptomic approaches have shown that it is possible to accurately to distinguish patients with confirmed bacterial infection from those with clear viral infection by studying expression of a small gene signature. However, these approaches are less effective at deciphering aetiology in children where the cause of infection is less clear.

Conventional diagnostic techniques, on which the discovery of these gene expression signatures have been based, omit a large amount of pathogen data, and present an overly simple representation of infectious diseases. The addition of data from CMTs reveals the complexity of infectious diseases, with pathogen co-detections common in some types of infections. However, with the exception of some respiratory assays and blood assays for

*S.pneumoniae* and *N.meningitidis* it is unclear how to relate pathogen detection by CMT to the aetiology of a child's illness or the need for antimicrobial intervention.

This analysis suggests that the development of novel diagnostics in infectious diseases needs to move away from categorising patients into those with bacterial and those with viral disease and move to a model that highlights patients based on their need for intervention, such as antibiotic treatment, or other interventions such as respiratory or cardiovascular support.

Gene expression analysis, even on a small 12-gene signature is able to identify differences in the host response to different pathogens, combinations of pathogens and clinical syndromes. The detection of some pathogens was only associated with altered gene expression at higher viral loads, suggesting gene expression analysis can distinguish between active infection and incidental detection of some pathogens. The differences in host response when pathogens are co-detected, such as the more bacterial-like host response associated with *S.pneumoniae* when co-detected with viruses, suggests that viruses may predispose to or act synergistically with bacteria.

This gene expression signature and accompanying statistical model assigned every febrile patient a numerical output that represented the predicted probability of bacterial infection (PPBI). DV patients were associated with lower PPBIs and DB patients were associated with higher PPBIs, suggesting that low PPBIs indicated a viral-like host response and high PPBIs indicate a bacterial-like host response. PPBIs in febrile children in our cohort, regardless of phenotype or clinical syndrome showed a bimodal distribution, with very high or very low

PPBIs in most patients. This model can assign most patients, based on host gene expression, as having a viral-like or bacterial-like host response, although it remains unclear whether these assignments are correct. Furthermore, many clinical syndromes and the detection of many pathogens were associated with heterogenous PPBIs. This may be because this gene expression model was based on a binary classification of infection and discovered on data that does not represent the full complexity of febrile illnesses in children.

This report shows the utility of gene expression studies in understanding infections – both the link between pathogen and host response, but also the processes by which one pathogen facilitates infection by another. Further analysis of additional pathogen data from patients with less common infections and expression data from a wider set of genes at multiple timepoints, will allow for better understanding of when pathogen detection requires intervention. This could form the foundation for novel diagnostics for infectious diseases which are discovered on data that is more representative of the true state of infection, and may therefore have the ability to direct the optimal management of febrile patients whose aetiology is unclear. To maximise their clinical utility, these novel diagnostics should incorporate febrile illnesses and inflammatory diseases which may mimic but are not directly related to infection, such as KD and IgAV.

# 8. References


1. Eilís Keeble & Lucia Kossarova. *Focus on: Emergency hospital care for children and young people*. https://www.nuffieldtrust.org.uk/research/focus-on-emergency-hospital-care-for-children-and-young-people (2017).
2. Child mortality – RCPCH – State of Child Health. https://stateofchildhealth.rcpch.ac.uk/evidence/mortality/child-mortality/.
3. World health statistics 2022: monitoring health for the SDGs, sustainable development goals. https://www.who.int/publications-detail-redirect/9789240051157.
4. Bryce, J., Boschi-Pinto, C., Shibuya, K. & Black, R. E. WHO estimates of the causes of death in children. *The Lancet* **365**, 1147–1152 (2005).
5. Biserni, G. B. *et al.* Potential Diagnostic and Prognostic Biomarkers for Adenovirus Respiratory Infection in Children and Young Adults. *Viruses* **13**, (2021).
6. Chuang, Y.-Y. & Huang, Y.-C. Enteroviral infection in neonates. *J. Microbiol. Immunol. Infect. Wei Mian Yu Gan Ran Za Zhi* **52**, 851–857 (2019).
7. Diaz-Diaz, A. *et al.* Viral Bacterial Interactions in Children: Impact on Clinical Outcomes. *Pediatr. Infect. Dis. J.* **38**, S14–S19 (2019).
8. Sidoti, F., Rittà, M., Costa, C. & Cavallo, R. Diagnosis of viral gastroenteritis: limits and potential of currently available procedures. *J. Infect. Dev. Ctries.* **9**, 551–561 (2015).
9. Gutierrez Sanchez, L. H. *et al.* A Case Series of Children with Acute Hepatitis and Human Adenovirus Infection. *N. Engl. J. Med.* **387**, 620–630 (2022).
10. Desforges, M. *et al.* Human Coronaviruses and Other Respiratory Viruses: Underestimated Opportunistic Pathogens of the Central Nervous System? *Viruses* **12**, (2019).

11. Whittaker, E. *et al.* Clinical Characteristics of 58 Children With a Pediatric Inflammatory Multisystem Syndrome Temporally Associated With SARS-CoV-2. *JAMA* **324**, 259 (2020).
12. Op de Beeck, A. & Eizirik, D. L. Viral infections in type 1 diabetes mellitus — why the β cells? *Nat. Rev. Endocrinol.* **12**, 263–273 (2016).
13. Beigelman, A. & Bacharier, L. B. The role of early life viral bronchiolitis in the inception of asthma. *Curr. Opin. Allergy Clin. Immunol.* **13**, 211–216 (2013).
14. Zhu, T., Zhang, L., Qu, Y. & Mu, D. Meta-analysis of antenatal infection and risk of asthma and eczema. *Medicine (Baltimore)* **95**, e4671 (2016).
15. Bansal, A. S., Bradley, A. S., Bishop, K. N., Kiani-Alikhan, S. & Ford, B. Chronic fatigue syndrome, the immune system and viral infection. *Brain. Behav. Immun.* **26**, 24–31 (2012).
16. Hickie, I. *et al.* Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. *BMJ* **333**, 575 (2006).
17. Sudre, C. H. *et al.* Attributes and predictors of long COVID. *Nat. Med.* **27**, 626–631 (2021).
18. Hanada, S., Pirzadeh, M., Carver, K. Y. & Deng, J. C. Respiratory Viral Infection-Induced Microbiome Alterations and Secondary Bacterial Pneumonia. *Front. Immunol.* **9**, (2018).
19. Rynda-Apple, A., Robinson, K. M. & Alcorn, J. F. Influenza and Bacterial Superinfection: Illuminating the Immunologic Mechanisms of Disease. *Infect. Immun.* **83**, 3764–3770 (2015).
20. Davidson, S., Maini, M. K. & Wack, A. Disease-promoting effects of type I interferons in viral, bacterial, and coinfections. *J. Interferon Cytokine Res. Off. J. Int. Soc. Interferon Cytokine Res.* **35**, 252–264 (2015).
21. Klein, E. Y. *et al.* The frequency of influenza and bacterial coinfection: a systematic review and meta-analysis. *Influenza Other Respir. Viruses* **10**, 394–403 (2016).

22. Everard, M. L. Paediatric respiratory infections. *Eur. Respir. Rev. Off. J. Eur. Respir. Soc.* **25**, 36–40 (2016).

23. Greenhow, T. L., Hung, Y.-Y., Herz, A. M., Losada, E. & Pantell, R. H. The Changing Epidemiology of Serious Bacterial Infections in Young Infants. *Pediatr. Infect. Dis. J.* **33**, 595–599 (2014).

24. Sanneh, B. *et al.* Declining Trends of Pneumococcal Meningitis in Gambian Children After the Introduction of Pneumococcal Conjugate Vaccines. *Clin. Infect. Dis.* **69**, S126–S132 (2019).

25. Acute bacterial meningitis in infants and children | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S1473309909703068?token=AE8D97C02183DDCE2905A42613BF2A70F20D9C12558AF7241984D9E7ED25F90AFA0AD82BC130D0658FDECFE4E0060113&originRegion=eu-west-1&originCreation=20220705125748 doi:10.1016/S1473-3099(09)70306-8.

26. Bleeker, S., Moons, K., Derksen-Lubsen, G., Grobbee, D. & Moll, H. Predicting serious bacterial infection in young children with fever without apparent source. *Acta Paediatr.* **90**, 1226–1231 (2001).

27. Takata, J., Kelly, D. F., Sadarangani, M., Jeffery, K. & Drysdale, S. B. 14-year trends and resistance patterns of blood and cerebrospinal fluid cultures in children under three years old. *J. Infect.* **83**, 533–541 (2021).

28. Manzano, S. *et al.* Markers for bacterial infection in children with fever without source. *Arch. Dis. Child.* **96**, 440–446 (2011).

29. Huber, S., Hetzer, B., Crazzolara, R. & Orth-Höller, D. The correct blood volume for paediatric blood cultures: a conundrum? *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **26**, 168–173 (2020).

30. Astete, J. A. *et al.* Blood culture diagnostic yield in a paediatric emergency department. *Eur. J. Emerg. Med.* **21**, 336–340 (2014).

31. Buller, R. S. Molecular detection of respiratory viruses. *Clin. Lab. Med.* **33**, 439–460 (2013).

32. Echavarría, M. *et al.* Clinical impact of rapid molecular detection of respiratory pathogens in patients with acute respiratory infection. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **108**, 90–95 (2018).

33. Martinón-Torres, F. *et al.* Life-threatening infections in children in Europe (the EUCLIDS Project): a prospective cohort study. *Lancet Child Adolesc. Health* **2**, 404–414 (2018).

34. Hagedoorn, N. N. *et al.* Variation in antibiotic prescription rates in febrile children presenting to emergency departments across Europe (MOFICHE): A multicentre observational study. *PLOS Med.* **17**, e1003208 (2020).

35. Paterson, I. K., Hoyle, A., Ochoa, G., Baker-Austin, C. & Taylor, N. G. H. Optimising Antibiotic Usage to Treat Bacterial Infections. *Sci. Rep.* **6**, 37853 (2016).

36. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629–655 (2022).

37. Langdon, A., Crook, N. & Dantas, G. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Med.* **8**, 39 (2016).

38. Hviid, A., Svanström, H. & Frisch, M. Antibiotic use and inflammatory bowel diseases in childhood. *Gut* **60**, 49–54 (2011).

39. Vetter, P., Schibler, M., Herrmann, J. L. & Boutolleau, D. Diagnostic challenges of central nervous system infection: extensive multiplex panels versus stepwise guided approach. *Clin. Microbiol. Infect.* **26**, 706–712 (2020).

40. Atallah, J. & Mansour, M. K. Implications of Using Host Response-Based Molecular Diagnostics on the Management of Bacterial and Viral Infections: A Review. *Front. Med.* **9**, (2022).

41. Chiu, R. W. K., Lo, Y. M. D. & Wittwer, C. T. Molecular diagnostics: a revolution in progress. *Clin. Chem.* **61**, 1–3 (2015).

42. Grumaz, S. *et al.* Next-generation sequencing diagnostics of bacteremia in septic patients. *Genome Med.* **8**, 73 (2016).

43. Ganova-Raeva, L. M. & Khudyakov, Y. E. Application of mass spectrometry to molecular diagnostics of viral infections. *Expert Rev. Mol. Diagn.* **13**, 377–388 (2013).

44. Jartti, T., Söderlund-Venermo, M., Hedman, K., Ruuskanen, O. & Mäkelä, M. J. New molecular virus detection methods and their clinical value in lower respiratory tract infections in children. *Paediatr. Respir. Rev.* **14**, 38–45 (2013).

45. Yang, S. & Rothman, R. E. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect. Dis.* **4**, 337–348 (2004).

46. Kash, J. C. & Taubenberger, J. K. The role of viral, host, and secondary bacterial factors in influenza pathogenesis. *Am. J. Pathol.* **185**, 1528–1536 (2015).

47. Rowe, H. M. & Rosch, J. W. Close Encounters of the Viral Kind: Cross-Kingdom Synergies at the Host-Pathogen Interface. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **41**, e1900128 (2019).

48. Lian, S., Liu, J., Wu, Y., Xia, P. & Zhu, G. Bacterial and Viral Co-Infection in the Intestine: Competition Scenario and Their Effect on Host Immunity. *Int. J. Mol. Sci.* **23**, (2022).

49. Berger, A. K. & Mainou, B. A. Interactions between Enteric Bacteria and Eukaryotic Viruses Impact the Outcome of Infection. *Viruses* **10**, (2018).

50. Karppinen, S., Toivonen, L., Schuez-Havupalo, L., Waris, M. & Peltola, V. Interference between respiratory syncytial virus and rhinovirus in respiratory tract infections in children. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **22**, 208.e1-208.e6 (2016).

51. Kaforou, M. *et al.* Detection of Tuberculosis in HIV-Infected and -Uninfected African Adults Using Whole Blood RNA Expression Signatures: A Case-Control Study. *PLoS Med.* **10**, e1001538 (2013).

52. 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).

53. Lee, H. J. *et al.* Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. *Sci. Transl. Med.* **10**, eaar3619 (2018).

54. Wright, V. J. *et al.* Diagnosis of Kawasaki Disease Using a Minimal Whole-Blood Gene Expression Signature. *JAMA Pediatr.* **172**, e182293–e182293 (2018).

55. 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 (2016).

56. 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 (2017).

57. 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**, 8043 (2018).

58. 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. U. S. A.* **110**, 12792–12797 (2013).

59. Bhattacharya, S. *et al.* Transcriptomic Biomarkers to Discriminate Bacterial from Nonbacterial Infection in Adults Hospitalized with Respiratory Illness. *Sci. Rep.* **7**, 6548 (2017).

60. Lydon, E. C. *et al.* Validation of a host response test to distinguish bacterial and viral respiratory infection. *EBioMedicine* **48**, 453–461 (2019).

61. Fraunberger, P. *et al.* PROGNOSTIC VALUE OF INTERLEUKIN 6, PROCALCITONIN, AND C-REACTIVE PROTEIN LEVELS IN INTENSIVE CARE UNIT PATIENTS DURING FIRST INCREASE OF FEVER. *Shock* **26**, 10–12 (2006).

62. Wu, C.-W. *et al.* Does procalcitonin, C-reactive protein, or interleukin-6 test have a role in the diagnosis of severe infection in patients with febrile neutropenia? A systematic review and meta-analysis. *Support. Care Cancer* **23**, 2863–2872 (2015).

63. MeMed. MeMed Receives CE Mark for Two Pioneering Technologies: The MeMed BV™ Test and MeMed Key™ Point-of-Need Platform. *GlobeNewswire News Room* https://www.globenewswire.com/news-release/2020/06/02/2042093/0/en/MeMed-Receives-CE-Mark-for-Two-Pioneering-Technologies-The-MeMed-BV-Test-and-MeMed-Key-Point-of-Need-Platform.html (2020).

64. Leticia Fernandez-Carballo, B., Escadafal, C., MacLean, E., Kapasi, A. J. & Dittrich, S. Distinguishing bacterial versus non-bacterial causes of febrile illness – A systematic review of host biomarkers. *J. Infect.* **82**, 1–10 (2021).

65. Ducharme, J. *et al.* A Multi-mRNA Host-Response Molecular Blood Test for the Diagnosis and Prognosis of Acute Infections and Sepsis: Proceedings from a Clinical Advisory Panel. *J. Pers. Med.* **10**, 266 (2020).

66. Safarika, A. *et al.* A 29-mRNA host response test from blood accurately distinguishes bacterial and viral infections among emergency department patients. *Intensive Care Med. Exp.* **9**, 31 (2021).

67. Chiu, C. Y. & Miller, S. A. Clinical metagenomics. *Nat. Rev. Genet.* **20**, 341–355 (2019).

68. Chen, Y., Fan, L., Chai, Y. & Xu, J. Advantages and challenges of metagenomic sequencing for the diagnosis of pulmonary infectious diseases. *Clin. Respir. J.* **16**, 646–656 (2022).

69. Lim, H. J., Saha, T., Tey, B. T., Tan, W. S. & Ooi, C. W. Quartz crystal microbalance-based biosensors as rapid diagnostic devices for infectious diseases. *Biosens. Bioelectron.* **168**, 112513 (2020).

70. Soler, M., Estevez, M. C., Cardenosa-Rubio, M., Astua, A. & Lechuga, L. M. How Nanophotonic Label-Free Biosensors Can Contribute to Rapid and Massive Diagnostics of Respiratory Virus Infections: COVID-19 Case. *ACS Sens.* **5**, 2663–2678 (2020).

71. R Core Team (2020). R: A language and environment for statistical computing. (2020).

72. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J. Stat. Softw.* **39**, (2011).

73. Boppana, S. B. Cytomegalovirus. in *Congenital and Perinatal Infections* (ed. Hutto, C.) 73–86 (Humana Press, 2006). doi:10.1385/1-59259-965-6:073.

74. Pass, R. F. HHV6 and HHV7: Persistence and vertical transmission. *J. Pediatr.* **145**, 432–435 (2004).

75. Thorley-Lawson, D. A., Hawkins, J. B., Tracy, S. I. & Shapiro, M. The pathogenesis of Epstein–Barr virus persistent infection. *Curr. Opin. Virol.* **3**, 227–232 (2013).

76. Morris, D. E., Cleary, D. W. & Clarke, S. C. Secondary Bacterial Infections Associated with Influenza Pandemics. *Front. Microbiol.* **8**, (2017).

77. Smith, A. M. & McCullers, J. A. Secondary Bacterial Infections in Influenza Virus Infection Pathogenesis. in *Influenza Pathogenesis and Control - Volume I* (eds. Compans, R. W. & Oldstone, M. B. A.) vol. 385 327–356 (Springer International Publishing, 2014).

78. Morgan, D. J. *et al.* Innate Immune Cell Suppression and the Link With Secondary Lung Bacterial Pneumonia. *Front. Immunol.* **9**, 2943 (2018).

79. McCullers, J. A. The co-pathogenesis of influenza viruses with bacteria in the lung. *Nat. Rev. Microbiol.* **12**, 252–262 (2014).

80. Hernández, S. *et al.* Invasive Pneumococcal Disease and Influenza Activity in a Pediatric Population: Impact of PCV13 Vaccination in Pandemic and Nonpandemic Influenza Periods. *J. Clin. Microbiol.* **57**, e00363-19, /jcm/57/8/JCM.00363-19.atom (2019).

81. Dangor, Z. *et al.* Temporal Association in Hospitalizations for Tuberculosis, Invasive Pneumococcal Disease and Influenza Virus Illness in South African Children. *PLoS ONE* **9**, e91464 (2014).

82. Berry, I. *et al.* Association of Influenza Activity and Environmental Conditions With the Risk of Invasive Pneumococcal Disease. *JAMA Netw. Open* **3**, e2010167 (2020).

83. Talbot, T. R. *et al.* Seasonality of invasive pneumococcal disease: Temporal relation to documented influenza and respiratory syncytial viral circulation. *Am. J. Med.* **118**, 285–291 (2005).

84. Salomon, A. *et al.* Influenza increases invasive meningococcal disease risk in temperate countries. *Clin. Microbiol. Infect.* **26**, 1257.e1-1257.e7 (2020).

85. Jacobs, J. H. *et al.* The Association of Meningococcal Disease with Influenza in the United States, 1989–2009. *PLoS ONE* **9**, e107486 (2014).

86. Brundage, J. F. & Shanks, G. D. Deaths from Bacterial Pneumonia during 1918–19 Influenza Pandemic. *Emerg. Infect. Dis.* **14**, 1193–1199 (2008).

87. Jansen, A. G. S. C. *et al.* Invasive pneumococcal and meningococcal disease: association with influenza virus and respiratory syncytial virus activity? *Epidemiol. Infect.* **136**, 1448–1454 (2008).

88. Glennie, S. *et al.* Modulation of nasopharyngeal innate defenses by viral coinfection predisposes individuals to experimental pneumococcal carriage. *Mucosal Immunol.* **9**, 56–67 (2016).

89. Wolter, N. *et al.* High Nasopharyngeal Pneumococcal Density, Increased by Viral Coinfection, Is Associated With Invasive Pneumococcal Pneumonia. *J. Infect. Dis.* **210**, 1649–1657 (2014).

90. Vu, H. T. T. *et al.* Association Between Nasopharyngeal Load of Streptococcus pneumoniae, Viral Coinfection, and Radiologically Confirmed Pneumonia in Vietnamese Children. *Pediatr. Infect. Dis. J.* **30**, 11–18 (2011).

91. Weinberger, D. M., Klugman, K. P., Steiner, C. A., Simonsen, L. & Viboud, C. Association between Respiratory Syncytial Virus Activity and Pneumococcal Disease in Infants: A Time Series Analysis of US Hospitalization Data. *PLoS Med.* **12**, e1001776 (2015).

92. Cebey-López, M. *et al.* Bacteremia in Children Hospitalized with Respiratory Syncytial Virus Infection. *PLOS ONE* **11**, e0146599 (2016).

93. Klugman, K. P., Madhi, S. A., Ginsburg, A. S. & Rodgers, G. L. The role of bacterial vaccines in the prevention of influenza mortality. *Lancet Glob. Health* **6**, e1268–e1269 (2018).

94. Simonsen, L. *et al.* Impact of Pneumococcal Conjugate Vaccination of Infants on Pneumonia and Influenza Hospitalization and Mortality in All Age Groups in the United States. *mBio* **2**, e00309-10 (2011).

95. Domínguez, A. *et al.* Benefit of Conjugate Pneumococcal Vaccination in Preventing Influenza Hospitalization in Children: A Case-control Study. *Pediatr. Infect. Dis. J.* **32**, 330–334 (2013).

96. Abadom, T. R. *et al.* Risk factors associated with hospitalisation for influenza-associated severe acute respiratory illness in South Africa: A case-population study. *Vaccine* **34**, 5649–5655 (2016).

97. Madhi, S. A., Klugman, K. P., & The Vaccine Trialist Group. A role for Streptococcus pneumoniae in virus-associated pneumonia. *Nat. Med.* **10**, 811–813 (2004).

98. Cebey-López, M. *et al.* Does Viral Co-Infection Influence the Severity of Acute Respiratory Infection in Children? *PLOS ONE* **11**, e0152481 (2016).

99. Parsons, K. S., Hsu, A. C. & Wark, P. A. B. TLR3 and MDA5 signalling, although not expression, is impaired in asthmatic epithelial cells in response to rhinovirus infection. *Clin. Exp. Allergy* **44**, 91–101 (2014).

100. Sun, K. & Metzger, D. W. Influenza Infection Suppresses NADPH Oxidase–Dependent Phagocytic Bacterial Clearance and Enhances Susceptibility to Secondary Methicillin-Resistant *Staphylococcus aureus* Infection. *J. Immunol.* **192**, 3301–3307 (2014).

101. Diavatopoulos, D. A. *et al.* Influenza A virus facilitates *Streptococcus pneumoniae* transmission and disease. *FASEB J.* **24**, 1789–1798 (2010).

102. Goulding, J. *et al.* Lowering the Threshold of Lung Innate Immune Cell Activation Alters Susceptibility to Secondary Bacterial Superinfection. *J. Infect. Dis.* **204**, 1086–1094 (2011).

103. Mina, M. J., Klugman, K. P. & McCullers, J. A. Live Attenuated Influenza Vaccine, But Not Pneumococcal Conjugate Vaccine, Protects Against Increased Density and Duration of

Pneumococcal Carriage After Influenza Infection in Pneumococcal Colonized Mice. *J. Infect. Dis.* **208**, 1281–1285 (2013).

104. Iverson, A. R. *et al.* Influenza Virus Primes Mice for Pneumonia From Staphylococcus aureus. *J. Infect. Dis.* **203**, 880–888 (2011).
105. D'Acremont, V. *et al.* Beyond Malaria — Causes of Fever in Outpatient Tanzanian Children. *N. Engl. J. Med.* **370**, 809–817 (2014).
106. Zar, H. J. *et al.* Aetiology of childhood pneumonia in a well vaccinated South African birth cohort: a nested case-control study of the Drakenstein Child Health Study. *Lancet Respir. Med.* **4**, 463–472 (2016).
107. Bénet, T. *et al.* Severity of Pneumonia in Under 5-Year-Old Children from Developing Countries: A Multicenter, Prospective, Observational Study. *Am. J. Trop. Med. Hyg.* **97**, 68–76 (2017).
108. Bénet, T. *et al.* Microorganisms Associated With Pneumonia in Children <5 Years of Age in Developing and Emerging Countries: The GABRIEL Pneumonia Multicenter, Prospective, Case-Control Study. *Clin. Infect. Dis.* (2017) doi:10.1093/cid/cix378.
109. Selwyn, B. J. The Epidemiology of Acute Respiratory Tract Infection in Young Children: Comparison of Findings from Several Developing Countries. *Clin. Infect. Dis.* **12**, S870–S888 (1990).
110. O'Brien, K. L. *et al.* Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *The Lancet* **394**, 757–779 (2019).
111. Respiratory syncytial virus (RSV) and its propensity for causing bronchiolitis - Pickles - 2015 - The Journal of Pathology - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/path.4462.

112. Tzanakaki, G. & Mastrantonio, P. Aetiology of bacterial meningitis and resistance to antibiotics of causative pathogens in Europe and in the Mediterranean region. *Int. J. Antimicrob. Agents* **29**, 621–629 (2007).

113. Escherichia coli mediated urinary tract infections: Are there distinct uropathogenic E. coli (UPEC) pathotypes? | FEMS Microbiology Letters | Oxford Academic. https://academic.oup.com/femsle/article/252/2/183/528405?login=false.

114. Kloepfer, K. M. *et al.* Detection of pathogenic bacteria during rhinovirus infection is associated with increased respiratory symptoms and asthma exacerbations. *J. Allergy Clin. Immunol.* **133**, 1301–1307, 1307.e1–3 (2014).

115. Ballinger, M. N. & Standiford, T. J. Postinfluenza Bacterial Pneumonia: Host Defenses Gone Awry. *J. Interferon Cytokine Res.* **30**, 643–652 (2010).

116. Chapter 9 - Respiratory System Diseases of Nonhuman Primates | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/B9780123813664000092?token=994BF4593C3E8CB97DCC7D6652837A2C51689ED0348A68499AC1131119AB532088AEEECB9C14D3E99CCC52DB51B01C1F&originRegion=eu-west-1&originCreation=20220905102732 doi:10.1016/B978-0-12-381366-4.00009-2.

117. Hennigs, J. K. *et al.* Characterization of Enterobacter cloacae Pneumonia: A Single-Center Retrospective Analysis. *Lung* **189**, 475–483 (2011).

118. Almand, E. A., Moore, M. D. & Jaykus, L.-A. Virus-Bacteria Interactions: An Emerging Topic in Human Infection. *Viruses* **9**, (2017).

119. Goka, E. A., Vallely, P. J., Mutton, K. J. & Klapper, P. E. Single and multiple respiratory virus infections and severity of respiratory disease: a systematic review. *Paediatr. Respir. Rev.* **15**, 363–370 (2014).

120. Guido, M. *et al.* Human bocavirus: Current knowledge and future challenges. *World J. Gastroenterol.* **22**, 8684–8697 (2016).

121. Ong, D. S. Y., Schuurman, R. & Heikens, E. Human bocavirus in stool: A true pathogen or an innocent bystander? *J. Clin. Virol.* **74**, 45–49 (2016).

122. Schildgen, O. & Schildgen, V. The Role of the Human Bocavirus (HBoV) in Respiratory Infections. in *Advanced Techniques in Diagnostic Microbiology: Volume 2: Applications* (eds. Tang, Y.-W. & Stratton, C. W.) 281–301 (Springer International Publishing, 2018). doi:10.1007/978-3-319-95111-9_12.

123. Agin, K., Rezaee, I., Heydarifard, Z. & Jazayeri, S.-M. Co-detection of bocavirus and bacteria in a respiratory specimen from a pregnant woman using multiplex real time PCR; a pathogenic role, or a bystander? *Iran. J. Microbiol.* **12**, 70–72 (2020).

124. Weiss, P. F. Pediatric Vasculitis. *Pediatr. Clin. North Am.* **59**, 407–423 (2012).

125. Rowley, A. H. & Shulman, S. T. The Epidemiology and Pathogenesis of Kawasaki Disease. *Front. Pediatr.* **6**, (2018).

126. McCrindle, B. W. *et al.* Diagnosis, Treatment, and Long-Term Management of Kawasaki Disease: A Scientific Statement for Health Professionals From the American Heart Association. *Circulation* **135**, (2017).

127. Yang, Y.-H. *et al.* The immunobiology of Henoch–Schönlein purpura. *Autoimmun. Rev.* **7**, 179–184 (2008).

128. Kainth, R. & Shah, P. Kawasaki disease: origins and evolution. *Arch. Dis. Child.* archdischild-2019-317070 (2019) doi:10.1136/archdischild-2019-317070.

129. Burns, J. C. & Glodé, M. P. Kawasaki syndrome. *The Lancet* **364**, 533–544 (2004).

130. Saulsbury, F. T. Clinical update: Henoch-Schönlein purpura. *The Lancet* **369**, 976–978 (2007).

131. Saulsbury, F. T. Henoch-Schönlein purpura. *Curr. Opin. Rheumatol.* **13**, 35–40 (2001).

132. Brown, T. J. *et al.* CD8 T Lymphocytes and Macrophages Infiltrate Coronary Artery Aneurysms in Acute Kawasaki Disease. *J. Infect. Dis.* **184**, 940–943 (2001).

133. Rowley, A. H. *et al.* The transcriptional profile of coronary arteritis in Kawasaki disease. *BMC Genomics* **16**, 1076 (2015).

134. Rowley, A. H., Baker, S. C., Orenstein, J. M. & Shulman, S. T. Searching for the cause of Kawasaki disease — cytoplasmic inclusion bodies provide new insight. *Nat. Rev. Microbiol.* **6**, 394–401 (2008).

135. Rivas, M. N. *et al.* Intestinal permeability and IgA provoke immune vasculitis linked to cardiovascular inflammation. *Immunity* **51**, 508-521.e6 (2019).

136. Bloom, J. L., Darst, J. R., Prok, L. & Soep, J. B. A case of Henoch-Schonlein Purpura with dilated coronary arteries. *Pediatr. Rheumatol. Online J.* **16**, 54 (2018).

137. Stewart, M., Savage, J. M., Bell, B. & McCord, B. Long term renal prognosis of Henoch-Schönlein purpura in an unselected childhood population. *Eur. J. Pediatr.* **147**, 113–115 (1988).

138. Koskimies, O., Mir, S., Rapola, J. & Vilska, J. Henoch-Schonlein nephritis: long-term prognosis of unselected patients. *Arch. Dis. Child.* **56**, 482–484 (1981).

139. Gordon, J. B., Kahn, A. M. & Burns, J. C. When children with Kawasaki disease grow up: Myocardial and vascular complications in adulthood. *J. Am. Coll. Cardiol.* **54**, 1911–1920 (2009).

140. Mukhopadhyay, S., Nair, S. & Ghosh, S. Pathogenesis in tuberculosis: transcriptomic approaches to unraveling virulence mechanisms and finding new drug targets. *FEMS Microbiol. Rev.* **36**, 463–485 (2012).

141. Anderson, S. T. *et al.* Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa. *N. Engl. J. Med.* **370**, 1712–1723 (2014).

142. Hoang, L. T. *et al.* Global gene expression profiling identifies new therapeutic targets in acute Kawasaki disease. *Genome Med.* **6**, 541 (2014).

143. Jaggi, P. *et al.* Whole blood transcriptional profiles as a prognostic tool in complete and incomplete Kawasaki Disease. *PLOS ONE* **13**, e0197858 (2018).

144. Jackson, H. *et al.* Kawasaki Disease Patient Stratification and Pathway Analysis Based on Host Transcriptomic and Proteomic Profiles. *Int. J. Mol. Sci.* **22**, 5655 (2021).

145. López, C. *et al.* Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. *Nat. Commun.* **10**, 1459 (2019).

146. Mullighan, C. G. *et al.* Next Generation Transcriptomic Resequencing Identifies Novel Genetic Alterations in High-Risk (HR) Childhood Acute Lymphoblastic Leukemia (ALL): A Report From the Children's Oncology Group (COG) HR ALL TARGET Project. *Blood* **114**, 704–704 (2009).

147. Lawrance, I. C., Fiocchi, C. & Chakravarti, S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. *Hum. Mol. Genet.* **10**, 445–456 (2001).

148. Mandel, M. & Achiron, A. Gene Expression Studies in Systemic Lupus Erythematosus. *Lupus* **15**, 451–456 (2006).

149. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinforma. Oxf. Engl.* **24**, 1547–1548 (2008).

150. 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).

151. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

152. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

153. Domaszewska, T. *et al.* Concordant and discordant gene expression patterns in mouse strains identify best-fit animal model for human tuberculosis. *Sci. Rep.* **7**, (2017).

154. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

155. Safran, M. *et al.* GeneCards Version 3: the human gene integrator. *Database J. Biol. Databases Curation* **2010**, baq020 (2010).

156. Weiner 3rd, J. & Domaszewska, T. *tmod: an R package for general and multivariate enrichment analysis*. https://peerj.com/preprints/2420v1 (2016) doi:10.7287/peerj.preprints.2420v1.

157. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinforma. Oxf. Engl.* **30**, 523–530 (2014).

158. Biezeveld, M. H. *et al.* Sustained activation of neutrophils in the course of Kawasaki disease: an association with matrix metalloproteinases. *Clin. Exp. Immunol.* **141**, 183–188 (2005).

159. Rocanin-Arjo, A. *et al.* A meta-analysis of genome-wide association studies identifies ORM1 as a novel gene controlling thrombin generation potential. *Blood* **123**, 777–785 (2014).

160. Yilmaz, D., Kavakli, K. & Ozkayin, N. The elevated markers of hypercoagulability in children with Henoch-Schönlein purpura. *Pediatr. Hematol. Oncol.* **22**, 41–48 (2005).

161. Brendel-Müller, K., Hahn, A., Schneppenheim, R. & Santer, R. Laboratory signs of activated coagulation are common in Henoch-Schönlein purpura. *Pediatr. Nephrol.* **16**, 1084–1088 (2001).

162. Guo, M. M.-H., Chang, L.-S., Huang, Y.-H., Wang, F.-S. & Kuo, H.-C. Epigenetic Regulation of Macrophage Marker Expression Profiles in Kawasaki Disease. *Front. Pediatr.* **8**, (2020).

163. Li, H. *et al.* Genetic polymorphisms in arginase I and II and childhood asthma and atopy. *J. Allergy Clin. Immunol.* **117**, 119–126 (2006).

164. Yang, C. *et al. Identifying hub genes associated with clinical characteristics in IgA nephropathy by WGCNA*. https://www.researchsquare.com/article/rs-37768/v1 (2020) doi:10.21203/rs.3.rs-37768/v1.

165. Potla, U. *et al.* Podocyte-specific RAP1GAP expression contributes to focal segmental glomerulosclerosis–associated glomerular injury. *J. Clin. Invest.* **124**, 1757–1769 (2014).

166. Watanabe Ryu *et al.* MMP (Matrix Metalloprotease)-9–Producing Monocytes Enable T Cells to Invade the Vessel Wall and Cause Vasculitis. *Circ. Res.* **123**, 700–715 (2018).

167. Takeshita, S. *et al.* Elevated serum levels of matrix metalloproteinase-9 (MMP-9) in Kawasaki disease. *Clin. Exp. Immunol.* **125**, 340–344 (2001).

168. Sun, Y. *et al.* MMP-9 and IL-6 are potential biomarkers for disease activity in Takayasu's arteritis. *Int. J. Cardiol.* **156**, 236–238 (2012).

169. Mattos, W. *et al.* Matrix Metalloproteinase-9 Expression in Asthma: Effect of Asthma Severity, Allergen Challenge, and Inhaled Corticosteroids. *Chest* **122**, 1543–1552 (2002).

170. Hu, H.-M., Du, H.-W., Cui, J.-W., Feng, D.-Q. & Du, Z.-D. New biomarkers of Kawasaki disease identified by urine proteomic analysis. *FEBS Open Bio* **9**, 265–275 (2019).

171. Banskar, S., Detzner, A. A., Hozo, I. & Gupta, D. The Pglyrp1-Regulated Microbiome Enhances Experimental Allergic Asthma. 14.

172. Pietzsch, J. & Hoppmann, S. Human S100A12: a novel key player in inflammation? *Amino Acids* **36**, 381–389 (2009).

173. Liu, W. *et al. Olfm4* deletion enhances defense against *Staphylococcus aureus* in chronic granulomatous disease. *J. Clin. Invest.* **123**, 3751–3755 (2013).

174. Liu, W. *et al.* Olfactomedin 4 down-regulates innate immunity against Helicobacter pylori infection. *Proc. Natl. Acad. Sci.* **107**, 11056–11061 (2010).

175. Mejias, A. *et al.* Whole Blood Gene Expression Profiles to Assess Pathogenesis and Disease Severity in Infants with Respiratory Syncytial Virus Infection. *PLOS Med.* **10**, e1001549 (2013).

176. Li, S., Zhu, A., Ren, K., Li, S. & Chen, L. DEFA1B inhibits ZIKV replication and retards cell cycle progression through interaction with ORC1. *Life Sci.* **263**, 118564 (2020).

177. Magistroni, R., D'Agati, V. D., Appel, G. B. & Kiryluk, K. New developments in the genetics, pathogenesis, and therapy of IgA nephropathy. *Kidney Int.* **88**, 974–989 (2015).

178. Saulnier, N. *et al.* Gene expression profiling of patients with latex and/or vegetable food allergy. 14.

179. Wen, T. *et al.* Carbonic Anhydrase IV Is Expressed on IL-5–Activated Murine Eosinophils. *J. Immunol.* **192**, 5481–5489 (2014).

180. Hand, L. E. *et al.* The circadian clock regulates inflammatory arthritis. *FASEB J.* **30**, 3759–3770 (2016).

181. Weintraub, Y. *et al.* Clock Gene Disruption Is an Initial Manifestation of Inflammatory Bowel Diseases. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **18**, 115-122.e1 (2020).

182. Yu, Y., Li, Y., Jiang, M. & Zhao, J. *SPRR2C, DEFB4A, WIF1, CRY2, and KRT19 are correlated with the development of atopic eczema*. https://www.researchsquare.com/article/rs-57989/v1 (2020) doi:10.21203/rs.3.rs-57989/v1.

183. He, H.-J., Bing, H. & Liu, G. TSR2 Induces laryngeal cancer cell apoptosis through inhibiting NF-κB signaling pathway. *The Laryngoscope* **128**, E130–E134 (2018).

184. Matsuo, Y. *et al.* The Protective Role of the Transmembrane Thioredoxin-Related Protein TMX in Inflammatory Liver Injury. *Antioxid. Redox Signal.* **18**, 1263–1272 (2012).

185. Novick, D. *et al.* Interleukin-18 Binding Protein: A Novel Modulator of the Th1 Cytokine Response. *Immunity* **10**, 127–136 (1999).

186. Pays, E. The function of apolipoproteins L (APOLs): relevance for kidney disease, neurotransmission disorders, cancer and viral infection. *FEBS J.* **288**, 360–381 (2021).

187. Skorecki, K. L. *et al.* A null variant in the apolipoprotein L3 gene is associated with non-diabetic nephropathy. *Nephrol. Dial. Transplant.* **33**, 323–330 (2018).

188. Bian, P. *et al.* RIPK3 Promotes JEV Replication in Neurons via Downregulation of IFI44L. *Front. Microbiol.* **11**, (2020).

189. Busse, D. C. *et al.* Interferon-Induced Protein 44 and Interferon-Induced Protein 44-Like Restrict Replication of Respiratory Syncytial Virus. *J. Virol.* **94**, (2020).

190. Riley, J. L. PD-1 signaling in primary T cells. *Immunol. Rev.* **229**, 114–125 (2009).

191. Tammaro, A. *et al.* TREM-1 and its potential ligands in non-infectious diseases: from biology to clinical perspectives. *Pharmacol. Ther.* **177**, 81–95 (2017).

192. Specific over-expression of deltex and a new Kelch-like protein in human germinal center B cells - ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0161589003000026?via%3Dihub.

193. Hattori, A. *et al.* Characterization of Long cDNA Clones from Human Adult Spleen. *DNA Res.* **7**, 357–366 (2000).

194. SCCRO3 (DCUN1D3) Antagonizes the Neddylation and Oncogenic Activity of SCCRO (DCUN1D1). https://www.jbc.org/content/289/50/34728.long.

195. Proteomic mapping of ER–PM junctions identifies STIMATE as a regulator of Ca 2+ influx | Nature Cell Biology. https://www.nature.com/articles/ncb3234.

196. Querec, T. D. *et al.* Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat. Immunol.* **10**, 116–125 (2009).

197. Diebolder, C. A. *et al.* Complement is activated by IgG hexamers assembled at the cell surface. *Science* **343**, 1260–1263 (2014).

198. Rissoan, M.-C. *et al.* Subtractive hybridization reveals the expression of immunoglobulin-like transcript 7, Eph-B1, granzyme B, and 3 novel transcripts in human plasmacytoid dendritic cells. *Blood* **100**, 3295–3303 (2002).

199. Chen, A.-C. *et al.* Association between allergic diseases and risks of HSP and HSP nephritis: a population-based study. *Pediatr. Res.* **79**, 559–564 (2016).

200. Brown, A. Henoch-Schönlein Purpura and Acute Nephritis Due to Food Allergy. *Glasg. Med. J.* **27**, 84–86 (1946).

201. BURKE, D. M. & JELLINEK, H. L. NEARLY FATAL CASE OF SCHOENLEIN-HENOCH SYNDROME FOLLOWING INSECT BITE. *AMA Am. J. Dis. Child.* **88**, 772–774 (1954).

202. Zhang, L. *et al.* Differential protein analysis of serum exosomes post-intravenous immunoglobulin therapy in patients with Kawasaki disease. *Cardiol. Young* **27**, 1786–1796 (2017).

203. Hara, T. *et al.* Kawasaki disease: a matter of innate immunity. *Clin. Exp. Immunol.* **186**, 134–143 (2016).

204. Huang, Y. & Kawasaki Disease Center, K. C. G. M. H. 4311Decreased CD177 expression and its association with intravenous immunoglobulin resistance in Kawasaki disease. *Eur. Heart J.* **40**, (2019).

205. Huang, Y.-H., Lo, M.-H., Cai, X.-Y., Liu, S.-F. & Kuo, H.-C. Increase expression of CD177 in Kawasaki disease. *Pediatr. Rheumatol. Online J.* **17**, (2019).

206. Ko, T.-M. *et al.* Genome-wide transcriptome analysis to further understand neutrophil activation and lncRNA transcript profiles in Kawasaki disease. *Sci. Rep.* **9**, 328 (2019).

207. Abe, J. Immunological Abnormalities and Use of Biomarkers and Cytokines to Predict the Severity of Kawasaki Disease. in *Kawasaki Disease: Current Understanding of the Mechanism and Evidence-Based Treatment* (eds. Saji, B. T., Newburger, J. W., Burns, J. C. & Takahashi, M.) 31–37 (Springer Japan, 2017). doi:10.1007/978-4-431-56039-5_5.

208. Deng, M. *et al.* Involvement of p53, p21, and Caspase-3 in Apoptosis of Coronary Artery Smooth Muscle Cells in a Kawasaki Vasculitis Mouse Model. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **26**, e922429-1-e922429-8 (2020).

209. Orenstein, J. M. *et al.* Three Linked Vasculopathic Processes Characterize Kawasaki Disease: A Light and Transmission Electron Microscopic Study. *PLOS ONE* **7**, e38998 (2012).

210. Kaufmann, S. H. E., Dorhoi, A., Hotchkiss, R. S. & Bartenschlager, R. Host-directed therapies for bacterial and viral infections. *Nat. Rev. Drug Discov.* **17**, 35–56 (2018).

211. Busse, D. C. *et al.* Interferon-Induced Protein 44 and Interferon-Induced Protein 44-Like Restrict Replication of Respiratory Syncytial Virus. *J. Virol.* **94**, e00297-20 (2020).

212. Ullah, H. *et al.* Antiviral Activity of Interferon Alpha-Inducible Protein 27 Against Hepatitis B Virus Gene Expression and Replication. *Front. Microbiol.* **12**, (2021).

213. Ramilo, O. & Mejias, A. Host transcriptomics for diagnosis of infectious diseases: one step closer to clinical application. *Eur. Respir. J.* **49**, (2017).

214. Ramilo, O. *et al.* Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **109**, 2066–2077 (2006).

215. Waddell, L. A. *et al.* ADGRE1 (EMR1, F4/80) Is a Rapidly-Evolving Gene Expressed in Mammalian Monocyte-Macrophages. *Front. Immunol.* **9**, (2018).

216. Pflanz, S. *et al.* IL-27, a Heterodimeric Cytokine Composed of EBI3 and p28 Protein, Induces Proliferation of Naive CD4+ T Cells. *Immunity* **16**, 779–790 (2002).

217. Pflanz, S. *et al.* WSX-1 and Glycoprotein 130 Constitute a Signal-Transducing Receptor for IL-27. *J. Immunol.* **172**, 2225–2231 (2004).

218. Wong, H. R. *et al.* Interleukin-27 is a novel candidate diagnostic biomarker for bacterial infection in critically ill children. *Crit. Care* **16**, R213 (2012).

219. Morita, Y., Masters, E. A., Schwarz, E. M. & Muthukrishnan, G. Interleukin-27 and Its Diverse Effects on Bacterial Infections. *Front. Immunol.* **12**, 678515 (2021).

220. Peters, V. A., Joesting, J. J. & Freund, G. G. IL-1 receptor 2 (IL-1R2) and its role in immune regulation. *Brain. Behav. Immun.* **32**, 1–8 (2013).

221. Lang, Y. *et al.* Interleukin-1 Receptor 2: A New Biomarker for Sepsis Diagnosis and Gram-Negative/Gram-Positive Bacterial Differentiation. *Shock* **47**, 119–124 (2017).

222. Yamashita, N. *et al.* Pathogenic mechanisms of influenza A(H1N1)pdm09 infection elucidated on gene expression profiling. *Pediatr. Int.* **55**, 572–577 (2013).

223. Ghosh, M. *et al.* Trappin-2/Elafin: a novel innate anti-human immunodeficiency virus-1 molecule of the human female reproductive tract. *Immunology* **129**, 207–219 (2010).

224. Becker, T., Gerke, V., Kube, E. & Weber, K. S100P, a novel Ca(2+)-binding protein from human placenta. cDNA cloning, recombinant protein expression and Ca2+ binding properties. *Eur. J. Biochem.* **207**, 541–547 (1992).

225. Arumugam, T., Simeone, D. M., Schmidt, A. M. & Logsdon, C. D. S100P Stimulates Cell Proliferation and Survival via Receptor for Activated Glycation End Products (RAGE) *. *J. Biol. Chem.* **279**, 5059–5065 (2004).

226. Hjertner, B. *et al.* Development of a 3-transcript host expression assay to differentiate between viral and bacterial infections in pigs. *PLOS ONE* **16**, e0256106 (2021).

227. Schoggins, J. W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011).

228. DeDiego, M. L., Martinez-Sobrido, L. & Topham, D. J. Novel Functions of IFI44L as a Feedback Regulator of Host Antiviral Responses. *J. Virol.* **93**, e01159-19 (2019).

229. Diamond, M. S. IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation. *Cytokine Growth Factor Rev.* **25**, 543–550 (2014).

230. Tang, B. M. *et al.* A novel immune biomarker *IFI27* discriminates between influenza and bacteria in patients with suspected respiratory infection. *Eur. Respir. J.* **49**, 1602098 (2017).

231. Hovnanian, A. *et al.* The human 2',5'-oligoadenylate synthetase locus is composed of three distinct genes clustered on chromosome 12q24.2 encoding the 100-, 69-, and 40-kDa forms. *Genomics* **52**, 267–277 (1998).

232. Zhou, S. *et al.* A Neanderthal OAS1 isoform protects individuals of European ancestry against COVID-19 susceptibility and severity. *Nat. Med.* **27**, 659–667 (2021).

233. Gizzi, A. S. *et al.* A naturally occurring antiviral ribonucleotide encoded by the human genome. *Nature* **558**, 610–614 (2018).

234. Honarmand Ebrahimi, K. A unifying view of the broad-spectrum antiviral activity of RSAD2 (viperin) based on its radical-SAM chemistry. *Metallomics* **10**, 539–552 (2018).

235. Kurokawa, C., Iankov, I. D. & Galanis, E. A key anti-viral protein, RSAD2/VIPERIN, restricts the release of measles virus from infected cells. *Virus Res.* **263**, 145–150 (2019).

236. Thiele, C. & Hirschfeld, G. cutpointr: Improved Estimation and Validation of Optimal Cutpoints in R. *J. Stat. Softw.* **98**, 1–27 (2021).

237. Meier, U. Nonparametric equivalence testing with respect to the median difference. *Pharm. Stat.* **9**, 142–150 (2010).

238. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses - Daniël Lakens, 2017. https://journals.sagepub.com/doi/10.1177/1948550617697177.

239. Steinberg, B. E., Goldenberg, N. M. & Lee, W. L. Do viral infections mimic bacterial sepsis? The role of microvascular permeability: A review of mechanisms and methods. *Antiviral Res.* **93**, 2–15 (2012).

240. Miller, C. S. & Danaher, R. J. Asymptomatic shedding of herpes simplex virus (HSV) in the oral cavity. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endodontology* **105**, 43–50 (2008).

241. Abbott, R. J. *et al.* Asymptomatic Primary Infection with Epstein-Barr Virus: Observations on Young Adult Cases. *J. Virol.* **91**, e00382-17 (2017).

242. Brealey, J. C. *et al.* Streptococcus pneumoniae colonization of the nasopharynx is associated with increased severity during respiratory syncytial virus infection in young children. *Respirol. Carlton Vic* **23**, 220–227 (2018).