

UNIVERSITY OF OXFORD

Transcriptomic analysis of biomarkers of respiratory syncytial virus disease

DPhil thesis
(Michaelmas term 2020)

Joseph McGinley

DPhil candidate reading in Paediatrics

Green Templeton College

Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom

Supervisors: Andrew Pollard, Daniel O'Connor, Simon Drysdale

Abstract

Transcriptomic analysis of biomarkers of respiratory syncytial virus disease

Joseph McGinley, Green Templeton College, D.Phil, Michaelmas Term 2020

RSV is one of the most prevalent causes of respiratory disease in children and the elderly worldwide. Despite this there is still a poor understanding of the role of host factors in severity of RSV disease. There is a significant unmet need to fully understand the relationship between host responses to viral pathogens and disease severity, both to enable more efficient stratification of patients and to assist in the development of novel treatments.

This thesis had three primary aims.

1: To characterise the host transcriptomic response against RSV.

2: To link the host transcriptomic response to RSV disease severity in infants.

3: To develop a biomarker capable of robustly identifying patients with more severe disease.

Microarrays were used to determine gene expression of RSV infected infants at the time of infection and in convalescence seven weeks later. Gene expression between these states were compared, associations between gene expression and severity were described using gene set enrichment analysis, and a panel of machine learning algorithms employed to identify the gene signature with the highest power to characterise severe RSV.

Key findings of this thesis include a description of the immune response against RSV, and the refinement of a novel gene signature that can correctly classify individuals into disease severity groups.

The immune response to RSV was characterised as being primarily defined by strong neutrophil and monocyte associated gene expression, alongside a downregulation of

lymphocytes and a strong upregulation of interferon related genes. Genes with expression patterns that correlate most strongly with RSV also relate to neutrophils and monocytes, with higher expression of interferon, lymphocyte, B and plasma cell genes relating most strongly to milder disease. A novel feature selection technique was developed to identify a five gene signature (made up of *ORF7*, *MMP8*, *IL1R2*, *ANXA3* and *BCL6*), that when used in conjunction with the age of a participant at the time of infection could correctly categorise severity of participants in this study at an AUC of .88, and in a completely independent validation cohort at an AUC of .6. The next challenge of this work is to be the refinement and validation of this biomarker, to assess the value that it can bring to physicians in decision making in the clinic.

Publications

Published

Lin G, **McGinley J**, Drysdale S, Pollard A. **Epidemiology and Immune Pathogenesis of Viral Sepsis**. *Front Immunol*. 2018;9:2147

Jefferies K, Drysdale S, Robinson H, Clutterbuck E, Blackwell L, **McGinley J**, Lin G, Galal U, Nair H, Aerssens J, Öner D, Langedijk A, Bont L, Wildenbeest J, Martinon-Torres F, Sánchez C, Nadel S, Openshaw P, Thwaites R, Widjojoatmodjo M, Zhang L, Nguyen T, Giaquinto C, Giordano G, Baraldi E, Pollard A, Respiratory Syncytial Virus Consortium in Europe (RESCEU) Investigators, **Presumed Risk Factors and Biomarkers for Severe Respiratory Syncytial Virus Disease and Related Sequelae: Protocol for an Observational Multicenter, Case-Control Study From the Respiratory Syncytial Virus Consortium in Europe (RESCEU)**, *The Journal of Infectious Diseases*

Submitted or in preparation

Green C, **McGinley J**, Sande C, Capone S, Vitelli A, Silva-Reyes L, Bibi S, Otasowie C, Thompson A, Dold C, Klenerman P, Dorrel L, Pollard A, O'Connor D. **An analysis of the host transcriptomic response to chimpanzee adenovirus- and MVA viral-vectored vaccines for**

RSV with correlation to clinical and immunological responses in healthy adults. *In preparation*

McGinley J, Thwaites R, Brebner W, Greenan-Barrett L, Aerssens J, Öner D, Bont L, Wildenbeest J, Martinón-Torres F, Nair H, Pollard A, Openshaw P, Drysdale S, RESCEU investigators. **A systematic review and meta-analysis of animal studies investigating the relationship between serum antibody, T lymphocytes and respiratory syncytial virus (RSV) disease**. *Article submitted*

McGinley J, McPherson C, Drysdale S, Oner D, Aerssens J, Thwaites R, Openshaw P, Pollard A. **Biomarkers of respiratory syncytial virus (RSV) disease severity in animal models**. *In preparation.*

The Impact of Covid-19 on this work

The presented work experienced some problems due to the Covid-19 pandemic. The largest impact of the pandemic on this project was in the number of participants with available microarray data. It was initially planned to run an additional set of samples gathered in the final year of the study. However, this underwent delays due to shutdowns. Additionally, the availability of further immunological data was hampered by the pandemic, meaning that it was not possible at the time of writing this thesis to examine the relationship between gene expression and antibodies, lymphocyte subtype populations, or any of the many other immunological outcomes for which data would have been generated.

Acknowledgements

I would like to thank Andrew Pollard, for giving me the opportunity to conduct this project. It is safe to say that without his advice and guidance this work would not have reached nearly the standard it has. Likewise I would like to thank Daniel O'Connor, without whom I fear my own skills as a scientist would not have risen to the task, and who took the time every week to listen to my progress, read my work and provide well thought out feedback. I would also like to thank Simon Drysdale, who tirelessly provided advice and feedback on every piece of work I did.

I am eternally grateful to everyone at OVG for their help and support over the last three years. Sagida Bibi for listening to all my frustrations, advising me prudently and for quickly and reliably correcting my mistakes in the lab. Liz Clutterbuck for her huge knowledge of all things lab and eternal willingness to have a coffee and a chat. Gu-lung Lin, who lightened my administrative workload immeasurably, and who was always there to step in if I need a break. Ruth Drury, who took it upon herself to teach me everything she knew about data science. Dylan Sheerin, who provided fantastic feedback and who was always a good friend. Amy Beveridge, who was always there for a chat and a pint, and who also lightened my workload immeasurably. Deniz, Jeroen and everyone else in the RESCEU consortium, for making the scope of this work achievable. Hannah, Matthew, Katja and everyone on the RESCEU team at OVG for always being there when I badgered them with emails. Chris, Christine, and all the pre-clin team for their huge help on my side projects.

I of course must thank Ruth Collins; without whose support I fear I would not have made it through the first year. As my one and only companion through the last 8 months of writing this thesis she was almost single-handedly responsible for maintaining my sanity, and ensuring I was able to finish this work. My siblings, Jaime, Alan, Conor, Dan and John, who were always free to talk when I needed it. Nat, for always taking the time to give me advice when I needed it. My parents, Kitty and Hughie who were always there to calm me down, and my grandfather, Joe, without whom I may never have decided to become a scientist.

Statement of responsibility

This thesis is based on the results of a large observational clinical trial. With the exception of the literature search underlying the results of chapter 6 and the running of the microarrays that provided the central results of this thesis, all work presented in this thesis was conducted by the author. The systematic literature searches were conducted by William Brebner and Lewis Greenan-Barrett. The microarrays were run by Janssen Pharmaceutica. This work would not have been possible without the extensive support of the industry and academic partners who collaborated on this trial. In particular I want to acknowledge the work of the clinical staff (in particular Hannah Robinson and Katja Pfafferot) who supported the trial, as well as the industry partners who ensured the generation of data (in particular Deniz Öner and Jeroen Aerssens) and the research staff who ensured the smooth running of everything from the point of sample collection, to the final assays. In particular I would like to thank Liz Clutterbuck, Amy Beveridge and everybody else at OVG for their essential contributions.

Contents

Abstract 2
Publications 3
Acknowledgements 5
Statement of responsibility 6
Index of figures 13
Index of tables 18
Chapter 1: Respiratory Syncytial Virus 19
1.1 RSV virology 19
1.2 Burden of RSV, seasonality, and epidemiology 22
1.3 RSV clinical presentation 25
1.4 Pathogenesis of RSV 27
1.5 Immune response against RSV 31
1.5.1 Humoral immunity 31
1.5.2 Cellular immunity 33
1.5.3 Innate immunity 41
1.5.4 Pathogen detection and antigen presentation 46
1.6 Transcriptomic response to RSV infection 51
1.7 Treatment of RSV 65
1.8 Biomarkers for RSV disease 68
1.9 Aims and objectives of this thesis 70
Chapter 2: General materials and methods 71
2. 1 Buffers 71
2.2 RESCEU Case control study 73
2.2.1 Participants 73
2.2.2 Power calculations 75
2.2.3 RESCEU infant Case control study objectives and outcome measures 76
2.2.4 Study participants 77
2.2.5 Inclusion and exclusion criteria 78
2.2.6 Recruitment sites 79
2.2.7 Recruitment 80
2.2.8 Screening and eligibility assessment 80
2.2.9 Informed Consent 81
2.2.10 Assessments and sampling 81

2.2.11 Patient testing with Alere I point of care test 84
2.2.12 Withdrawal from study 84
2.2.13 Reporting and storage of clinical information 85
2.3 Sample collection and processing 86
2.3.1 Whole blood collection and storage 86
2.3.2 RNA extraction from whole blood 86
2.4 Microarray preparation 88
2.4.1 cDNA preparation 88
2.4.2 Array Hybridisation 91
2.5 Assessment of disease severity 92
2.5.1 ReSVinet score 92
2.5.2 Requirement for respiratory support 93
2.5.3 Requirement for hospitalization 94
Chapter 3: Transcriptomic response to RSV infection in infants 95
3.1 Introduction 95
3.2 Methods 96
3.2.1 Demographic information for samples used in transcriptomic analysis 96
3.2.2 Removal of outliers 97
3.2.3 Gene filtering 97
3.2.4 Normalization 98
3.2.5 Identification of differentially expressed probes on infection 98
3.2.6 Gene set enrichment analysis with GSEA 98
3.2.7 Principal component analysis 99
3.2.8 CIBERSORTx deconvolution of transcriptomics data 99
3.2.9 ComBat batch correction 99
3.2.10 PVCA (Principal Variance Component Analysis) 100
3.3 Quality control of microarray data reveals signal associated with RSV infection 100
3.3.1 Low quality samples removed from analysis 100
3.3.2 The majority of variation between samples prior to quality control is attributable to technical variation 101
3.3.3 Normalization of array data reveals distinct signal of RSV infection 106
3.3.4 Batch correction with ComBat removes effects of microarray batches and cDNA yields 113
3.4 RSV infection is associated with neutrophil and interferon signalling 117
3.4.1 Infection with RSV induces significant differential expression of 1707 genes 117
3.4.2 Gene set enrichment analysis reveals neutrophil degranulation and decreased ribosomal activity to be associated with RSV infection 119

3.4.3 Increased populations of neutrophils, eosinophils and monocytes associate with acute RSV infection, while increased proliferation of B cells and CD8+ T cells occurs in convalescence.... 127
3.5 Discussion.... 129
3.5.1 Summary of main findings.... 129
3.5.2 Quality control of microarray analysis.... 129
3.5.3 Differential expression and gene set enrichment.... 130
3.5.4 Deconvolution of cell types in whole blood.... 131
3.5.5 General limitations.... 132
3.5.6 Conclusions.... 133
Chapter 4: Associations of changes in gene expression with RSV disease.... 134
4.1 Introduction.... 134
4.2 Methods.... 135
4.2.1 Disease severity assessment.... 135
4.2.2 Correlative analyses.... 135
4.2.3 Venn diagrams.... 135
4.2.4 PLS-DA.... 136
4.3 Distribution of RSV disease severity among participants.... 136
4.3.1 Proportions of participants were hospitalised and required supplemental oxygen.... 136
4.4 Correlations of gene expression with clinical disease.... 137
4.4.1 Neutrophil and Monocyte gene signatures correlate significantly with invasiveness of respiratory support.... 137
4.5 Patterns of gene expression are distinct between participants with severe and mild disease 144
4.5.1 Genes on first component of invasiveness of respiratory intervention PLS-DA associate with upregulated neutrophils and monocytes, and downregulated lymphocytes.... 144
4.5.2 Genes on second component of invasiveness of respiratory intervention PLS-DA associate with upregulated neutrophils and monocytes, and downregulated lymphocytes.... 150
4.6 Discussion.... 154
4.6.1 Summary of main findings.... 154
4.6.2 Correlations between gene expression and disease severity.... 154
4.6.3 Patterns of gene expression that have significant association with RSV disease severity 155
4.6.4 Limitations.... 157
4.6.5 Conclusions.... 158
Chapter 5: Biomarkers for severe RSV disease.... 160
5.1 Introduction.... 160
5.2 Methods.... 161
5.2.1 SIMON.... 161
5.2.2 Area under the receiver operating characteristic curve (AUC) analysis.... 162

5.2.3 PCA feature selection 162
5.2.4 Validation in published data 164
5.3 Preparation of gene set 165
5.4 Multiple algorithms are capable of predicting RSV disease severity using all selected genes 170
5.5 100 gene signature is predictive of more severe RSV disease 173
5.6 A five gene signature is predictive of requirement for invasive ventilation 176
5.7 Validation in published data 179
5.7 Discussion 181
5.7.1 Quality control 181
5.7.2 Models and validation 182
5.7.3 Genes 182
5.7.4 Model performance 186
5.7.5 Limitations 188
5.7.6 Conclusions 188
Chapter 6: Systematic review of the role of Antibodies and T lymphocytes in RSV disease severity in mice and cattle 189
6.1 Introduction 189
6.2 Methods 190
6.2.1 Search strategy 190
6.2.2 Data extraction 191
6.2.3 Standardised mean difference 192
6.2.4 Odds ratio calculation 192
6.2.5 Predictive analyses 192
6.3 Studies included 192
6.4 Antibody titres are not indicative of protection in cattle 193
6.5 Mouse antibody titres associate with less severe disease 198
6.6 Mouse lymphocytes associate with protection 200
6.6.1 Concurrent biomarker analyses 200
6.6.3 Predictive biomarker analyses 209
6.7 Discussion 212
6.7.1 Summary of main findings 212
6.7.2 Lymphocytic responses 212
6.7.3 Antibodies and RSV disease severity 213
6.7.4 Limitations 214
6.7.5 Conclusions 215
Chapter 7: Discussion 217

7.1 Introduction 217
7.2 Summary of the principal findings of this study 217
7.2.1 The immune response to RSV 217
7.2.2 RSV and severity 218
7.2.3 Biomarkers for severe RSV 220
7.3 Limitations of this study 221
7.4 Future directions 222
7.5 Conclusions 223
Appendix chapter 1: Supplemental materials for chapter 3 224
S1.1 Quality control of microarray data reveals signal associated with RSV infection 224
S1.1.1 The majority of variation between samples prior to quality control is attributable to technical variation 224
S1.1.2 Normalization of array data reveals distinct signal of RSV infection 227
S1.1.3 Batch correction with ComBat removes effects of microarray batches and cDNA yields. 229
S1.2 Differential expression of genes between participants with RSV infection and healthy controls 231
Appendix Chapter 2: Supplemental material for chapter 4 234
S2.1 Introduction 234
S2.2 Distributions of disease severity 234
S2.3 Correlations of gene expression with clinical disease 235
S2.3.1 Gene expression correlates significantly with clinical disease 235
S2.3.2 Majority of genes correlated with each severity metric are also correlated with others 236
S2.3.3 Neutrophil and Monocyte gene signatures correlate significantly with ReSVinet score 239
S2.3.4 Neutrophil and Monocyte gene signatures correlate significantly with requirement for supplemental Oxygen 246
S2.3.5 Neutrophil and Monocyte gene signature correlates significantly with requirement for hospitalisation 252
S2.4 Patterns of gene expression are distinct between participants with severe and mild disease 257
S2.4.1 Machine learning algorithms can robustly separate mild and severe disease as categorised by ReSVinet score 257
S2.4.2 Machine learning algorithms can robustly separate mild and severe disease as categorised by ReSVinet score 260
S2.4.3 First and second components of PLS-DA show distinct patterns of gene expression that correlate with ReSVinet score 263
S2.5 Discussion 272
Appendix chapter 3: Supplemental materials for chapter 5 274
3.1 Acute samples prior to normalization and batch correction 274

3.2 List of models used in biomarker discovery 277
References 278

Index of figures

Figure 1.1.1 Structure of the RSV genome **21**
Figure 1.1.2 Antigenic sites of the RSV F protein in the pre-fusion (left) and post fusion (right) conformations of the protein **22**
Figure 1.2 Components of the human immune response found to have a role in the immune response against RSV **50**
Figure 1.3 Components of the immune response found to have a role in the immune response against RSV in the mouse model **51**
Figure 1.7.1 Path diagram of all RSV vaccine candidates currently in clinical or pre-clinical trials. Error! Bookmark not defined.
Figure 3.3.1.1 Distribution of array probe intensities before and after removal of outliers with significantly different distribution of probe intensities across the whole array **100**
Figure 3.3.2.1 Boxplot of raw array probe intensities for each sample included in microarray analysis after removal of outliers and prior to any other QC **102**
Figure 3.3.2.2 Scree plot of loadings of principal components in raw data showing proportion of variance attributable to each principal component **103**
Figure 3.3.2.3 Principal component analysis of samples before quality control **105**
Figure 3.3.2.4 PVCA of raw data of all samples after removal of outliers **106**
Figure 3.3.4.1 PCA mapping of samples from all patients based on expression of sex chromosome genes on the first and second principal components **108**
Figure 3.3.4.2 Boxplot of normalized array probe intensities for each sample included in microarray analysis **109**
Figure 3.3.4.3 Scree plot of loadings of principal components in quantile normalized data showing proportion of variance attributable to each principal component **110**
Figure 3.3.4.4 PCA mappings of samples after RMA normalization, sex chromosome removal and probe filtration **112**
Figure 3.3.4.5 PVCA of quantile normalized data of all samples showing the proportion of variation in the data set **113**
Figure 3.3.5.1 Scree plot of loadings of principal components in quantile normalized and batch corrected data showing proportion of variance attributable to each principal component **114**
Fig 3.3.5.2 PCA mappings of samples after normalization and background correction **116**
Figure 3.3.5.3 PVCA of batch corrected data of all samples after removal of outliers showing the proportion of variation in the data set attributable to patient characteristics and technical effects **117**
Figure 3.4.1.1 Volcano plot illustration of transcriptional response to RSV infection for each gene when comparing RSV infected children to convalescent children **118**
Figure 3.4.2.1 Gene set enrichment analysis of A Positively regulated and B negatively regulated biological processes from the set of all genes in analysis **122**
Figure 3.4.2.2 Gene set enrichment analysis of A Positively regulated and B negatively regulated cellular components from the set of all genes in analysis **123**
Figure 3.4.2.3 Gene set enrichment analysis of Positively regulated hallmark gene expression terms from the set of all genes in analysis **124**
Figure 3.4.2.4 Gene set enrichment analysis of A Positively regulated and B negatively regulated immunological database terms from the set of all genes in analysis **125**
Figure 3.4.2.5 Gene set enrichment analysis of A Positively regulated and B negatively regulated molecular functions from the set of all genes in analysis **126**
Figure 4.3.1.2 Distribution of participants recruited to the RESCEU case control trial according to methods of oxygen supplementation required **137**
Figure 4.4.1.2 Gene set enrichment analysis of A – biological process terms correlated with more severe and B – biological process terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence **140**

Figure 4.4.1.3 Gene set enrichment analysis of A – cellular component terms correlated with more severe and B – cellular component terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 141
Figure 4.4.1.4 Gene set enrichment analysis of A – molecular function terms correlated with more severe and B – molecular function terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 142
Figure 4.4.1.5 Gene set enrichment analysis of A – immunological terms correlated with more severe and B – immunological terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 143
Figure 4.5.1.1 Gene set enrichment analysis of A – biological process terms correlated with more severe and B – biological process terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 146
Figure 4.5.1.2 Gene set enrichment analysis of A – cellular component terms correlated with more severe and B – cellular component terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 147
Figure 4.5.1.3 Gene set enrichment analysis of A – molecular function terms correlated with more severe and B – molecular function terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 148
Figure 4.5.1.4 Gene set enrichment analysis of A – immunological terms correlated with more severe and B – immunological terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 149
Figure 4.5.2.1 Gene set enrichment analysis of A - cellular components associated with more severe from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 151
Figure 4.5.2.2 Gene set enrichment analysis of A - molecular function terms correlating with more severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 152
Figure 4.5.2.3 Gene set enrichment analysis of A – immunological terms correlated with more severe and B – immunological terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 153
Figure 5.2.3.1 Flow chart of feature selection methods demonstrating A: methods by which genes are ranked for selection of relevant features in an unbiased manner and B: the use of this method in biomarker discovery. 164
Figure 5.3.1 Scree plot of the first ten principal components of the samples from RSV infected patients at the time of acute disease after normalization and batch correction 166
Figure 5.3.2 Mappings of samples at the time of acute disease on the first and second principal components 167
Figure 5.3.3 Mappings of samples at the time of acute disease on A the first and second and B the second and third principal components 168
Figure 5.3.4 Loadings of genes on A: the first principal component, B: the second principal component and C: the third principal component 169
Figure 5.4.1 AUROC statistics for predictions of disease severity in the reserved test fraction of the population for highest performing machine learning models where models were trained on top 2500 selected genes with age used as a cofactor 171
Figure 5.6.3 Expression of the six features in final signature for both mild and severe disease 179
Figure 5.7.2 Receiver operating characteristic curve of predicted disease severity based on random forest model of 5 genes and age in the independent validation cohort 181
Figure 6.3.1 Literature Search Flowchart, based on PRISMA methods for conducting systematic reviews for A: systematic literature search for contribution of lymphocyte and cellular responses and their

relevance to animal disease severity and B: systematic literature search for contribution of immunoglobulin responses and their influence on animal disease severity 193
Figure 6.4.2 Forest plots of disease severity outcome data from calf studies that used; A: a continuous outcome measure of disease severity, such as a multifactorial scoring system, or a measure with some other non-binary outcome such as temperature; B: a dichotomous outcome measure of disease severity, where cattle were defined as severe or mild based on the development of any observable form of respiratory disease 197
Figure 6.5.2 Forest plot of disease severity outcome versus neutralising antibody levels data from mouse studies that used a continuous severity outcome 200
Figure 6.6.1.3 – Forest plots presenting random-effects meta-analysis with standard mean difference of disease severity measures between high cell and low cell number groups in mice 209
Figure 6.6.2.1 - Scatter graphs showing the relationship between various T lymphocytes (x-axis) and disease severity measure 211
Figure S1.1.1.1 Principal component analysis of samples before quality control Grouping of samples according to expression of probes 226
Figure S1.1.3.1 PCA mappings of samples after RMA normalization, sex chromosome removal and probe filtration Grouping of samples according to expression of probes 229
Figure S1.3.1 PCA mappings of samples after normalization and background correction Grouping of samples according to expression of probes 231
sFigure 1.2.1 Volcano plot illustration of transcriptional response to RSV infection when comparing RSV infected children to healthy controls 232
Sfigure 1.2.2 Volcano plot illustration of transcriptional response to RSV infection when comparing children in convalescence to healthy controls 233
Figure S2.2.1 Distribution of disease severity scores in patients recruited to the RESCEU case control study 235
Figure S2.2.2 Number of patients recruited to the RESCEU case control study who were hospitalised and who were not hospitalised 235
Figure S2.3.2.1 Venn diagram of genes that significantly correlate with disease severity as measured by ReSVinet score, invasiveness of supplemental oxygen treatment requirement for supplemental oxygen and requirement for hospitalisation 237
Figure S2.3.2.2 Agreement plots of genes with significant correlation with disease severity by some metric 238
Figure S2.3.3.2 Gene set enrichment analysis of A – biological processes associated with more severe disease and B - biological processes associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 242
Figure S2.3.3.3 Gene set enrichment analysis of A – cellular components associated with more severe disease and B – cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 243
Figure S2.3.3.4 Gene set enrichment analysis of A – molecular functions associated with more severe disease and B – molecular functions associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 244
Figure S2.3.3.5 Gene set enrichment analysis of A – immunological terms associated with more severe disease and B – immunological terms associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 245
Figure S2.3.4.2 Gene set enrichment analysis of A – biological processes associated with more severe disease and B - biological processes associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 248
Figure S2.3.4.3 Gene set enrichment analysis of A – cellular components associated with more severe disease and B – cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 249

Figure S2.3.4.4 Gene set enrichment analysis of A – molecular functions associated with more severe disease and B – molecular functions associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 250
Figure S2.3.4.5 Gene set enrichment analysis of A – immunological tems associated with more severe disease and B – immunological terms associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 251
Figure S2.3.5.2 Gene set enrichment analysis of A – biological processes associated with more severe disease and B - biological processes associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 254
Figure S2.3.5.3 Gene set enrichment analysis of A – cellular components associated with more severe disease and B – cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 255
Figure S2.3.5.4 Gene set enrichment analysis of A – molecular functions associated with more severe disease and B – molecular functions associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 256
Figure S2.4.1.1 Individual mappings on the first two components of pls-da analysis of disease severity defined by RESViNET score 258
Figure S2.4.1.2 Variable mappings on the first two components of pls-da analysis defined by RESViNET score 259
Figure S2.4.1.3 The loadings of genes on A: the first component and B: the second component of PLS-DA plots defined by RESViNET score 259
Figure S2.4.2.1 Individual mappings on the first two components of pls-da analysis according to RESViNET score 261
Figure S2.4.2.2 Variable mappings on the first two components of pls-da analysis according to RESViNET score 262
Figure S2.4.2.3 The loadings of genes on A: the first component and B: the second component of PLS-DA plots according to RESViNET score 262
Figure S2.4.3.1.1 Gene set enrichment analysis of A – biological processes associated with more severe disease and B - biological processes associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 264
Figure S2.4.3.1.2 Gene set enrichment analysis of A – cellular components associated with more severe disease and B – cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 265
Figure S2.4.3.1.3 Gene set enrichment analysis of A – molecular functions associated with more severe disease and B – molecular functions associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 266
Figure S2.4.3.1.4 Gene set enrichment analysis of A – immunological terms associated with more severe disease and B – immunological terms associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 267
Figure S2.4.3.2.1 Gene set enrichment analysis of biological processes associated with more severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 269
Figure S2.4.3.2.2 Gene set enrichment analysis of cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence 270
Figure S2.4.3.2.3 Gene set enrichment analysis of A – immunological terms associated with more severe disease and B – immunological terms associated with less severe disease from the set of differentially

expressed genes between patients at the time of RSV infection and the same patients in convalescence 271
Figure S3.1.1 Scree plot of the first ten principal components of the samples from RSV infected patients at the time of acute disease before normalization and batch correction 275
Figure S3.1.2 Mappings of samples at the time of acute disease on the first and second principal components before normalization and batch correction 276

Index of tables

Table 1.6.1 Transcriptomic studies using samples taken from patients experiencing an ongoing RSV infection 55
Table 1.8.1 Summary of studies that produced a predictive transcriptomic biomarker 69
Table 2.1.1 Buffers and reagents used for RNA extractions 71
Table 2.1.2 Buffers and reagents used in synthesis of cDNA for microarray hybridisation 72
Table 2.1.3 Buffers and reagents used in microarray hybridisation 72
Table 2.2.3.1 Objectives and outcome measures of the RESCEU case control study 77
Table 2.2.5.2 Exclusion criteria for healthy controls in the RESCEU case control study 79
Table 2.2.10.1 Samples taken from patients recruited to the study 82
Table 2.2.10.2 Samples taken at convalescent time point from patients recruited to the study 83
Table 2.2.10.3 Information taken on study participants after completion of the study 83
Table 2.2.12.1 Potential reasons for withdrawal from the study 84
Table 2.2.13.1 Fields in study CRF and description of information recorded in each CRF field 86
Table 2.5.1.1 ReSVinet scoring system for RSV disease severity 93
Table 2.5.2.1 Scoring system of respiratory support methods used in treatment of RSV used for analysis 94
Table 3.2.1 Breakdown of demographic information for all patients included in microarray analysis 96
Table 3.2.2 Breakdown of participants recruited to the RESCEU case control study for each participating study site 97
Table 4.4.1.1 Table of the top 30 genes that were most positively and negatively correlated with graded requirements for supplemental oxygen 139
Table 5.4.2 Genes ranked by model contribution rank for the top 5 best performing models in the test set where models were trained on top 2463 selected genes 172
Table 5.5.2 Genes ranked by model contribution rank for the top 5 best performing models in the test set where models were trained on top 100 selected genes 175
Table 5.6.2 Genes ranked by model contribution rank for the top 5 best performing models in the test set where models were trained on top 5 selected genes with age as a covariate. 178
Table 5.7.1 Distribution of disease severities in population used for validation study 180
Table 6.4.1 A summary of calf studies, with details of observed correlations and study limitations 196
Table 6.5.1 A summary of mouse studies, with details of observed correlations and study limitations 199
***Table 6.6.1.1 Table of all studies used in analysis of relationship of lymphocytes with disease severity in this review* 207**
Table 6.6.1.2 - Summary of reported associations with disease severity and lymphocytes in papers included in the systematic review 208
Table S2.3.1.1 Summary table of genes that have a statistically significant correlation (FDR <0.05) with disease severity as measured by ReSVinet score, requirement for respiratory support, a scoring of invasiveness of respiratory support, and requirement for hospitalisation 236
Table S2.3.3.1 Table of the top 30 genes that were most highly associated with more and less severe disease according to the ReSVinet disease severity score 241
Table S2.3.4.1 Table of the top 30 genes that were most positively and negatively correlated with the requirement for supplemental oxygen of any kind 247
Table S2.3.5.1 Table of the top 30 genes that were most positively and negatively correlated with the requirement for hospitalisation of participants 253
Table S3.2.1 List of models used in biomarker discovery 277

Chapter 1: Respiratory Syncytial Virus

1.1 RSV virology

Respiratory Syncytial Virus (RSV) is a single stranded RNA virus of the genus *Orthopneumovirus*. The RSV virion is a nucleocapsid within a lipid envelope. The virions can be spherical, or long and filamentous. Spherical virions usually range between 100-350 nm in diameter, while filamentous virus particles can be 60-200 nm in diameter and up to 10 μm in length. The majority of viral particles produced in RSV infection are filamentous, with the conditions that result in spherical virions remaining unclear(1)(2).

The genome of RSV is approximately 15,000 bases long and contains 10 genes, with 11 open reading frames (3). These genes encode proteins responsible for viral attachment to the host (fusion protein F, small hydrophobic protein SH, glycoprotein G)(4)(5)(6), non-structural (NS) proteins 1 and 2 which are involved in immune evasion(7), polymerase and polymerase function proteins (polymerase L and phosphoprotein P)(8)(9), structural proteins (nucleoprotein N and matrix protein M)(10)(11) and regulatory proteins (anti-transcription protein M2-1 and transcription regulation protein M2-2)(12)(13). Two main subspecies of RSV are currently in yearly circulation, RSV-A and RSV-B(14). RSV-A and RSV-B differ primarily in sequence of the attachment glycoprotein G. RSV infection occurs first through attachment of the surface glycoprotein G to CX3CR1 receptors in ciliated airway epithelial cells(15), followed by fusion with the host cells which is mediated by the fusion protein F(15). The fusion protein of RSV has multiple functional confirmations which play different roles in infection. The F protein is first synthesised as an inactive precursor form called F_0, that requires cleavage before it is competent for fusion with host cells. After cleavage the F protein takes a functional form that facilitates fusion with the host cell(16). The RSV G gene is the most highly variable gene in the RSV genome(17). The G gene of RSV has translational variants that are expressed in different cell types when infected with RSV(18). Different functions outside of the traditional attachment function associated with the G protein have been observed for these alternative forms. The sG form of the G protein (which is translated from an alternative start codon, and is shorter than the standard G protein)(19) is secreted as opposed to being expressed on the surface of the virion, and is thought to act as a decoy

receptor for antibodies directed against the G protein, decreasing potential Fc mediated killing of RSV infected cells(20). Several genotypes of RSV have emerged over the years due to replication events within ectodomains of the G gene. Two notable emergent strains are the Ontario genotype (ON1)(21) (primarily identifiable by a 72 nucleotide long duplication near the carboxyl end of the mucin domain of the gene), which emerged in 2010, and the Buenos Aires genotype (BA)(22) (primarily identifiable by a 60 nucleotide long duplication near the carboxyl terminus of the gene) which emerged in 1999, and which has since become the predominant B genotype of RSV in circulation in some parts of the world(22).

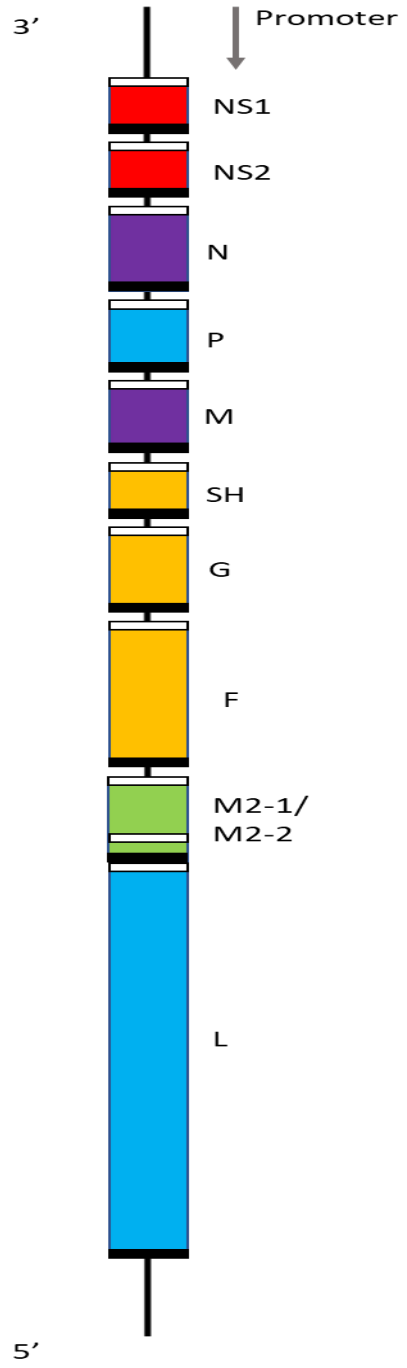


Figure 1.1.1 Structure of the RSV genome Genes of the RSV genome arrange from 3’ to 5’, black and white rectangles at the beginning and end of each gene represent conserved start and end signals, genes are coloured according to protein function. Red represents non-structural proteins involved in immune evasion, yellow represents surface proteins involved in attachment and fusion, purple represents structural proteins, blue represents proteins involved in viral genome replication, green represents regulatory proteins. M2 gene contains 2 open reading frames dividing the protein structure between M2-1 and the M2-2. Sizes of blocks are proportional to relative size of gene within the genome.

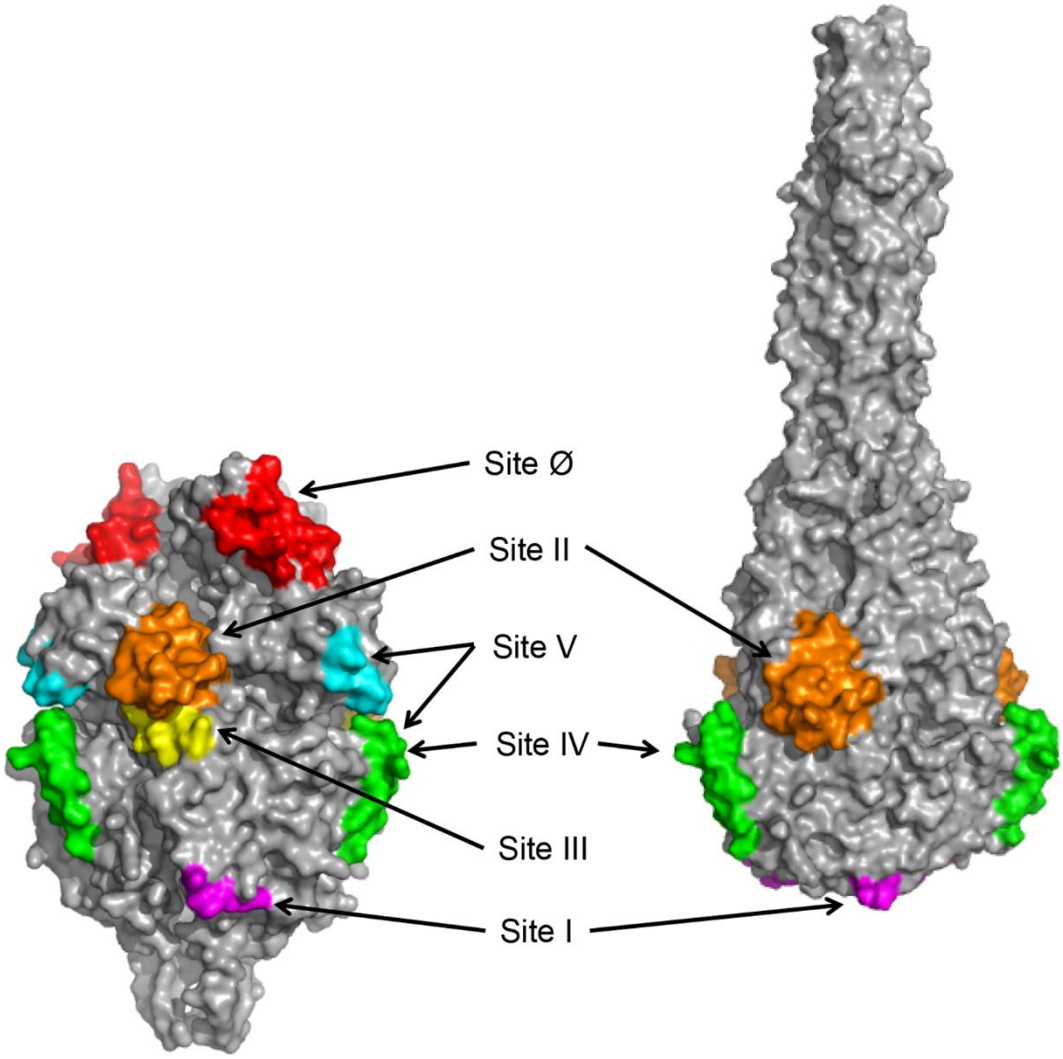


Figure 1.1.2 Antigenic sites of the RSV F protein in the pre-fusion (left) and post fusion (right) conformations of the protein. Site 0 and site V are hidden upon change in conformation to post fusion form. Figure taken with permission from Flynn et al (23).

1.2 Burden of RSV, seasonality, and epidemiology

RSV typically causes mild or no disease but is associated with severe acute respiratory infection in individuals at the extremes of age, and other at-risk groups. There is still little understanding of why some patients develop more severe disease while most of those infected outside of the highest risk groups remain asymptomatic(24). RSV is one of the largest causes of childhood mortality worldwide and is responsible for upwards of 33 million acute lower respiratory tract infections, 3.2 million hospitalisations and as many as 59600 deaths in children under five per year. It has been estimated that overall mortality due to RSV-ARTI (acute respiratory tract infection) may be as high as 118 200 deaths per year(25), with 99% of the deaths occurring in low- and middle-income countries (LMICs). RSV is

responsible for approximately 2.3% of all deaths of children within the first month of life, and 6.7% of all deaths between the ages of one month and one year (fig 1.4.1) (26). RSV has a distinct seasonality in temperate climates, with infections tending to occur in autumn and winter months each year, with one of the two main strains (RSV-A and RSV-B) taking precedence each year. The dominant circulating subtype of RSV in any given season varies from region to region, and dominant circulating strains have been seen observed to vary within a season even in regions with close geographical proximity to each other(27). RSV seasonality is less predictable in more tropical climates, with seasons tending to extend over longer periods, with less sharp and distinct peaks(28). Peaks occurring between May and June are common in tropical climates, although multiple peaks can occur throughout the year between spring and autumn.

Previously a number of risk factors for severe RSV infection in infants have been characterised(29), such as heart defects, premature birth and neuromuscular impairments. These risk factors contribute to risk for severe RSV in a number of ways related to dysregulation of cardiovascular development, which can result in increased risk of hypoxia. Congenital heart disease is a common risk factor contributing towards likelihood of severe RSV disease(30), and has been shown to do so by limiting oxygen delivery(31), resulting in a higher likelihood of hypoxia in cases where blood oxygenation is further inhibited by respiratory viral infection. Similarly, respiratory complications are common in people with neuromuscular disorders, with respiratory issues being one of the leading causes of death in people with neurological disorders(32). Premature birth has also been associated with the development of respiratory complications(33). It is possible that many risk factors for more severe RSV infection result in an increased tendency towards respiratory disorders, allowing the virus to further exacerbate existing respiratory complications.

RSV causes a significant healthcare burden worldwide, with roughly 450,000 GP visits, 29,160 hospitalisations and 80 deaths per year in the UK alone. This constitutes a higher disease burden than influenza in children below the age of one(34). RSV is not routinely tested for in much of the world, even in cases where a patient has developed severe bronchiolitis, meaning it is possible that RSV is underdiagnosed in cases of children and older adults presenting in the clinic with respiratory infections. Despite this difficulty dedicated clinical studies conducting tests for RSV in ALRI cases presenting in hospital have

allowed for some understanding of the incidence rate of RSV, and how it differs from region to region.

The age groups with the highest burden of RSV vary by study, and possibly by region. Most studies indicate that the highest burden of disease lies in the 2 to 6 months of age group(35)(36)(37), possibly due to a skew towards Th2 biased immunity in this age group(38), however there is some evidence suggesting a higher burden of disease in children between 1 and 2 years of age(39)(40). Another group with a significant burden of disease are older adults, who are responsible for approximately 1.5 million acute lower respiratory tract infections per year, 336,000 hospitalisations and 14,000 in hospital deaths, making up a significant portion of RSV deaths and hospitalisations worldwide.

Seroprevalence studies use the development of neutralising antibody for a disease in a population after the period in which the disease circulates can be used to determine the exposure rate of an infectious disease, including in those who do not go on to develop symptomatic disease. One such prospective study conducted in Rochester in the United States observed a 29% infection rate in infants for RSV within a given season, although in patients with a family member that experienced an RSV infection this rate was 62%(41). Other studies have observed similar attack rates for RSV in young children, with 32% being observed in the UK(38), 29% in Germany(42), while a rate of 48% was observed in children between the age of 4 months and 1 year in a Japanese study(43). Rates of exposure to RSV can vary quite highly from year to year and from population to population. Statistics from China show a higher exposure rate than other populations with 71% of children between the age of 1 month and 6 months showing seropositivity for RSV(44). Exposure rates vary significantly from year to year and are much higher in years with a larger epidemic. Extremely high exposure rates approaching 100% were observed in years with strong circulation, while in years with less severe epidemics exposure rates were as low as 40% in this population(37).

RSV is generally transmitted through the eyes and nose, during close contact via either droplets or by large particle aerosol(45). While initially thought incapable of being transmitted by small particle aerosols(46), more recent evidence has demonstrated RSV to be capable of small particle aerosol transmission, however this method is highly

inefficient(47). RSV also remains infectious on surfaces, suggesting that they may be another major means of disease spread.

1.3 RSV clinical presentation

RSV disease ranges from asymptomatic infection to fatal respiratory disease. Typically, less severe RSV infections are restricted to the upper respiratory tract, with common symptoms being wheezing, crackles, coughing, sneezing, congestion, rhinorrhoea, and occasionally fever and myalgia. The symptoms of RSV are most commonly characterised as bronchiolitis, which while lacking strict definitions, is most often characterised by rhinorrhoea, along with dry coughing, wheezing and tachypnea and dyspnea. Fever and Apnea may be observed in bronchiolitis, however high fevers are generally uncommon. More severe RSV infections usually involve more extensive colonisation and inflammation of the lower respiratory tract, potentially causing bronchiolitis, alongside more general symptoms such as increased breathing difficulty, apnea, tachypnea, high fevers and respiratory failure in extreme cases(48). Typically, RSV disease will begin in the upper respiratory tract and move to the lower respiratory tract over the course of the infection. More severe symptoms are much more common in younger children, with over 20% of infants younger than 6 months of age hospitalised with RSV experiencing apnea.

The development of RSV bronchiolitis usually begins with light cold like symptoms in the upper respiratory tract for 2-4 days, following which patients will either recover, or develop a lower respiratory tract infection. Development of lower respiratory tract symptoms is often signalled by the development of a more severe cough. Commonly used clinical indicators of requirement for hospitalisation include younger age (< 2 months), dehydration, feeding intolerance, tachycardia, low oxygen saturation and breathing difficulties(49).

In the case of development of bronchiolitis, a direct diagnosis of RSV is not common, as bronchiolitis may be attributed to a number of viral and bacterial pathogens. Bronchiolitis is often diagnosed based on common respiratory symptoms, along with other common symptoms such as feeding intolerance, lethargy, or irritability. Due to similarities between bronchiolitis and pneumonia X-rays are often used in diagnosis. Bacterial testing is not recommended in cases of bronchiolitis as bacterial infections are uncommon in bronchiolitis. Virological testing to identify specific viruses is not recommended in most

cases of bronchiolitis in the United States, as knowledge of the viral pathogen responsible for the condition does not often alter clinical management decisions(50). Diagnosis of RSV for the purpose of epidemiological studies, and in cases where it is desirable to know the aetiology of a viral pathogen is most often performed by using viral PCR, but previously fluorescence assays and viral culture have been used(51).

Infants have been recognised as a high-risk group for severe respiratory infections, including RSV. It has been estimated that up to 95% of children will have been infected with RSV before the age of 2(52). Children below the age of 1 are at the highest risk of hospitalisation with severe RSV. Other well characterised risk factors that predispose towards more severe disease include prematurity, congenital heart and lung diseases. Low birth weight has also been associated with a higher risk of RSV disease, particularly in patients with a birth weight below the 10^{th} percentile. Congenital and acquired immunodeficiencies both predispose towards more severe RSV disease. Many birth defects also contribute towards significantly increased risk for developing RSV LRTI. Down's Syndrome, along with other chromosomal abnormalities and conditions are associated with up to a twofold increase in hospitalisation risk, generally attributed to anatomic abnormalities in the respiratory tract that contribute towards susceptibility to respiratory disorders and increased risk of severe respiratory infection(53). In particular, narrowed airways are common in children with Down's Syndrome(54), which have been shown to contribute towards risk of respiratory disease(55). Cystic fibrosis shows the highest increase in risk of any risk factor and contributes towards a 4.3 fold increase in risk hospitalisation due to RSV LRTI. These underlying conditions in addition to increasing risk of hospitalisation also increase risk of developing particularly severe infections. Patients with these conditions are more likely to require use of the ICU, have an increased risk of mortality, are more likely to require mechanical ventilation, and experience longer periods of hospitalisation(51).

Despite the abundance of risk factors predisposing towards hospitalisation with RSV, more than half of all patients hospitalised with RSV do not have any of these, although many will have more common risk factors that do not contribute to likelihood of hospitalisation to the same extent, such as age of below 6 weeks, male sex, and exposure to other children who have developed the disease, such as siblings or other children in a day care(51).

In LMICs severe disease is most common in patients below the age of 6 months, with over 39% of hospitalisations with RSV occurring in this age group(56). RSV disease in infants is associated with the development of long term respiratory sequelae such asthma later in life, however the reasons for this association remain unclear, and it is not yet known whether RSV infection is predisposing children towards respiratory sequelae, or if children with a predisposition towards respiratory sequelae are more likely to develop severe RSV.

One of the reasons why infants are more susceptible to severe respiratory disease than older age groups is the relative narrowness of airways in younger children which makes them prone to obstruction(57). Additionally the smooth muscle of the respiratory tract is prone to hyperreactivity during airway inflammation, further predisposing children to symptoms associated with narrowed airways, such as coup, wheezing(55) and crackles(58).

Re-infection with RSV in the first two years of life is common, and occurs in between 30-75% of children who experienced an RSV infection within the first 12 months of life, typically in the following RSV season(37). Re-infection rates vary from season to season based on scale of the current epidemic. Generally subsequent infections after the first are less severe, with rates of bronchiolitis, fever and other symptoms decreasing upon re-exposure. Re-infection occurs frequently throughout the remainder of life, however the majority of infections in non-risk groups are subclinical or asymptomatic(37). Re-infection poses a serious health risk in certain high-risk groups however, such as the elderly or the immunosuppressed.

1.4 Pathogenesis of RSV

The pathogenesis of RSV disease remains poorly characterised, with conflicting evidence on the main causes of severe disease. While there is some evidence for direct viral cytopathy as a cause of or contributor towards more severe RSV(59), there is a larger body of evidence indicating that the host response is the primary cause of severe disease(60). Neutrophils have been associated with severe RSV disease in humans and are the predominant immune cell type found in the Bronchoalveolar Lavage Fluid (BALF) of children with more severe disease(61). Extensive neutrophil infiltration of lung tissue has also been observed in fatal cases of RSV LRTI(62).

The influence of total viral load on disease severity and disease pathogenesis is a topic that remains controversial. There is some evidence suggesting the total viral load during an RSV

infection has a significant impact on disease severity for the worse(63)(64)(65)(66), either due to heightened immune responses to compensate for a higher load of virus (and therefore more immunopathogenesis) or increased viral cytopathy due to a higher load of virus. There is also evidence suggesting that the total viral load does not have a significant impact on the severity of RSV disease(67)(68). A common issue with many studies that have attempted to understand the relationship between viral load and disease is that viral load readings are taken at a single time point during the course of disease(69). One study followed patients in PICU (paediatric intensive care unit) wards over the course of infection and observed the change in viral load dynamics with disease severity and time, and found that patients with a higher initial viral load and with a faster clearance speed showed decreased disease severity(69). It is possible that a higher initial viral load induces a more robust immune response that results in faster viral clearance and less severe disease. This hypothesis is corroborated by other studies which observed higher viral loads in patients that did not require hospitalisation alongside higher concentrations of antiviral cytokines in the respiratory tract in this population(70)(71). One of these studies observed that more severe disease was associated with lower viral load and decreased expression of type-I IFN, and increased expression of IL-17A, a neutrophil associated cytokine(72).

Overall it is difficult to draw a final conclusion as to the relationship between viral load and disease severity, as thus far there has been a great degree of heterogeneity in design of studies attempting to address the question, with differences in age, time of infection, sample collection methods and assays all creating significant variability(69). Larger and better controlled studies (in particular for comorbidities) comparing the effects of viral load between age groups and across the time period of RSV disease would be desirable to provide more authoritative answers to this question.

Whether or not viral load has a significant influence on disease pathogenesis and severity, it remains only one factor contributing to disease severity, and while viral cytopathy may contribute to disease pathogenesis, it is also clear that immunopathogenesis is a major factor in disease severity and pathogenesis in both humans and in animal models.

The first step in infection of host cells by RSV is the attachment of the virion to the host cell via the G protein, usually in differentiated airway epithelial cells(73). The G protein binds to glycosaminoglycans (GAGs), a type of disaccharide polymer that is common on the surface

of many mammalian cell types(73). The primary GAGs that RSV binds to are negatively charged regions of heparan sulfate(74). Other GAGs have some role in facilitating RSV infection, particularly those containing iduronic acid(75). There are conflicting reports on how highly expressed heparan sulfate domains are on the surface of airway epithelial cells, with evidence in A549 immortalised human cells(76) suggesting a lack of expression in the human airway epithelium, while studies in BEAS-2b cells(77) noted drastically reduced infection following cleavage of heparan domains by heparinase. These data cast doubts as to whether binding to GAGs by the G protein is relevant in human clinical infections. It is possible that these properties may be unique to the immortalised cell lines used in these experiments. Further research using either primary cells or expression data from primary cells would be useful to better understand these interactions. While fusion of RSV with host cells via the G protein dependent on the heparin-binding domain of the G protein (a string of positively charged amino acids found between the two mucin-like domains of the G protein) has been shown, RSV variants lacking this region of the G protein still demonstrate some heparinase-sensitive binding ability, suggesting that other regions of the G protein may have some binding capability against host cells independent of the heparin binding domain(78).

Another target for RSV attachment via the G protein is the CX3 chemokine receptor 1 (CX3CR1), the receptor for the chemokine fractalkine(79). There is a body of evidence that suggests some interaction between the RSV G protein and CX3CR1. Cells with higher expression of CX3CR1 showed a proportionately higher rate of infection with RSV, while blocking of binding domains on CX3CR1 reduced RSV infection rates(77)(80). It has also been demonstrated that blocking of the G protein CX3CR1 in the mouse model reduced ability of mice to be infected with RSV(80). RSV has also been shown to co-localise with CX3CR1 motifs on cell surfaces(81). Despite this evidence however, a direct interaction between the G protein and CX3CR1 has not yet been demonstrated.

While RSV interaction with the CX3CR1 protein has been shown not to influence attachment, this receptor has been shown to be involved in some other aspects of RSV disease, in particular immune evasion(82).

Despite the importance of G protein in infection, it is not a full requirement for colonisation in cell culture. RSV mutants lacking both the G and SH protein can infect cells in culture and

proliferate, while RSV mutants lacking the F protein cannot(83). While the primary role of the F protein is fusion with the host cell membrane after attachment, there is evidence to suggest that F has an additional role in attachment. The F protein has also been shown to bind to heparin and heparan sulfate in vitro. This interaction was shown to facilitate binding to and infection of cells in vitro(84), with virions that only express the F protein (and not G or SH) retaining half of their infectivity in immortalised cell culture. There are a number of other proteins that F has been hypothesised to bind to during attachment, however like the binding targets of the G protein, no direct biochemical or physical interaction between the F protein and these proteins has been demonstrated using purified proteins(85). The proteins that have been associated with attachment via the F protein are nucleolin(86) (which has been shown to have interactions with a number of other viruses), epidermal growth factor receptor (EGFR)(87) and intercellular adhesion molecule 1 (ICAM1)(88).

The primary role of the F protein is facilitating fusion of the virion with the hosts cell membrane. RSV fusion with the host membrane is now thought to be a two stage process, where the virion is initially taken in by micropinocytosis, where it is contained in Rab5-positive fluid filled macropinosomes. RSV fuses with the membrane of these macropinosomes(89). It was also demonstrated that after internalization into macropinosomes, the F protein underwent a further cleavage step, removing a part of the F1 region, and making the virus infectious(89). There is also evidence for direct fusion of RSV with the plasma membrane of host cells, wherein RSV infection of the host plasma membrane is dependent on its interaction with cholesterol-rich microdomains (lipid rafts) in the host cell plasma membrane. Docking with these rafts appears to allow the virion to begin fusion with the plasma membrane(90). It is possible that both of these models for fusion of RSV exist in parallel in vivo depending on the host cell being infected and the local conditions at the time of infection, however at this time the conditions that would cause the virion to favour one method over the other are not yet known.

Induction of fusion with host membranes remains a poorly understood process. Fusion of RSV with the host cell is dependent on the conversion of the F protein to its “post-fusion” state. Many varieties of fusion proteins in common viruses have some chemical trigger for conversion of the fusion protein into its active form, such as pH, direct binding to a receptor, or activation by some other viral glycoprotein(91). PH induced fusion is unlikely in the case

of RSV. The existence of a host molecule that specifically induces fusion of RSV with host cells has been proposed, as species specificity of RSV is dependent on the F2 subunit of the F protein in vitro(92) (chimeric human and bovine F proteins containing a human F1 and a bovine F2 region had the capability to infect bovine cells but not human cells) but such a molecule has yet to be identified or characterised. Induction by another surface glycoprotein is unlikely given that RSV retains its capacity to attach to and fuse with cells even in G and SH knockout conditions.

It has also been hypothesised that fusion is mediated by spontaneous conversion of the F protein to the post-fusion form. Under this model, attachment to the host cell by the G protein or F protein would bring the virion in contact with the host cell, wherein spontaneous conversion of the F protein to the post-fusion form would commence fusion with the membrane. The F protein has been observed to spontaneously convert to the post-fusion form of the protein(93), although the rate at which this occurs and the factors affecting it have not yet been well characterised. One factor found to influence conversion to the post-fusion form of the protein is temperature, which causes a slight increase in conversion rate(94).

1.5 Immune response against RSV

The host immune response to RSV is highly associated with severity of disease, although in paediatric patients other factors such as viral cytopathy of host epithelial cells may have a role. A comprehensive understanding of the host immune response to RSV is essential for further vaccination efforts, as well as the development of treatments and biomarkers for disease.

1.5.1 Humoral immunity

B cells produce specific antibody in response to infection, antibodies with neutralising capacity are thought to be one of the most important aspects of protective immunity against RSV(95)(96). B cell proliferation is induced by RSV infection. Populations of circulating B cells and B cell precursors increase in infants with RSV LRTI(97), and high frequencies of plasma cells have been observed in the lungs of children with fatal cases of RSV LRTI(97). One method by which RSV may induce differentiation and proliferation of B cells is through IFN-α, which has been implicated in B cell differentiation and proliferation,

and which has been found to accumulate in the bronchial epithelial tissues of RSV infected patients(98). Other markers of B cell proliferation and differentiation are found in higher concentrations in the respiratory tissues of RSV infected patients co-localising around RSV infected cells(99). B-cell-activating factor (BAFF) and a B cell stimulating ligand (APRIL) levels are increased in the respiratory epithelium of RSV infected children(99). These ligands can induce the proliferation and encourage survival of plasma cells, resulting in heightened production of antibodies(100)(101).

There is evidence that antibody production in response to RSV infection can occur by T cell independent routes(98) and is linked to expression of APRIL and BAFF. RSV neutralising antibodies of various isotypes, including IgG, IgA and IgM in the respiratory epithelium correlate with markers of B cell activity and activation, such as APRIL and BAFF, but correlate with lower quantities of T cell dependent cytokines, including IL-2, IL-10 and IL-4(98). It is possible that the expression of B cell stimulating factors underpinning this T cell independent antibody response are induced by RSV infection of epithelial cells, as *BAFF* mRNA levels are found upregulated in bronchial brushings(99), although the mechanism of this induction remains uncharacterised, the fact that the BAFF production in the upper and lower airways of children with a range of viral and bacterial infections suggests that this is a common aspect of the immune response in the airways(99).

Antibody neutralisation of RSV increases over the course of infection, with neutralising antibodies against the RSV F and G proteins steadily increasing over the course of infection(102). Neutralising antibody against RSV (in particular IgG and IgA titres) is one of the most effective forms of longer-term protection against infection, with higher titres of RSV neutralising antibody correlating with decreased risk of infection, decreased viral replication and a decreased risk of developing more severe disease(95)(96), although they are not wholly protective against infection. There is evidence from a controlled human infection model (CHIM) suggesting that nasal IgA may confer more protection against reinfection with RSV than serum neutralising antibody titres(103), which is supported by further data from paediatric infections suggesting that IgA may have a greater contribution to recovery than other isotypes(104). It has also been observed that during RSV convalescence IgG RSV-specific memory B cells were detectable, but IgA RSV-specific memory B cells were not, suggesting that these may be present in lower numbers(103). The

lack of detectable IgA RSV-specific memory B cells may be important in explaining why RSV can recurrently re-infect certain individuals, and further suggests an important role for IgA in controlling initial RSV infection.

Maternally derived antibodies are a source of protection against RSV that takes precedence very early in life. The highest levels of RSV IgG are generally found in children <1 month old(105). By 3 months this antibody will have waned(106). Likelihood and severity of early childhood infections correlates with the avidity of the maternal antibody population against RSV(107)(108). Avidity of maternal antibodies against RSV is lower in patients with symptomatic RSV disease than in age-matched controls(109), and it is likely that low avidity maternal antibodies are associated with an increased risk of developing disease. It is possible that a high titre of maternal IgG may hinder the development of more effective IgA immunity(110).

1.5.2 Cellular immunity

T helper type 1 (Th1) responses are critical in the defence against RSV infection(111), and have strong relationships with speed of viral clearance(112) and development of symptoms(113). One of the most prominent Th1 cytokines is IFN-γ(114). IFN-γ is involved in reducing disease severity(113), either in limiting a skew towards harmful Th2 biased immunity, or in limiting viral infection, resulting in a shorter course of disease(112), and less chances for damage due to harmful inflammatory responses or viral cytopathy. Lower levels of IFN-γ are seen in the supernatant of stimulated PBMCs(115) from patients experiencing more severe RSV disease. Lower levels of Th1 skewed $CD4^+$ lymphocytes are also observed in patients experiencing more severe RSV disease(115). Additionally, certain high risk groups, in particular children <6 months of age have lower serum levels of IFN-γ(116), although the effect of decreased IFN-γ in this group is hard to separate from the overall decreased capacity to mount a competent immune response, although the lower IFN-γ in this group may contribute towards the harmful Th2 bias that has been shown to occur in very young patients(117). In particular IFN-γ has shown correlations with requirement for ventilation in patients with bronchiolitis, with a particularly high correlation existing between low levels of circulating IFN-γ and the requirement for mechanical ventilation(118). IL-12 is another Th1 lymphocyte response related cytokine that has been demonstrated to induce expression of IFN-γ. Levels of circulating IL-12 have been found to be lower in

patients with severe RSV RTIs when compared with both controls and patients with mild RSV disease(119)(115).

IFN-γ responses also influences other symptoms of RSV, such as hypoxia and wheezing. A low ratio of IFN-γ to IL-10 has been associated with development of hypoxia and wheeze(120), while hypoxic patients have an overall lower expression of IFN-γ mRNA in PBMCs(121).

These relationships between IFN-γ and disease are also observable in the nasal mucosa and in the lungs, with levels of IFN-γ being increased during RSV infection in both tissues. A weaker mucosal IFN-γ response to RSV has also been associated with numerous measures of disease severity, including need for ventilation, hypoxia and wheezing among other measures of severity(113), further supporting a protective role for IFN-γ in infection. The build-up of IFN-γ over the course of disease is indicative of the induction of Th1 responses, which commence after some period of infection, and typically peak late in infection(122).

RSV has numerous means by which it can attenuate the host interferon response. RSV inhibits the type I interferon response by modulation of the suppressor of cytokine signalling (SOCS) proteins, SOCS1 and SOCS3 in a G protein dependent manner in bronchial epithelial cell lines(123). The inhibition of type I interferons by the viral proteins NS1 and NS2 also hints at a potential capacity for type I IFNs to control RSV infection that is being inhibited by viral activity(124).

One of the key functions of interferon activity in the lungs during RSV infection is the recruitment of inflammatory monocytes to the respiratory tract. Inflammatory monocytes are the second cell type to be recruited to the lungs after neutrophils(125). Recruitment of inflammatory monocytes is mediated by the chemokine CCL2, which is induced by interferon signalling(126). The role of these monocytes in RSV remains ambiguous. In mouse models they have been shown to have an anti-viral effect, and contribute towards viral clearance in the lower respiratory tract(126). In humans high levels of CCL2 in the nasopharyngeal wash and bronchioalveolar lavage have been associated with the development of more severe disease(127)(128), and in models of influenza infection, have been shown to contribute to lung pathology(129).

Evidence from the mouse model also suggests that interferon responses may be important for limiting the development of airway hyperreactivity (AHR) and respiratory sequelae. One study of IFN-γ(-/-) mice challenged with RSV found that viral load in the lungs was significantly higher than in wild-type mice, with increased AHR. In studies of IFN-γR(-/-) mice, IFN receptor deficiency resulted in aggravated pulmonary pathology, enhanced disease and reduced viral clearance(130)(131), leading to increased expression of IL-4, IL-5 and IL-13 and a shift towards a Th2-dominated response(130). It has therefore been suggested that IFN-γ is an important aspect of protection from AHR and lung pathology(132). Transcriptomics studies have produced conflicting evidence on the role of IFN-γ in RSV disease in mice, with some studies showing an association between *ifng* and more severe disease(133), while others have associated IFN-γ with more severe disease(134). However, van Schaik and colleagues demonstrated less severe obstructive airways disease during RSV infection in IFN-γ(-/-) mice compared with wild-type controls, and treatment with IFN-γ in IFN-γ(-/) mice resulted in greater AHR(135). The role of IFN-γ is thus ambiguous in RSV infection in the mouse model.

In murine model of RSV infection type I interferons are primarily produced by plasmacytoid dendritic cells (pDCs)(136). Mice depleted of pDCs treated with IFN-α exhibited decreased viral load, while still exhibiting an increased tendency towards AHR compared with controls, suggesting that IFN-α is not the sole effector molecule through which pDCs may have an effect on RSV infection. It has been shown that neonatal mice produce IFN-α in response to RSV infection(137) and that treatment of mice with IFN-α or adoptively transferred pDCs, from a mature animal capable of producing IFN-α, reduced viral load and Th2 based immunopathology on repeated infections with RSV. It is unclear if the same results would be observed upon IFN-α treatment of mice prior to the first infection. Other studies have shown a correlation between increased IFN-α expression and improved outcomes (in particular decreased airway inflammation and hypersensitivity). Mice re-infected with RSV and treated with Flt3 ligand, a molecule that causes expansion of pDC and myeloid DC populations, and strongly induces IFN-α production in these cells(138). Showed significantly reduced viral load when compared with controls(139). It has been hypothesized that this is due to the increased production of RSV specific IgA(140), although no other measures of

disease severity were taken. Altogether IFN-α in the mouse model seems to aid in prevention of re-infection and correlates with reduced viral load.

Other Th1 cytokines also have associations with disease severity. The cytokines IL-1α and IL-18 have been associated with the development of more severe bronchiolitis, and are expressed at higher levels in children experiencing more severe disease(141)(142). TNF-α expression is highest at the peak of disease and decreases throughout convalescence(143).

$CD8^+$ lymphocytes are cytotoxic lymphocytes that induce apoptosis in cells that are expressing certain markers induced by either abnormal regulation within the cell, or infection of the cell by an intracellular pathogen. $CD8^+$ lymphocytes produce a number of cytokines also produced by Th1 lymphocytes in response to viral infection, in particular IFN-γ(144). $CD8^+$ T cell populations also negatively correlate with markers of Th2 activity, such as IL-4(144). Circulating $CD8^+$ T cell counts increase over the course of RSV disease(145), and associate with faster recovery (145). Systemic RSV specific $CD8^+$ T cell counts are at their lowest at the height of both symptoms and viral load, and peak during convalescence, and tend to remain elevated at time of discharge from ICU(146). $CD8^+$ lymphocytes display an effector phenotype during acute RSV infection(147), and cytotoxic CD8 responses are found to be higher in infants with less severe RSV infection, possibly indicating a protective role(148). In human challenge models in adults, $CD8^+$ lymphocyte influx into the lungs has been associated with decreased viral loads, and pre-existing RSV-specific $CD8^+$ cell counts in the lungs at the time of infection associated with decreased viral loads on infection and less severe symptoms(149).

Evidence from animal models has been used to try to understand the relationships between these cell populations and disease severity in ways that are difficult to observe in humans due to the relative difficulty of conducting interventions in humans. In animal models, it has been suggested that T-cell function may be actively dysregulated(150) by RSV infection, which may result in a higher degree of airway inflammation(151).

$CD4^+$ cells were found to contribute to increased lung pathology in one study of mice infected with RSV(152). Further research has demonstrated that selective deletion of IL-4Rα, expressed on $CD4^+$ cells, prevented immunopathology on reinfection in mice which had been originally infected as neonates(153).

In neonatal cattle, the immune response in the lungs is characterised by $CD8^+$ T-cell infiltration, with these T-cells outnumbering $CD4^+$ cells by a ratio of 3:1 by day 10 of infection(154). Selective depletion of T-lymphocyte subsets in gnotobiotic calves using monoclonal antibodies resulted in $CD4^+$-depleted calves exhibiting poor neutralising antibody production, without a decrease in rates of viral clearance(155). While $CD8^+$ T-cell depletion did not have an effect on antibody production, time to viral clearance in $CD8^+$-depleted calves was greater. However, $CD4^+$ and $CD8^+$ depletion *in vivo* resulted in a reduction in weight loss during RSV challenge(156).

A similar study in mice found that control mice developed airway eosinophilia and AHR in response to RSV infection, while $CD8^+$-depleted mice did not(157). It was also found $CD8^+$ cells enhanced disease during RSV re-infection in a study of BALB/c mice(158). One study that used a prime boost vaccination strategy to induce a memory $CD8^+$ T cell response in mice found that this $CD8^+$ memory resulted in increased weight loss in mice, alongside increased mortality and immunopathology mediated by IFN-γ(159). Other studies have demonstrated similar findings, where $CD8^+$ T cells have been shown to be key drivers of immunopathology while being critical for timely viral clearance(160)(156).

$CD8^+$ T cell depletion has significantly also had effects on disease in bovine infections, with multiple studies showing increased pneumonic consolidation in $CD8^+$ depleted cattle alongside a delay in viral clearance(155)(161)(162). It is likely that $CD8^+$ T cells contribute towards recovery from disease in cattle(163). A study of the mouse transcriptomic response to RSV infection found that an age associated decrease in $CD8^+$ T cell populations was associated with more severe disease(133).

Memory $CD8^+$ T cells are a subset of $CD8^+$ T cells with a considerably longer lifespan that are involved in long term memory responses against previously detected pathogens. Higher numbers of memory CD8 cells in the airways associated with protection against RSV infection in human adults(149). In the mouse model adoptive transfer of RSV-specific memory CD8 T cells reduced disease severity upon viral challenge(164), although infection of mice with an existing population of RSV specific CD8 T cells has also been associated with more severe disease(159) as a result of early and excessive IFN-γ production, a phenomenon that has not been described in response to repeat infections with RSV in humans.

T helper type 2 (Th2) lymphocyte responses have been extensively linked to both more severe RSV disease(165)(115)(166) and the development of ongoing respiratory sequelae(167). Th2 responses are primarily involved defence against parasites(168) and in allergy(169) and are linked to the proliferation of eosinophils(170). Th2 responses contribute to a skew away from Th1 responses, which are optimal for controlling viral infections(171). Th2 lymphocytes do not seem to have many beneficial roles in controlling RSV infection. Many aspects of Th2 biased immunity have also been associated with more severe RSV disease, further suggesting that these responses are not appropriate for controlling RSV infection and are may be deleterious. Levels of Th2 associated cytokines have been found to be elevated both locally and systemically during RSV LRTIs(165)(171)(172).

The balance between a Th1 and a Th2 biased response against RSV is arguably more important in determining disease severity than the change in a given response type considered independently, as each response type negatively regulates the other, and Th1 responses are better at controlling the virus than Th2 responses(166). A useful proxy for this balance is the ratio of IL-4/IFN-γ(166). A high IL-4/IFN-γ ratio has been associated with more severe RSV bronchiolitis(166)(171)(173). Circulating Th1 cell counts have also found to be diminished in severe RSV infection, while circulating Th2 cell counts were not(174). There have been studies that have contradicted these findings however, and indicated that a high ratio of IFN-γ/IL-4 is associated with more severe disease(175). However, overall a large majority of studies support the hypothesis that a Th2-biased response is deleterious in responses to RSV infections.

Regulatory T cells (Tregs) are a subset of $CD4^+$ T lymphocytes with immunoregulatory properties. Tregs also have the effect of further biasing the immune response away from other lymphocyte types. Tregs have a protective role against the development of severe disease in RSV infection and help to control harmful inflammation that may be contributing to tissue damage, however evidence for this hypothesis is lacking in humans. Frequency of activated Treg populations is reduced in the peripheral blood of patients with RSV infection(176)(177). Decrease in expression of IL-33, a cytokine involved in Treg accumulation in mucosal tissues(178) was observed in the lung mucosa, suggesting that the decrease in circulating T cells is not an artefact of their recruitment to the lungs, and may

represent an overall decrease in Tregs(176). Treg cell populations in humans have not been associated with any change in disease severity in humans(179).

There is evidence in the mouse model indicating a protective role for Tregs in RSV infection. Treg populations are associated with recovery from disease in the mouse model(180)(181), and appear to be both involved in the recruitment of $CD8^+$ T cells to the lung tissue(182) (which are associated with viral clearance and recovery from disease) and modulation of $CD8^+$ T cell induced inflammation(182). This relationship between $CD8^+$ T cells and Tregs has not yet been observed in human disease, and so may be a peculiarity of murine disease. Interestingly, IL-10 dependent immunomodulatory activity can occur via other T lymphocyte subtypes in the mouse model. IL-10 production during RSV infection has been observed in IFN-γ producing $FOXP3^-$ T cells(183). A Treg mediated skew away from harmful Th2 mediated inflammation and eosinophilia has also been observed in the mouse model(184). The differences between the roles observed for Tregs in the mouse model and what has thus far been observed in human studies vary significantly. While this may be due to differences in how the mouse model experiences RSV disease, it is not yet possible to categorically say that Tregs do not have a protective role in humans. More studies will be required to better understand this relationship in humans.

There is evidence for a role for Th2 responses in mice contributing towards the development of airway hyperresponsiveness. IL-13 is a critical cytokine in the allergic response, which is also associated with Th2 responses and associated with reduced RSV disease severity in the murine model. One study demonstrated that overexpression of IL-13 resulted in decreased viral titres, protection against RSV-induced weight loss and diminished lung IFN-γ production(185). However, RSV infection can induce an IL-13 dependent change in airway function, contributing to the development of severe allergic asthmatic responses(186), while a further study implicated IL-13 in AHR during RSV infection(187). It has been suggested that experimental use of different viral strains may account for this discrepancy(185).

Th2 responses have also been implicated in the development of enhanced disease following vaccination with the formalin inactivated RSV vaccine. Increased expression of IL-4 in the lungs was observed in mouse models of enhanced RSV disease(188), and depletion of IL-4 and IL-10 resulted in decreased disease severity(189)(190). Other studies in the mouse

model supported this evidence, showing an upregulation of Th2 associated cytokines IL-5 and IL-13 in enhanced RSV disease, indicating a role for a Th2 skew in contributing towards the development of enhanced disease(191). One explanation of this phenomenon is that a Th2 skew induced by the vaccine was incapable of properly controlling infection (due to a simultaneous decrease in expression of Th1 associated cytokines(191)), resulting in uncontrolled viral proliferation, which further induced an inappropriate immune response, resulting in spiralling uncontrolled inflammation and Th2 dependent eosinophilia(192).

Th17 responses are linked to neutrophil responses and have been implicated in contributing to more severe RSV disease. T helper type 17 (Th17) cells and associated cytokines (IL-17 in particular) are heightened in RSV bronchiolitis(193). IL-17 has been associated with increased risk of requiring ventilation(194) and IL-17A concentrations correlate with neutrophil counts in ventilated patients(195). There is some evidence suggesting that Th17 cells may have a protective role against severe RSV infection(176). It has also been shown that despite associations between IL-17A and more severe disease, it may be associated with faster clearance of disease and recovery in milder cases(194). Overall, Th17 responses are associated with harmful neutrophil responses and are often themselves associated with harm and worsened disease.

IL-17 effects the activity and differentiation of different types of lymphocytes that may have significance in RSV disease. IL-17 increases expression of IL-13 in the mouse model(196), which encourages Th2 differentiation. IL-17 can also inhibit the activity of $CD8^+$ T lymphocytes(197), inhibiting the ability of the patient to clear RSV infection.

A similar role for Th17 responses has been observed in animal models. IL-6 is up-regulated after infection in mice and mediates resolution by inducing production of IL-27. This suppresses pathogenic immune responses in RSV infection, and accordingly, IL-6 may be associated with reduced disease severity(198). Both neutralisation and knockout of IL-17 in mice have been demonstrated to reduce mucous production, inflammation and viral load in RSV infection, as well as causing the upregulation of a range of inflammatory molecules including IFN-γ and granzyme B(197). While IL-17 has also been demonstrated to have another potential role in RSV pathogenesis in mice through the regulation of IL-27(199). Mice with downregulated IL-17 (as a result of STAT-1 knockout) demonstrated increased

mucus production and airway dysfunction, although the extent to which this can be attributed to IL-17 is difficult to discern(200).

γδ T cells are a subset of T lymphocytes that do not require antigen processing and peptide presentation for activity, and that are involved primarily in immune responses in epithelial barriers(201). γδ T cells appear to produce more IFN-γ in RSV infection than in other viral infections of the respiratory tract(202). Children with RSV bronchiolitis have altered populations of γδ T cells in the blood compared with healthy controls(145). γδ T cells have been shown to be recruited to the lung during RSV infection in the mouse model, and additionally appear to be involved in viral clearance and decreased disease severity in this model in an IL-17A dependent manner (203). Which is the opposite of what has been observed in other mouse models of viral diseases, such as West Nile virus(204) and herpes simplex virus(205) where IL-17A expression has been associated with increased disease severity. γδ T cells have not been observed to influence the progression or severity of RSV disease in cattle(161), although they do produce cytokines when exposed to bovine RSV in vitro, suggesting that the relationship between γδ T cells and bovine RSV infection may not yet be fully understood(206).

Another potential role for γδ T cells in RSV infection is in the development of asthma and respiratory sequelae after infection. γδ T cells have been implicated in the development of allergic reactions in the respiratory mucosa(207) and in asthma(208). γδ T cells from infants with severe RSV showed decreased capacity to produce IFN-γ(202), which has been associated with increased risk of developing recurrent wheeze and asthma in later life(202), which is one possible way by which these cells may influence the development of respiratory sequalae, although their role remains poorly characterised in humans.

1.5.3 Innate immunity

The most prominent aspect of the early immune response to RSV is increased neutrophil recruitment to the lungs, as well as increased neutrophil activity in the lungs. Neutrophils are the most highly recruited cell type to the lungs in the early stages of RSV infection, based on cell counts from bronchiolar lavage (BAL) samples taken from children with both severe RSV disease and milder RSV disease(61). Extensive neutrophil recruitment has been

observed in the lungs of children with fatal RSV infections(209)(62). The timing of neutrophil responses also coincides with the peak of RSV symptoms. Neutrophil counts are typically highest at the peak of both viral load and RSV symptoms(146). There are also indicators that neutrophils experience increased activation during RSV infection, with increased production of neutrophil elastase during infection, and increased expression of markers of neutrophil activity, such as CD11b, CD54 and CD18(210)(211)(212).

There is evidence for direct interaction of RSV with neutrophils, as RSV genomic RNA and mRNA are present within neutrophils isolated from the BAL and from peripheral blood(213) however it remains unclear whether many of these interactions are due to neutrophil phagocytosis of RSV particles, or some other interaction. RSV F, G and N genes have been shown to be expressed in BAL and peripheral blood neutrophils(214), suggesting replication within neutrophils, rather than phagocytosis resulting in successful viral killing.

It has been demonstrated in recent years that viruses can stimulate neutrophil extracellular traps (NETs)(215) in a process called NETosis. NETosis is a form of programmed cell death in neutrophils involving the release of a web of DNA, histones and antibacterial proteins(216). The antifungal and antibacterial roles of NETs have been well characterised. Recently, it had been discovered that viruses induce NETosis, however the mechanisms by which NETosis could inhibit viral replication or protect from infection are not yet well understood. NETs have bene shown to bind to human immunodeficiency virus-1 particles and restrict their movement(215), while also destroying some viral particles by expression of antiviral molecules on NETs, such as myeloperoxidase and α-defensin(215). Histone particles in NETs also have a number of antiviral functions. Histones H3 and H4 can cause aggregation of influenza, inhibiting its ability to infect host cells(217). Histone protein H1 can bind to noroviruses and prevent attachment to and infection of intestinal epithelial cells(218). No role has yet been shown for NETs in RSV infection, but it is possible, particularly in light of RSVs induction of neutrophil recruitment and activation.

Neutrophil activity in the lungs during RSV infection has been associated with the type 1 interferon (IFN) response(219). Neutrophil recruitment during RSV infection has been found to be dependent on MYD88/TRIF signalling(220), while activation of neutrophils is mediated by MAVS signalling, which induced production of type I interferon. Neutrophil recruitment in MYD88/TRIF$^{-/-}$ mice can be restored through chemoattractants. Similarly, activation of

neutrophils in MAVS$^{-/-}$ mice can be restored through type I IFN administration to the lungs of infected mice(219). MYD88 and TRIF signalling were not found to be associated with neutrophil activation. Additionally, mice with decreased neutrophil recruitment to the lungs due to deficient MYD88 and TRIF activity did not have decreased viral loads, suggesting that neutrophils do not have a role in antiviral activity(219).

The results of studies showing antiviral activity for neutrophils in humans are inconsistent with what has been observed in mice, where no such role has been characterised. It is possible that RSV has developed some immune evasion mechanism that results in decreased neutrophil activity, similar to those hypothesised in HIV-1(215), Epstein Barr virus(221) and human cytomegalovirus(222). It is also possible that this is a feature of the model. Mice are not natural hosts for RSV, and so murine NETs may not be effective against the human virus. This is less likely than an immune evasion mechanism, as most heretofore characterised antiviral functions of NETs involve non-specific electrochemical interactions, which are highly generalizable across viruses, such as the interaction of NETs with negatively charged surface envelopes of viruses like HIV-1(215) or host defence peptides like α-defensins and cathelicidins which are found in NETs and which have shown activity against RSV(223)(224).

Eosinophils are another variety of polymorphonuclear cell that are active during acute RSV infection. Eosinophils may have a protective role during acute RSV infection. There is an increased level of eosinophils expressing the myeloid activation marker CD11b during RSV infection, and expression of this correlates with decreased disease severity as measured by time of requirement of supplemental oxygen(225). Additionally, when compared with other respiratory viruses, RSV infected patients do not appear to have higher eosinophil numbers at the time of acute infection, but do show increased eosinophil numbers during recovery from infection(226). Little is known about the recruitment of eosinophils to the respiratory tract during infection, but measures of activity do show increases in eosinophil activity during bronchiolitis(227)(228)(229). Elevated concentrations of eosinophil cationic protein (ECP) in the respiratory tract associate with upregulation of markers of neutrophil recruitment (specifically CCL-3)(227), as well as total systemic numbers of neutrophils, indicating that either neutrophil and eosinophil activity co-regulate, or simply that neutrophil recruitment and eosinophil activity coincide. There is more evidence suggesting

that eosinophils recruitment and activity either contributes to or coincides with recovery. Increase in concentrations of molecules such as RANTES (eosinophil chemoattractant), eotaxin and ECP all coincide with a transition from severe disease to recovery, suggesting that eosinophil activity may play a role in disease resolution(226)(230).

However, eosinophils also have been observed to contribute to the development of ongoing airway hyperreactivity (AHR), and so the development of further respiratory sequelae after RSV infection(231). These eosinophil responses are highly associated with a strong bias towards Th2 type lymphocyte responses, which has been well characterised as harmful in the context of RSV infection. Evidence from the mouse model suggests that eosinophils do have some role in the development of disease pathogenesis. Transfer of eosinophils from RSV-infected mice into naïve mice resulted in airway hyper-responsiveness (AHR)(232).

NK cell recruitment follows neutrophil and monocyte recruitment. There is evidence to suggest higher populations of NK cells associate with less severe RSV infection. Low numbers of NK cells are found in the lungs of children with fatal cases of RSV disease, and lower blood NK cell counts associate with more severe RSV disease(233)(234). Broadly, RSV infection seems to cause a reduction in total NK cell counts systemically, although there is an overall increase in activated NK cells lacking expression of CD94(145). NK cells may have some immunoregulatory role in RSV infection that contributes to their protective role against severe RSV infection. Blood NK cells in RSV infection show increased expression of inhibitory leukocyte immunoglobulin-like receptor subfamily B member (LILRB1)(235), which has well characterised immunoregulatory roles(236).

The earliest elements of the host immune system that have an effect on RSV infection are the non-specific innate elements of the nasal epithelium. These include the anti-microbial peptides and surfactant proteins contained within the mucus(224), which have been shown to have some ability to kill RSV. There is also some evidence that the composition of the airway microbiota has some effect on RSV infection, in particular it has been suggested that nasal microbiota can have some influence on whether infection progresses to the lower respiratory tract(237). It has also been suggested that the nasal microbiome influences the immune response against RSV, however it is not yet possible to determine whether the initial microbiome populations are influencing the development of RSV infection, or if RSV is causing dysregulation of the nasal microbiome in some patients. Additionally, it is possible

that changes in nasal microbiome post RSV infection (in particular the massive proliferation of certain bacterial clades, such as *Streptococcus pneumoniae*) that highly correlate with more severe infection are contributing towards the development of neutrophil responses that contribute to severity(238)(239). There is not yet enough evidence to determine whether this is simply an incidental side effect of RSV infection, or whether it is a central part of the pathogenesis of RSV disease.

The strong interferon response is one of the defining and identifying features of RSV infection(60). Interferon proteins in humans are a family of cytokines that are categorised as type I, type II or type III, with classes being defined by the set of receptors bound by the proteins. Interferons are broadly categorised as having antiviral activity, although they have been implicated in other processes. The most prominent type I interferon proteins are IFN-α and IFN-β, which are produced by and responded to by a large number of cell types(240). The most prominent type II interferon is IFN-γ, which is primarily produced by T lymphocytes and NK cells, and primarily causes signalling related to defence against intracellular pathogens. Type I IFN and type III IFN signalling is induced in response to viral infection by signalling by plasmacytoid dendritic cells(241).

RSV has demonstrated relatively weak induction of type I interferon activity. Trials have found extremely low or undetectable amounts of type I interferons in nasal secretions of children infected with RSV, and have not detected any significant influence for these proteins on disease severity or resolution of disease(242)(243). Expression of type I interferons was also found to have no influence on amount of viral shedding in RSV infected infants(244). IFN-α expression in response to RSV is higher in adult patients and in older children, who have a lower risk of developing severe RSV disease(245). An increased impairment of the capacity for the production of type I IFNs was seen in young children when compared with adults, and this increase was associated with disease severity in the infants, however infants with more severe disease in this study also showed decreased RIG-I and MAVS signalling, which could also be responsible for the differences in severity(245). Trials that have aimed to treat RSV by supplemental type I IFN application have found some success in limiting viral shedding and controlling infection(246), but have also induced more severe disease and exacerbated some symptoms(247).

Contrary to the relatively weak induction of type I interferons by RSV, type II interferons are strongly induced by the virus(248), and have well characterised roles in infection, and in the response to intracellular pathogens. IFN-γ has a strongly protective role against more severe RSV infection, although it has been implicated in increased disease severity. IFN-γ also strongly contributes towards increasing the bias towards a Th1 lymphocyte response, and indeed to the development of a cell mediated immunity in general, which is essential in controlling RSV infection and in developing ongoing immunity against the virus.

Innate lymphoid cells (ILCs) have been associated with RSV disease severity in both mouse models and in humans. There are relatively few data on the relationship between ILCs and RSV in humans. ILC2 has been associated with greater disease severity in humans(249), and have been associated with increased risk of being admitted to the ICU. One study found that mice infected with RSV demonstrated elevated IL-13 producing ILC2 cells at 4 days post infection(250). A mutation causing a loss of IL-13 producing ILC2 responses was also characterised. Decreased IL-13 levels in lungs observed in these mutants correlated with attenuated weight loss and decreased mucus production in infected mice. ILC2 responses have also been associated with increased asthma exacerbation in the RSV mouse model(251).

Complement C3a and tachykinins have been shown to have a role in RSV pathogenesis, as well as in the development of AHR post infection(252). Knockout of tachykinin receptors simultaneously reduced the development of AHR while also causing a decrease in the expression of Th17 cytokines(252) suggesting type-17 cytokines as a possible medium for the reduction in AHR development. Depletion of neutrophils by anti-Ly6G neutralising antibody was found to induce significantly higher weight loss, lung inflammation and AHR and delayed viral clearance, alongside a significant increase in the expression of type-17 cytokines, including IL-17, IL-22 and IL-23p19(253).

1.5.4 Pathogen detection and antigen presentation

The initial immune response against RSV is induced by binding of RSV to host pattern recognition receptors (PRRs) in the respiratory epithelium. The receptors most likely to first encounter the virus and induce initial signalling are RIG-I-like receptors expressed on the interior of alveolar macrophages(254) and in the epithelial cells of the respiratory tract(255). Signalling by RIG-I-like receptors is one of the primary methods by which viral infections are

detected by the host immune system, and are essential for induction of type I interferon signalling in RSV infection(256). Alveolar macrophages that are incapable of inducing type I interferon signalling in vitro still show some ability to control RSV infection(254), indicating some other role in viral control in addition to activation of anti-viral responses by type I interferon signalling.

Other PRRs have been implicated in the response to RSV, in particular the toll like receptor (TLR) family. TLRs 2, 3, 4 and 7 have been shown to recognise RSV(256). TLR3 senses double-stranded RNA in the endosomes, and so is capable of detecting RSV during replication and induces signalling via type I interferons(257). Induction of TLR3 in RSV infection takes place later than induction of other PRRs such as RIG-I and MDA5, and is regulated by interferon-β produced in response to RIG-I-like receptor signalling(257).

Interactions between the RSV F protein and TLR 4 have been observed in mouse models. These interactions are unexpected, due to the well characterised primary role of TLR4 in recognising bacterial lipopolysaccharide. Binding of RSV F to TLR4 was found to be necessary for the production of IL-6 in murine epithelial cells, and mice without TLR4 receptors took longer to clear RSV from the lungs(258). The involvement of TLR4 in inflammation has been challenged by other studies which have demonstrated a role for TLR4 in resolving and limiting RSV induced inflammation through induction of alternatively activated macrophages which limit harmful inflammation caused by other aspects of the immune response(259).

Alveolar macrophages are another antigen presenting cell that are involved in the immune response against RSV. Alveolar macrophages are one of the main inducers of the immune response against RSV in the respiratory epithelium, and are involved in early engagement of the inflammatory response by expression of cytokines such as TNF-α, IFN-γ and IL-6(260). In the mouse model macrophages have been shown to be essential for viral clearance(126), while in humans they have been associated with recovery from disease and improved time to viral clearance(112). Additionally alveolar macrophages in mice mediate the recruitment of monocytes to the lung via expression of type-I interferons(126).

Alternatively activated macrophages (M2 macrophages) are a subset of macrophages that are activated by cytokines associated with type-2 lymphocyte responses (specifically by

cytokines such as IL-4 and IL-13)(261) instead of the classical IFN-γ mediated activation pathway(262) and have distinct physiological roles from classically activated macrophages (M1), being more involved in repair, humoral immunity and the development of hypersensitivity and allergy, with subtypes of alternatively activated macrophages demonstrating immunoregulatory activity(263). Macrophages are polarized towards an M2 phenotype in RSV infection, and this polarization correlates with repair and recovery in the lungs(259).

The other main subset of macrophages (M1 macrophages or classical macrophages) are activated via IFN-γ, and have a pro-inflammatory phenotype(264). One study attempted to encourage the differentiation of M1 type macrophages in the mouse model by intranasal application of *Bacillus subtilis* found that the increased populations of M1 macrophages and associated inflammatory cytokines such as IFN-γ correlated with a reduction in viral titres and lung pathology(265).

There is evidence indicating that IL-6 may also contribute to the polarisation of alternatively activated macrophages(266), perhaps indicating that the role of TLR4 signalling in RSV infection overall may tend toward immunoregulatory and anti-inflammatory. There is conflicting evidence for a pro-inflammatory role for TLR4 signalling in RSV infection, with one study showing induction of NF-κB signalling by interaction between TLR4 and the RSV F protein(267), while another study has indicated the opposite, demonstrating no activation of NF-κB due to interaction between TLR4 and the RSV F protein(268). One difference between these two studies that may explain the discrepancy between the results, is that the study that indicated a pro-inflammatory function for TLR4 in RSV used a purified F protein, while the contradictory study used the full virus. It is possible that this interaction only happens between TLR4 and the purified F protein and does not happen in interactions with the full virus, suggesting that the induction of NF-κB via the F protein may not be physiologically relevant. The pro-inflammatory functions of TLR4 signalling in response to bacteria are well characterised(269), meaning a pro-inflammatory function for TLR4 in RSV is not impossible, however further study will be required to fully understand it's role in infection. The role of TLR4 in RSV infection remains ambiguous, but current evidence suggests a greater tendency towards immunoregulatory function than towards inflammatory.

There is evidence in mouse models for the induction of an adaptive immune response upon RSV infection even in the absence of TLRs and RLRs. It is possible that this occurs instead via induction of antigen uptake and recognition by DAMP (damage associated molecular pattern) associated signalling, as opposed to PAMP (pathogen associated molecular pattern) associated signalling, where RSV causes a degree of viral cytopathy and according mucus build-up in the airway, which would contain immunogenic PAMPs that would induce responses against RSV. The DAMPs HMGB1 and S100A9 have been shown to be released on RSV infection, lending credibility to this hypothesis(270)(271).

Interferon signalling in RSV is initiated via intermediate signalling molecules MAVS, MyD88 and TRIF, which respond directly to activation by PRRs. MAVS is signalled to by RIG-I-like receptors while MyD88 and TRIF are signalled to by TLRs(272).

Dendritic cells are recruited to the respiratory tract during infection, and present RSV antigen to T and B cells, allowing for the development of more specific responses to infection. Low numbers of dendritic cells are associated with the development of RSV bronchiolitis in humans(60)

Dendritic cells are also important in the response against RSV in animal models. MyD88, a cytosolic adapter downstream of toll-like receptors (TLRs) which enables signalling to activate transcription factors, including NF-kB and IFN regulatory factors (220). MyD88(-/-) mice exhibit exacerbated RSV pathophysiology, associated with enhanced Type-2 cytokine expression(273). Two further murine studies found DCs were involved in viral clearance and were protective against disease pathogenesis(136,274).

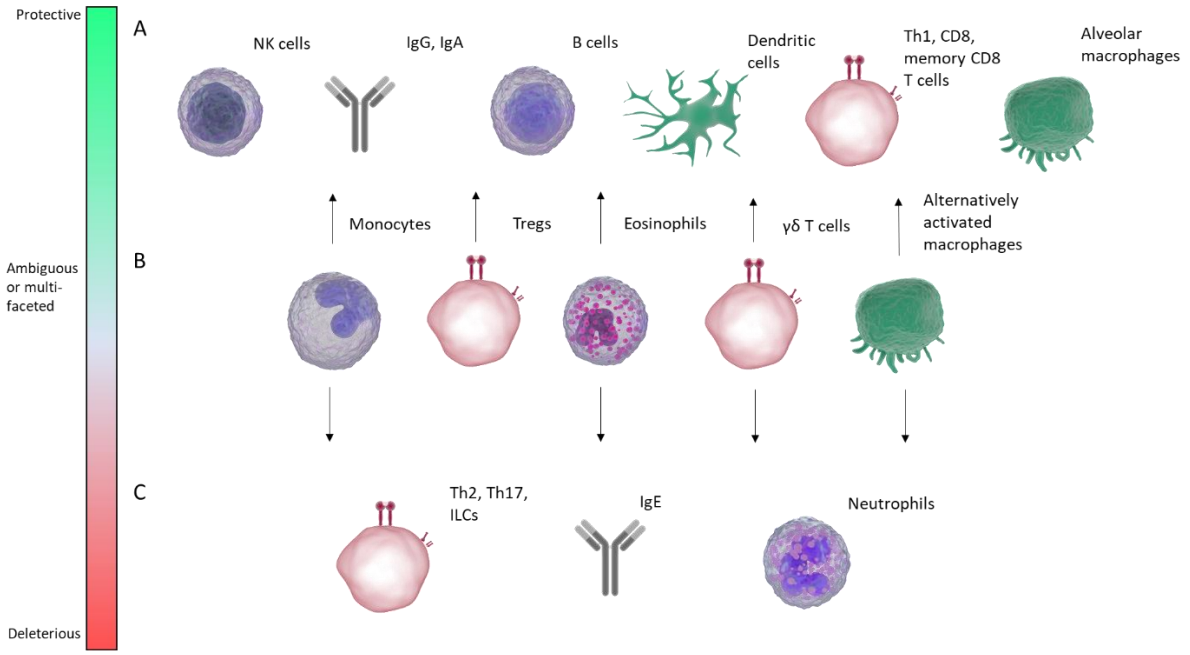


Figure 1.2 Components of the human immune response found to have a role in the immune response against RSV, colour gradient and arrangement indicate their potential role in pathogenesis. (A) Cell types unambiguously protective against severe disease, indicated by green background such as B cells, Dendritic cells and neutralising IgG and IgA, (B) ambiguous or multi-faceted roles in disease pathogenesis, indicated by the blue area on the gradient, arrows indicate roles for these cells in protection or pathogenesis, with arrows pointing upwards indicating a possible protective role and arrows pointing downwards indicating a possible deleterious role (C) cells with a primarily harmful role in disease pathogenesis, indicated by placement at the bottom of the figure and red area on the gradient

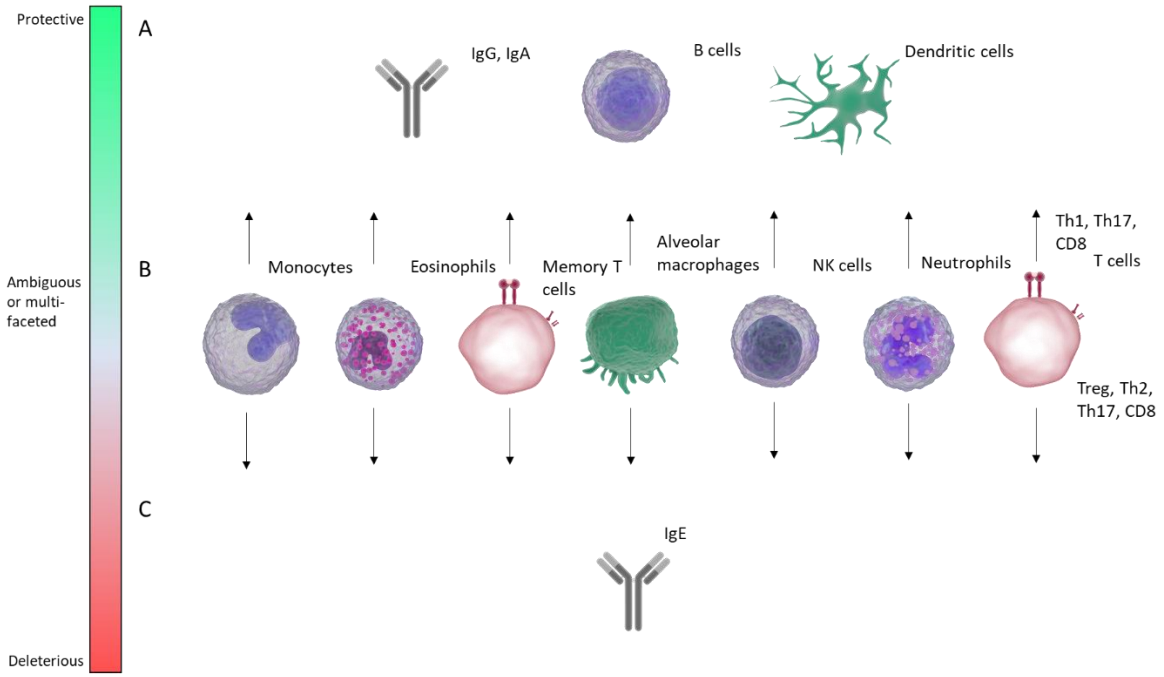


Figure 1.3 Components of the immune response found to have a role in the immune response against RSV in the mouse model, colour gradient and arrangement indicate their potential role in pathogenesis. (A) Cell types unambiguously protective against severe disease, indicated by green background such as B cells, Dendritic cells and neutralising IgG and IgA, (B) ambiguous or multi-faceted roles in disease pathogenesis, indicated by the blue area on the gradient, arrows indicate roles for these cells in protection or pathogenesis, with arrows pointing upwards indicating a possible protective role and arrows pointing downwards indicating a possible deleterious role (C) cells with a primarily harmful role in disease pathogenesis, indicated by placement at the bottom of the figure and red area on the gradient

1.6 Transcriptomic response to RSV infection

Transcriptomics is the study of changes in gene expression across a large set of genes aiming to understand the change in expression of thousands or tens of thousands of genes simultaneously using technologies such as microarrays and RNA sequencing (RNA-seq). Some of these methods (such as RNA-seq) indiscriminately capture every RNA transcript that can be amplified that is present in the sample being analysed, while other methods, such as microarrays take a more directed approach, and capture expression in a curated range of genes, that can either aim to be general and broadly representative of the global transcriptome, or to enrich for genes associated with certain tissues or functions, such as immunity. These technologies allow for a comprehensive understanding of changes in gene expression and cell composition (particularly immune recruitment) to a given tissue. A

notable limitation of these technologies is that they do not account for the potential for different translation rates for different RNA transcripts, meaning that the changes in gene expression do not necessarily totally reflect the change in the final amount of protein produced. Despite this, the scope and relatively lower cost of these technologies compared with similarly comprehensive proteomics methods makes them an invaluable tool in the development of our understanding of RSV disease. Other advantages such as the huge dynamic range in gene expression detectable by transcriptomic technologies make a desirable option.

The transcriptomic response to RSV infection has been previously characterised in a number of studies. Differences have been observed between the transcriptomic responses to RSV in the blood and in the upper and lower respiratory tract.

The tissue that transcriptomic analysis is performed on influences the sets of genes that are observed to be up-regulated in response to infection as immune cells will accumulate at the site of infection. Blood transcriptome responses in RSV infection show the effects of the transit of soluble or immunological factors to the respiratory tract, traffic towards the respiratory tract by immune cells or the response to any pathogens that are capable of replicating in the blood. While this may make whole blood an imperfect tissue for direct study of disease pathogenesis, whole blood can be an ideal tissue for biomarker development due to its relative ease of access.

Studies analysing the transcriptomic response to RSV infection						
Name	Authors/year	Tissue	Time points (from hospital admission)	Sample size	Outcome	Technology used
Nasopharyngeal Microbiota, Host Transcriptome, and Disease Severity in Children with Respiratory Syncytial Virus Infection(238)	Piters et al, 2016	Blood	<24h	132 total 106 RSV+ 26 healthy controls	Severity score (GRSS)	Microarray
Olfactomedin 4 Serves as a Marker for Disease Severity in Pediatric Respiratory Syncytial Virus (RSV) Infection(275)	Brand et al, 2015	PBMCs from whole blood	<24h	26 total Mild – 9 Moderate – 9 Severe – 8	Severity (supplemental oxygen)	Microarray
Whole Blood Gene Expression Profiles to Assess Pathogenesis and Disease Severity in Infants with Respiratory Syncytial Virus Infection(276)	Mejias et al, 2013	Blood		174 total RSV pos – 135 Healthy control – 39	Severity (severity score, hospitalisation, supplemental oxygen)	Microarray
Microbiome-Transcriptome Interactions Related to Severity of Respiratory Syncytial Virus Infection(277)	Sonawane et al, 2019	$CD4^+$ T cells from whole blood	When symptoms were present	58 total 22 for transcriptomics analysis	Severity score (GRSS)	RNA sequencing
Transcriptome assists prognosis of	Jong et al, 2016	Whole blood	<24 hours after	39 total, 7 mild, 14	Oxygen requirements	Microarray

disease severity in respiratory syncytial virus infected infant(278)			presentation in hospital	moderate, 18 severe		
Airway Gene-Expression Classifiers for Respiratory Syncytial Virus (RSV) Disease Severity in Infants(279)	Caserta et al, 2019	Nasal epithelial cells	<24 hours after presentation in hospital, again 28 days later	106 at acute time point 69 at convelescence	Severity score (GRSS)	RNA sequencing
Cord Blood Gene Expression in Infants Hospitalized with Respiratory Syncytial Virus Bronchiolitis(280)	Fjaerli O et al, 2007	Whole blood, cord blood	<24 hours after presentation in hospital	5 with acute disease	Hospitalisation	Microarray
Whole blood gene expression in infants with respiratory syncytial virus bronchiolitis(281)	Fjaerli O et al, 2006	Whole blood	<24 hours after presentation in hospital	5 with acute disease	Hospitalisation	Microarray
Transcriptomic Profiling in Childhood H1N1/09 Influenza Reveals Reduced Expression of Protein Synthesis Genes(282)	Herberg et al, 2013	Whole blood	<24 hours after presentation in hospital	25 RSV infected patients	N/A	Microarray
Nasopharyngeal gene expression, a novel approach to study the course of respiratory syncytial virus infection(283)	Kieboom et al, 2015	Nasopharngeal swabs	<24 hours after presentation in hospital	20 RSV infected patients 10 healthy controls	Requirement of respiratory support	Microarray

Host Transcription Profile in Nasal Epithelium and Whole Blood of Hospitalized Children Under 2 Years of Age With Respiratory Syncytial Virus Infection(284)	Do et al, 2017	Nasal cells, whole blood		83 total 38 acute RSV infected, 15 convalescent, 30 rhinovirus infected	ICU admission or requirement of respiratory support	Microarray
Immune profiles provide insights into respiratory syncytial virus disease severity in young children(285)	Heinonen et al, 2020	Whole blood		190 total 125 Acute RSV infected 65 Healthy controls	Severity score (CDSS) and hospitalisation	Microarray

Table 1.6.1 Transcriptomic studies using samples taken from patients experiencing an ongoing RSV infection.

One RSV study using microarray technology in whole blood(276) aimed to identify a multi gene signature that predicted outcome of infection. The identified genomic score correlated significantly with a clinical disease severity score, as well as duration of need for supplemental oxygen and requirement for hospitalisation. Additionally, multiple other factors were observed in this study that associated with an increased chance of developing more severe RSV disease. This study also aimed to compare RSV infection to infection by other respiratory infections, specifically human rhinovirus (HRV) (n=30) and human influenza (n=16) lower respiratory tract infections. Findings of this study were validated in independently recruited cohorts, with the genomic score identified in the initial training cohort of 46 patients recruited from Dallas, Texas, being tested in another 46 person cohort from the same site, and validated in two separate validation cohorts from Turku, Finland and Columbus, Ohio. Based on this study design, a genomic score was identified that could discriminate between infection with RSV, HRV and influenza with 95% accuracy, a sensitivity of 94% and a specificity of 98%. Gene set for discrimination between infections was developed through use of the k nearest neighbour algorithm on the normalized expression values of the gene sets from each cohort. The design of this study allowed it to identify gene

signatures associated with RSV infection, more severe RSV infection, the effect of age on response to RSV infection and the differences between RSV and other respiratory infections.

The general blood response to RSV infection identified in this study was defined by an upregulation of interferon, neutrophil and myeloid response genes, with downregulation of B and T cell associated genes. Many genes and pathways were shared between RSV and influenza in particular and differed primarily by the degree of induction or down-regulation. RSV could be discriminated from HRV and Influenza by a particularly high degree of neutrophil up-regulation when compared with the two other pathogens, and a particularly high degree of down-regulation of genes associated with B cell, T cell, lymphoid and antibacterial responses. More severe RSV disease was associated in this study with a particularly strong down-regulation of CD4, CD8, NK cell, B cell and plasma cell responses. Patients with moderate disease and severe disease experienced comparable levels of interferon up-regulation, which was higher than patients with mild disease. Neutrophil and inflammatory genes were consistently more highly expressed in patients with more severe disease. Downregulation of genes associated with adaptive immune responses were particularly down-regulated in younger patients (<6 months) suggesting this as a possible reason why this group experiences a higher burden of disease. Downregulation of B cell related genes persisted even a month post infection, while expression of other genes returned to baseline levels, while upregulation of interferon genes was found to be even higher in convalescence than during acute disease. This study lacked patients on the milder end of the RSV disease spectrum, with all patients involved being hospitalized.

Another study conducted by the same group(238) analysed the transcriptomic response to RSV in a cohort of 132 children, 106 of which had an RSV infection, and 26 of which did not. The population of children with an RSV infection was further subdivided into inpatients (n=84) and outpatients (n=22). The primary aim of this study was to determine the influence of the nasal microbiome on the transcriptomic response to RSV and on disease severity, and so a gene expression analysis analysing differences in expression between patients with severe and milder disease was not shown. In place of gene expression analyses, an integrative approach involving the compression of highly correlated gene expression patterns into a single, uncorrelated variables accounting for large bodies of co-correlated genes was used. This was accomplished in this study by use of principle component analysis

(PCA). Principal components were selected to account for over 50% of the variance in the data set. Multivariate linear models and calculated effect sizes were used to determine which gene components associated particularly highly with any given bacterial taxonomic groups. Pathway analysis was used to assign biological significance to each selected gene principal component. Additionally, a differential expression analysis was performed to compare groups of participants with different dominant bacterial taxa. Subsets of patients were selected using an unsupervised clustering algorithm based on their dominating bacterial taxa, and relative expression of genes was compared between these groupings. No direct comparisons were made in this study between gene expression and disease severity, rather the relationships between bacterial taxa and disease severity were examined, and gene clusters were found that associated more highly with those taxa.

After unsupervised clustering of patients by dominant taxa, gene expression was compared between patients with dominant *H. influenzae*– and *Streptococcus* taxa (which were found within the study to be associated with disease severity) and the patients with dominant *Corynebacterium, Moraxella*, and *S. aureus* clusters. The *H. influenzae*– and *Streptococcus* clusters were associated with increased expression of interferon, neutrophil and monocyte modules, and decreased expression of B cell, T cell, NK cell and plasma cell modules. Formal differential expression analysis was also performed between these two groupings, and found a particularly high degree of up-regulation of neutrophil and neutrophil activity related genes, such as genes associated with IL-8 and IL-17A signalling. A core signature of RSV infection that was independent of dominant bacterial taxa was also identified, which was primarily characterised by overexpression of genes related to interferon responses, pattern recognition receptors and general inflammatory responses, such as response to NF-κB.

The integrative approach used in this study has some inherent limitations. It is difficult based on this approach to understand the causes of more severe disease. By primarily analysing the interaction between two elements that both may have some bearing on disease severity, it is difficult to determine how either influence disease severity. It is also impossible without further studies to determine which variable may be inducing the other, or the extent to which they influence each other. It is equally possible that changes in dominant bacterial taxa are induced by recruitment of immune cells to the respiratory tract

in response to infection. It is possible that the dominant taxa in the respiratory tract have some degree of influence over the cells which are recruited and their activity. This study was also limited in that it lacked the power to make distinctions between different degrees of disease severity beyond the differences between hospitalised and non-hospitalised patients.

A third study conducted by this group(285) utilized a combination of this transcriptional module approach and cytometry data to understand the relationship between the immune response and disease severity in children under the age of two. The module based approach using transcriptomics data found modules related to interferon, CD8 T cells, NK cells and T cells to be related to less severe disease (defined by lack of admission to hospital) while modules associated with monocytes, neutrophils and inflammation were more highly expressed in more severe disease. Cytometry data from this study showed that association of genes related to severity is not necessarily mirrored in the abundances of cell types in the respiratory tract. Neutrophils did not show increased abundance in whole blood of children below the age of six months despite transcriptomic associations between neutrophil signatures and disease severity in those patients, indicating that the relationship between neutrophils and severity in this age group may be more related to function than abundance.

Another study(277) took a similar approach using a transcriptional module based approach to analyse the interactions between the host microbiome and transcriptome, and the influence that these interactions have on disease severity. This study analysed the nasal microbiome and the transcriptome of $CD4^+$ T cells isolated from whole blood in 58 patients with a confirmed RSV infection. Of the 58 patients recruited to this study, only 22 had transcriptomics samples taken at both baseline and at follow-up two weeks later. Seven of these patients had mild RSV disease and 15 had severe RSV disease according to the GRSS clinical severity score used in this study(286).

Twenty seven differentially expressed genes were identified in this study between acute infection and convalescence (FDR <.05). All differentially expressed genes found to be more highly expressed at the acute time point were genes involved in interferon responses, such as *IFI27*, *IFITM1*, *IFITM2*, *MX1* and *ISG15*, while at the convalescent time point two weeks later more genes associated with B cell responses were more highly expressed, such as *MS4A1*(287). Additionally, a differential expression analysis was performed between patients with severe disease and patients with milder disease. This analysis found 600 genes

to be differentially expressed using a nominal p value cut off of <.05. *EZH1* was identified as the gene most highly associated with RSV severity by this analysis. *EZH1* is associated with TLR induced production of pro-inflammatory cytokines, such as IL-6, TNF-α and IFN-β, suggesting that excessive early inflammatory cytokine released mediated by PRRs might contribute towards increased disease severity.

A further novel method of analysis was employed by this study, similar to that used in Piters et al(238) to associate gene expression (summarized in principal components) and bacterial expression by correlating gene principal component expression for each patient with bacterial principal component expression, and further to estimate these relationships in each individual patient through use of the LIONESS algorithm. By estimating these correlations in individual patients, it was then possible to associate relationships between genes and bacterial clades with disease severity.

A study conducted by Jong et al(278) identified an 84 gene signature that could discriminate between hospitalized patients that will go on to develop severe RSV and those that will only experience mild disease. Disease severity was defined in this study by the requirement for supplemental oxygen. All patients in this study were hospitalised, meaning this study was unable to capture the lowest end of the disease severity spectrum. Patients were classified as experiencing mild disease if they did not have hypoxia, moderate if patients received some supplemental oxygen, and severe if the patients required mechanical ventilation. Patients experiencing moderate and mild disease were grouped together, and gene expression compared to those with severe disease. The patients were divided between training and test sets using leave one out cross validation (LOOCV) and the support vector machines algorithm (SVM)(288) was used to identify highly co-correlated gene expression features that could predict more severe disease. Biomarkers were independently validated in a previously published data set(276).

Ninety-five genes were found to be differentially expressed between severe cases and the combined set of patients experiencing mild and moderate disease (FDR <.05, absolute fold change >2). One hundred and seventy eight were observed between severe and mild disease, 49 between severe and moderate disease and 221 between severe disease and healthy controls. No genes were found to be differentially expressed between patients with mild disease and healthy controls. Only 1 gene was differentially expressed between

convalescence and healthy controls. Gene enrichment on the set of genes differentially expressed between healthy controls and severe disease revealed upregulation of terms relating to neutrophils activity and signalling as well as neutrophil activity and signalling. Machine learning prediction using SVM on a training set was able to correctly classify all but 5 patients disease severity status correctly. This gene set was reduced to an 84 gene signature that could predict progression to more severe disease with an LOOCV area under curve (AUC) of 0.96 in the patients recruited to the study, and an AUC of 0.85 in the independent validation cohort.

This study suffered limitations similar to other studies by only analysing RSV disease in hospitalised patients. Additionally, patients up to the age of two were recruited to this study, which may hurt comparability with studies that restricted recruitment to children below the age of one. Gene expression was standardised for age and sex in order to remove potential confounding effects of these variables on gene expression to ensure comparability between this study and others.

One smaller study using microarray technology in a cohort of 26 children with either severe, moderate or mild RSV infection identified an overall trend towards lower $CD4^+$ T cells, $CD8^+$ T cells and NK cells in the peripheral blood of patients with more severe RSV(275). This study combined data from whole blood transcriptomics with confirmatory assays using RT-qPCR in larger cohort of 80 children to confirm the findings of selected genes. Data from flow cytometry experiments were also integrated into this analysis. Disease severity was determined by requirement of supplemental Oxygen or breathing support. Mild patients had no hypoxia or severe breathing problems, the moderate group required hospitalisation for supplemental oxygen or feeding support, while severe disease was classified as patients requiring mechanical ventilation. This study found 564 genes to be differentially expressed between mild and severe disease, and 808 genes to be differentially expressed between patients experiencing acute severe disease and the same patients in convalescence. Olfactomedin4 (*OLFM4*) was the most highly differentially expressed gene in all analyses and identified as a discriminatory biomarker for severe RSV in this study. Patients with elevated *OLFM4* expression were found to be 6 times more likely to develop severe disease. Prior to this study *OLFM4* had not been previously identified as being involved in respiratory tract infections. This study was limited by the relatively low sample size. Additionally,

microarray and PCR data were integrated for biomarker analysis, which are not necessarily comparable and may create additional variability. The discrepancies between this study and others in terms of most highly differentially expressed genes may be attributable to the purification of PBMCs from whole blood before RNA extraction and microarray analysis instead of directly from whole blood as in other studies. The use of PBMCs as opposed to whole blood or nasal epithelial cells makes this study difficult to compare to other transcriptomics studies.

Two studies conducted by Fjaerli et al(281)(280) used microarray technology to identify genes associating with RSV bronchiolitis (when compared with asymptomatic disease)(281) and genes associated with development of RSV bronchiolitis in the cord blood of infants who would go on to develop a severe RSV infection(280). Both of these studies were limited by a small sample size (5 cases and 5 controls in each), and also by a lack of mild RSV cases, as only patients with bronchiolitis and asymptomatic infection were included. While a validated gene signature could not be developed in these cohorts, a number of genes associating with more severe RSV disease were identified. *IFI27* and *CLC* were differentially expressed in the blood of patients experiencing bronchiolitis compared with asymptomatic controls, while *FAM102A*, *TNFRSF25*, and *STMN3* were significantly downregulated in the cord blood of patients who would develop severe RSV disease. The approach of using cord blood gene expression to develop a biomarker for future severe RSV infection is novel within the RSV field and may be a useful tool in biomarker development.

There are relatively few studies that have utilised RNA-sequencing to analyse the host response to RSV infection. This may be due to this being a relatively new technology, but also microarray technology costs considerably less. One study that analysed gene expression from cells isolated from the nasal epithelium made use of RNA sequencing to identify a biomarker for severe RSV disease(279).

This study was made up of three cohorts, one was a cohort made up of patients presenting in hospital with RSV disease, another cohort followed patients from birth with sampling at birth and upon infection with RSV. A third cohort was made up of patients experiencing an RSV infection that presented at local clinical practices or outpatient clinics. Samples for transcriptomic analysis were collected by brushing of the nasal turbinate with a flocked swab. Of the 106 samples used for transcriptomic analyses, 42 of these came from patients

with mild disease (GRSS ≤3.5) and 56 came from patients with severe disease (mean GRSS of 6.13, SD 0.22). The set of all genes 6844 that passed pre-processing and filtering were correlated with disease severity. Sixty-six genes were identified that correlated with disease severity (FDR < .05). A model was developed using these genes and some other genes that were not considered differentially expressed. This model could correctly classify patients as having severe or mild RSV disease with an accuracy of 89.6%. An additional analysis was performed on this patient cohort that aimed to develop a predictive biomarker for disease severity by using genes that associated with more severe infection but that also did not have a significant difference in expression between acute infection and convalescence. This predictive score correctly identified patients as having severe or mild disease 84% of the time. However, it was not tested in any patients before the development of RSV disease, meaning its efficacy has yet to be confirmed. Whether these genes are simply induced in severe infections and remain consistently expressed for a long period after infection or are highly expressed prior to infection in patients that will go on to develop severe infection remains to be seen.

This study also noted that gene expression in the upper respiratory tract was not significantly different from what had been previously observed in the lower respiratory tract. This observation would require a dedicated study to confirm; however, this data set provides a promising indication that these tissues may be comparable, which would make implementations of future clinical trials easier, as the upper respiratory tract is relatively more accessible than the lower.

A study that analysed the gene expression in response to RSV infection in both blood and nasal swabs that also included disease severity as an outcome was conducted by Do(284) et al. Thirty eight RSV infected children between the age of 1 month and 2 years were recruited to this study, with 15 convalescent samples, 30 human rhinovirus infected patients. Differentially expressed genes were identified for both blood and nasopharyngeal samples between acute infection and convalescence. Higher numbers of differentially expressed genes were observed on average for nasopharyngeal samples, indicating that a greater degree of differential gene expression in response to RSV occurs there. Interferon responses (α/β and γ), Th17 responses and general cytokine signalling responses were upregulated in both blood and nasal samples. Highly co-regulated gene sets were correlated

with outcome measures, including disease severity and viral load. The modules associated with severe RSV contained 309 genes in nasal swabs and 118 genes from blood. Genes highly associated with severity in both networks included *HIST1H4E*, *IL7R*, *ISG15* and *BCL6*. This study was limited in that it lacked patients with mild RSV infection, making it impossible to compare the differences between mild and severe RSV. The gene modules that differed between blood and nasopharyngeal RSV infection were not presented in this study, and would have provided a useful reference for understanding the relative limitations of the different tissues for transcriptomics analysis.

Van den Kieboom et al(283) conducted a transcriptomics study in nasopharyngeal swabs aimed at identifying a biomarker for severe RSV disease. Thirty patients were included in the analysis, with severity indicated by requirement of supplemental oxygen (n=10) or mechanical ventilation (n=10). Mild patients required no additional support. A number of genes were found to be associated with severe disease in this study, specifically ubiquitin D, tetraspanin 8, mucin 13, β-microseminoprotein and chemokine ligand 7. However, no robust feature selection and validation methods were applied in this study, so the usefulness of these identified biomarkers is limited. This is likely due to the impossibility of developing robust training and test data sets with the small number of patients recruited to this study. Additionally, no functional analysis was performed on the differentially expressed genes found identified in this study between severe and non-severe RSV disease.

There have been a number of studies that aimed to analyse the transcriptomic response to other viral respiratory pathogens such as influenza and rhinovirus, but which also performed transcriptomic analysis on samples from patients infected with RSV. One such study was conducted by Herberg et al(282), and compared the gene expression profile of 34 children infected with RSV to 25 with influenza infection and 33 healthy control children. This study only aimed to identify transcriptional signatures associated with disease for exploratory purposes, and so did not have robust outcome measures that would allow biomarker development or assembly of gene signatures associating with disease phenotypes. A signature was developed that aimed to distinguish between RSV infection and other infections using the SVM feature selection algorithm. The reported signature could distinguish between RSV and bacterial infection at 95% sensitivity. While this study reports robust numbers of differentially expressed genes, the numbers are higher than what has

been reported in other studies of similar sample size, and a lack of repeat sampling and outcome measures makes it impossible to associate the transcriptomic features identified here with any aspect of disease. This study was primarily focused on exploratory analysis of influenza signatures and did not present functional analysis of RSV induced gene expression, and overall was poorly suited to improving understanding of RSV infection.

Other studies have performed transcriptomic analysis of respiratory pathogens without a distinct focus on a specific pathogen, however many of these studies have an insufficient number of RSV infected patients to identify any differentially expressed genes or statistically robust signatures associated with RSV disease(289)(290).

A consistent problem in associating disease severity with changes in gene expression is the metric by which severity is measured. Numerous diverse methods of severity assessment have bene used in different gene expression studies, many of them varying hugely with each other. The three main metrics that have been used in these studies are severity scores, supplemental oxygen requirements and hospitalisation. Severity scores, while being heterogeneous, tend to consistently take into account a number of factors, such as receipt of supplemental oxygen, temperature, intolerance to feeding, respiratory difficulty and respiratory rate. Examples of such scores that have been used include GRSS(286) and the ReSVinet score(291). Supplemental oxygen requirement discriminants tend to either use the necessity of receipt of supplemental oxygen as a marker for severe disease, or some gradient of the degree of breathing support required. Many studies have used requirement for supplemental oxygen as a metric(292)(276)(278)(283)(284), making it useful for ensuring comparability between studies. Supplemental oxygen requirements also make up a part of the GRSS(286), which has also seen wide use in gene expression studies(238)(277)(279)(285). Finally, the requirement for hospitalisation is a simpler scoring system, used by a few studies. This metric, while useful in single site based studies with consistent hospitalisation criteria, will likely vary enough between hospitals that it may be a poor criteria by which to determine disease severity for the purpose of meta-analysis.

The most commonly identified signals detected in whole blood in response to infection with RSV are neutrophil related signals. A trend towards decreased expression of B and T cells in the whole blood has consistently been observed. It is possible this is due to some immune

dysregulation relating to severe RSV infection; however, it may also be attributable to an accumulation of these cells at the site of infection in the lower respiratory tract.

1.7 Treatment of RSV

Despite the large global burden of RSV, no effective dedicated treatment currently exists. The only drug specifically licensed for treatment of RSV infection is Ribavirin, which is not recommended for routine clinical usage due to limited efficacy and toxicity(293). The prophylactic monoclonal antibody palivizumab has also been licensed for RSV(294), however palivizumab is only prescribed to high risk patients, and is prohibitively expensive(295). Palivizumab binds to the RSV F protein and has demonstrated efficacy in the prevention of severe RSV infection. Palivizumab has been associated with a 39-78% decrease in hospitalisations(296).

Due to the lack of cheap and effective drugs for RSV infection, most treatment is supportive. Supportive treatment for RSV includes primarily hydration, feeding support and respiratory support. Supplemental oxygen treatment is used in severe RSV infection when a patient's oxygen saturation dips below 90%. In patients with apnea or respiratory failure mechanical ventilation may be employed. Around 5% of healthy infants hospitalized with an RSV infection will require mechanical ventilation, however the proportion of patients is higher in those with chronic underlying conditions, such as heart disease(297).

Efforts in developing an RSV vaccine were slowed due to the controversial formalin inactivated RSV vaccine trial in the 1960s, which resulted in two deaths, and hospitalisation of over 80% of recruits to the trial due to the vaccine causing the induction of enhanced disease upon exposure to RSV post vaccination(298). The causes of enhanced RSV disease in response to the formalin inactivated vaccine remain poorly characterised. Enhanced RSV disease took place in children who were seronegative for RSV prior to vaccination. Following vaccination these children experienced an increase in frequency and severity of respiratory tract infections. Enhanced RSV disease has been characterised by fever, wheezing and bronchopneumonia, as well as an increased rate of hospitalisation(298). While not experiencing a significant increase in the rate of respiratory infections, infants vaccinated with the formalin inactivated RSV vaccine experienced an 80% rate of hospitalisation in response to infection with RSV, while only 5% of those who were vaccinated with the

control vaccine strain were hospitalised within the same period(298). Infants who experienced enhanced disease had excessive neutrophil, eosinophil and mononuclear cell infiltration of the respiratory tract(299), as well as Th2 polarization of lymphocytes(189), ineffective antibody(300) and CTL responses post infection(301), and pulmonary immune complex deposition(302). One hypothesis explaining the causes of enhanced disease after vaccination with the formalin-inactivated RSV vaccine, is that vaccination failed to induce strongly neutralising responses against viral epitopes (due to poor affinity maturation), while simultaneously weakly activating TLRs results in a poorly protective response(300). Simultaneously the vaccine induces immune complex formation, where IgG and complement protein C3 are deposited in alveoli and bronchioles of vaccinated mice. While immune complexes were still present in $C3^{-/-}$ mice, the lung histopathology of these was similar to wild type mice, however airway hyperreactivity (AHR) was drastically reduced. Both $C3^{-/-}$ mice and B $cell^{-/-}$ mice experienced reduced bronchoconstriction on exposure to RSV after vaccination when compared with wild type mice(303). These immune complexes were also found in lung biopsies of children who died due to enhanced disease following vaccination. The reasons for the induction of immune complex formation are not yet well understood, however immune complex formation has previously been linked to the presence of low affinity antibody(304). The affinity maturation and immune complex hypothesis is supported by the fact that enhanced disease did not occur in patients who were seropositive for RSV, and who will have had some previous high avidity B cells in circulation that would likely outcompete the low avidity vaccine induced B cells upon RSV infection(305). Additionally, a formalin inactivated measles vaccine has been associated with both enhanced disease after exposure to the pathogen, and poor affinity maturation of B cells upon vaccination(302). In light of this, the generation of adequately neutralising antibodies on vaccination against RSV should be considered as an outcome in all future pre-clinical vaccine trials against RSV.

RSV vaccination efforts remained controversial for a number of years after this trial, however in more recent years candidate vaccines have begun to emerge (figure 1.4).

Maternal immunity (in particular through the medium of RSV neutralising antibodies) is passed on to the infant after birth and provides some protection up to four months of age. Maternal vaccination against influenza and Pertussis has shown protection up until two to

three months of age(306), mirroring the highest risk window for severe RSV infection(307). Transplacental antibody transfer reaches its highest efficiency during the third trimester of pregnancy, making maternal vaccination less effective in pre-term infants(308). The novavax F protein nanoparticle vaccine took this approach and demonstrated efficacy in clinical trials(309) despite not achieving its primary clinical outcome.

Infant vaccination is also being considered for RSV. Infant vaccination is an approach that has shown efficacy for many diseases such as *Bordetella pertussis* and hepatitis B(310). There are however some potential reasons why it may not be the most appropriate strategy for RSV vaccination. Infants are at the highest risk of severe RSV infection in the first months of life(307), where efficacy of vaccines is lower than later in life(311). Maternal vaccination aims to circumvent this problem by ensuring neutralising antibodies are acquired by the infant through placental transfer, ensuring protection in the first two months of life before the age where most patients begin vaccinations(310). Vaccination of infants after the first two months of life may also be effective in preventing RSV infections in older children, where vaccines demonstrate greater efficacy.

RSV vaccination efforts have taken a number of different strategic approaches to protecting against RSV infection through a diverse range of means. One of the most promising vaccines in recent years was the Novavax F protein nanoparticle vaccine. This maternal vaccine made it into stage three clinical trials before failing to meet its primary endpoint of prevention of medically significant RSV LRTI in the first 90 days of life, however it showed some efficacy in preventing more severe disease(309). Children of mothers vaccinated with the Novavax nanoparticle F protein vaccine had a 25% reduction in all cause LRTI hospitalisations and a 39% reduction in all cause LRTI severe hypoxemia, indicating some efficacy in preventing more severe RSV disease. An RSV F nanoparticle vaccine is also being trailed in the elderly and in children(312). These trials are in phase one and phase two respectively(312). No other RSV vaccines are currently in stage three of clinical trials.

A variety of viral vector vaccines, subunit vaccines and neutralizing antibodies are currently in clinical trials. There are two subunit vaccines for RSV in phase three of clinical trials(313)(314), and four in phase one. Six of these vaccines use the RSV F protein as the antigen(313)(314)(315), while one uses the G protein(316) and another uses the SH

protein(317). Subunit vaccines typically contain some antigenic protein for the pathogen being vaccinated against linked to some carrier protein(318).

Viral vector vaccines are a newer vaccine type that have emerged in years. These vaccines are recombinant viruses that express antigenic genes but which are not competent to cause disease(319). Three vaccines of this type are currently in stage 2 clinical trials, one of which used an MVA vector, and two of which use an adenoviral vector(320)(321)(322). The adenoviral study being conducted by GSK (ChAd-155) uses recombinant RSV Fusion (F), Nucleoprotein (N), transcription processivity factor (M2-1)(322), while the Janssen vaccine uses the pre-fusion form of the RSV F protein(320) and the Bavarian Nordic vaccine uses the F and G proteins(321).

Other approaches that have vaccines currently undergoing trial include nucleic acid based vaccines(323) (such as the ones in preclinical trials sponsored by Moderna(324) and CureVac), live attenuated viruses (numerous candidates in phase one and preclinical trials)(325)(326)(327)(328), and whole-inactivated viral vaccines (one of which is in pre-clinical trials being sponsored by Blue Willow Biologics).

Additionally, numerous groups have attempted to develop a more effective immunoprophylaxis against infection through neutralizing antibodies. The only currently market approved RSV specific treatment Palivizumab (synagis) takes this approach. Other promising treatments in this family include the monoclonal antibody being developed by Sanofi and Astra Zenica, MEDI8897(329), which targets the RSV F protein, and has shown efficacy in phase 1 and 2 clinical trials and has demonstrated a 78.4% reduction in the rate of hospitalisation due to RSV in infants(330). Another neutralizing antibody is also currently in trials in infants called MK-1654, which is being developed by Merck has shown promising safety in phase one trials and is currently in phase two(331).

1.8 Biomarkers for RSV disease

Due to the difficulty of predicting who will develop severe RSV disease or respiratory sequelae, and the expense of current prophylactic options, there is an unmet need for a clinical biomarker that could aid in predicting children who may develop severe RSV disease, either before infection or during the earlier stages of infection. Likewise, a biomarker that

could predict the likelihood of developing respiratory sequelae during an ongoing RSV infection would allow for adjustment of treatment to minimize the likelihood of developing sequelae, while also allowing for more objective assessments of whether a patient can be discharged.

No biomarker predicting the severity of infections or development of sequelae is currently licensed, however much work has been done in the development of biomarkers for viral infections. Studies have been performed using transcriptomic technologies to try to distinguish between bacterial and viral infections, and have demonstrated up to 100% sensitivity and 96.4% specificity in test sets, and sensitivity of 91.7% and 90% in independent validation cohorts(332).

There are many difficulties associated with the development of a predictive biomarker for disease severity. It is difficult to develop a predictive biomarker without conducting a prospective study, where patients are followed from birth and sampled upon RSV infection. Prospective trials have been conducted in RSV, however the difficulties in recruiting large numbers of RSV positive patients using this approach has prevented the development of a predictive biomarker using this approach thus far. There have been some attempts at using statistical modelling to create a predictive biomarker by using the status of genes that associate with severity, but which remain highly expressed in convalescence. Biomarkers developed by such routes may be promising in future, however such methods require validation in prospective cohorts before these methodologies could be widely adopted.

Severity biomarkers			
Study	**Biomarker**	**Efficacy**	**Severity metric**
Jong et al 2016	84 genes	AUC = .858	Oxygen requirement
Brand et al 2015	1 (OLFM4)	N/A	Mechanical ventilation
Wang et al 2019	41 genes	89.6% correct classification	GRSS
Discriminatory biomarkers			
Study	**Biomarker**	**Efficacy**	**Predicts**
Mejias et al 2013	70 genes	94% sensitivity, 98% specificity	Infecting pathogen

Table 1.8.1 Summary of studies that produced a predictive transcriptomic biomarker Studies producing biomarkers either for RSV severity or RSV diagnosis are shown, as well as the number of genes in the predictive signature, and the measure of efficacy used in the study, where applicable

1.9 Aims and objectives of this thesis

The aim of this thesis is to develop a biomarker for RSV disease severity and respiratory sequelae using changes in gene expression in whole blood, the central objectives of this thesis are:

1. To systematically review the relationship between the immune response to RSV and its relationship to disease severity in selected animal models

2. To further develop understanding of the immune response to RSV and how it relates to disease severity

3. To further develop understanding of the immune response to RSV and how it relates to development of respiratory sequelae

4. To develop a biomarker predictive of progression of RSV infection to more severe disease or development of respiratory sequelae

I aim to address each of these questions within a separate chapter of this thesis. Firstly, I aim to outline the primary methods that I will use to address these objectives, alongside an in depth description of the observational clinical trial that produced the data that will be used to address the first three objectives of this thesis (chapter 2). Next I will seek to demonstrate a descriptive analysis of the transcriptomic response to RSV in infants, as well as describe its relationship to disease severity (chapter 3). I will then seek to describe the relationship between changes in gene expression in RSV infected patients and the likelihood of developing respiratory sequelae after recovery from disease (chapter 4). I will then describe the development of a statistical model that will be used to develop a biomarker for both more severe disease and respiratory sequelae and demonstrate the efficacy of this model in test and validation datasets (chapter 5). Finally, I will aim to describe the relationship between disease severity in animal models and certain selected aspects of the immune response to RSV disease in these models (chapter 6). The final chapter will discuss this work and speculate on the further development of this work.

Chapter 2: General materials and methods

2. 1 Buffers

Buffer	Catalogue Number	Supplier
Paxgene RNA stabilization solution	762165	QIAgen
Buffer BR1 (Resuspension Buffer)	762164	QIAgen
Buffer BR2 (Binding Buffer)	762164	QIAgen
Buffer BR3 (Wash Buffer)	762164	QIAgen
Buffer BR4 (Wash Buffer)	762164	QIAgen
Buffer BR5 (Elution Buffer)	762164	QIAgen
RNase-free Water	762164	QIAgen
Proteinase K	762164	QIAgen
DNase I, RNase-free	762164	QIAgen
Pure ethanol (96-100%)	762164	QIAgen
Buffer RDD	762164	QIAgen

Table 2.1.1 Buffers and reagents used for RNA extractions

Buffer	Catalogue Number	Supplier
Nuclease-free Water	902927	AppliedBiosystems
Pico First-Strand Buffer	902927	AppliedBiosystems
Pico First-Strand Enzyme	902927	AppliedBiosystems
Pico 3' Adaptor Buffer	902927	AppliedBiosystems
Pico 3' Adaptor Enzyme	902927	AppliedBiosystems
Pico IVT Buffer	902927	AppliedBiosystems
Pico IVT Enzyme	902927	AppliedBiosystems
Pico 2nd-Cycle Primers	902927	AppliedBiosystems
Pico 2nd-Cycle ds-cDNA Buffer	902927	AppliedBiosystems
Pico 2nd-Cycle ds-cDNA Enzyme	902927	AppliedBiosystems
Hydrolysis buffer	902927	AppliedBiosystems
Neutralization buffer	902927	AppliedBiosystems
Purification beads	902927	AppliedBiosystems
80% ethanol wash solution	902927	AppliedBiosystems
Poly-A RNA Controls	902927	AppliedBiosystems
Pico cleanup reagent	902927	AppliedBiosystems
Pico Fragmentation & Label Buffer	902927	AppliedBiosystems
Pico Fragmentation & Label Enzyme	902927	AppliedBiosystems

Table 2.1.2 Buffers and reagents used in synthesis of cDNA for microarray hybridisation

Buffer	Catalogue Number	Supplier
DMSO	902927	AppliedBiosystems
5X WT Hyb Add 1	902927	AppliedBiosystems
Control Oligo B2 (3 nM)	902927	AppliedBiosystems
20X Eukaryotic Hybridization Controls (bioB, bioC, bioD, cre)	902927	AppliedBiosystems
15X WT Hyb Add 4	902927	AppliedBiosystems
Nuclease-free Water	902927	AppliedBiosystems
Hybridization Master Mix	902927	AppliedBiosystems
2.5X WT Hyb Add 6	902927	AppliedBiosystems

Table 2.1.3 Buffers and reagents used in microarray hybridisation

2.2 RESCEU Case control study

The RESCEU consortium (Respiratory Syncytial Virus in EUrope) is a large international consortium aiming to bring together experts in numerous fields across Europe. The RESCEU consortium aims to bring experts of many more disciplines to further the understanding of the impact of RSV disease, as well as it's epidemiology and pathogenesis. The RESCEU consortium is divided into work packages, each with a distinct task. The goals of these work packages range from establishing a comprehensive biobank of samples from RSV infected patients, to conducting systematic reviews, to the present biomarker study.

2.2.1 Participants

Within the clinical study work occurring within RESCEU, there are four primary clinical trials (figure 2.2.1) that each recruit from different populations.

2.2.1.1 Infant active study

The first is the infant active study, to which patients were recruited at birth, and followed for the first year of life, with samples taken at the time of recruitment and each time the participant underwent a respiratory infection, and seven weeks following a confirmed RSV infection. All patients recruited to the study were followed up one year, two years and three years after recruitment to the study to assess the development of respiratory sequelae after infection.

2.2.1.2 Case control study

The second study, that the present work is based upon was the infant case control study, to which children experiencing an ongoing RSV infection were recruited, either from paediatric intensive care units or from local healthcare centres. Additionally, healthy control patients who are not experiencing any respiratory symptoms were recruited to the study. These patients were sampled during acute disease, and once more during convalescence seven weeks later. Any ongoing symptoms were recorded during every sampling visit All patients recruited to the study were followed up one year, two years and three years after recruitment to the study to assess the development of respiratory sequelae after infection.

2.2.1.3 Older adult study

The third study was the older adults study, wherein adults older than the age of 60 were recruited from community healthcare centres before the start of RSV season, and were tested for RSV and sampled during every respiratory infection. Samples are taken at the time of recruitment and again after a confirmed RSV infection, and once more at the end of the RSV season.

2.2.1.4 COPD study

The final clinical study within the RESCEU consort is the COPD study, where older adult patients with COPD were regularly sampled and tested for RSV infection, and further samples were taken for analysis during a confirmed RSV infection and seven weeks later in convalescence. The present work is concerned with the infant case control cohort.

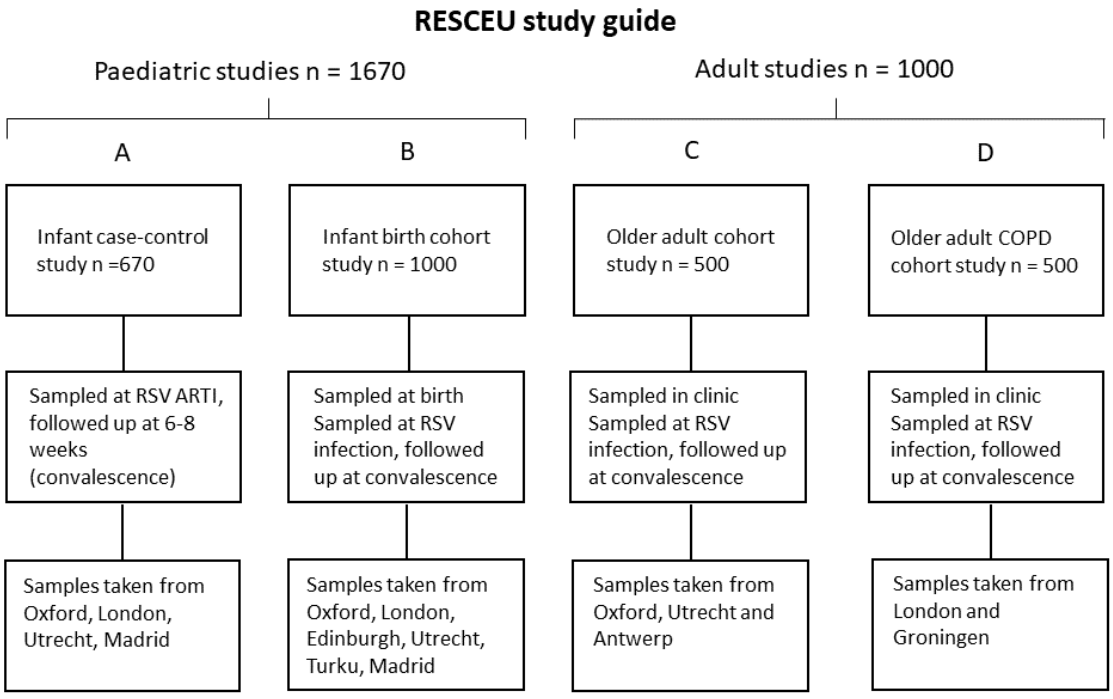


Figure 2.2.1.1 Description of RESCEU study cohorts Study is divided into four cohorts, two in adults and two in children.

2.2.2 Power calculations

Study sizes were designed according to power calculations designed to maximise capacity to create robust biomarkers. Case control study 670 children will be recruited, with 570 cases and 100 controls giving 90% power (SD 0.4), with a true biomarker rate of 0.5% and log odds ratio of 0.18. B: Infant cohort study aims to estimate the prevalence of RSV ALRI. A prospective cohort of 1000 children < 1 years is followed for 1 year assuming α=0.05; power ≥ 0.95 and assumed loss to follow-up = 10%. C/D: Older adult study and COPD study aim to estimate the prevalence of RSV ALRI. A prospective cohort of 1000 elderly > 60 years is followed for 1 year assuming α=0.05; power ≥ 0.95 and assumed loss to follow-up = 10%.

2.2.3 RESCEU infant Case control study objectives and outcome measures

As the work presented in this thesis deals exclusively with the infant case control study all further methods exclusively relate to that study

		Objectives	Outcome Measures
Primary	**1**	To establish biomarkers predictive of, or associated with, lower respiratory tract RSV infection and disease severity in infants	RSV disease severity will be assessed using the ReSVinet score and assessing the need for hospitalisation or not.
Secondary	**1**	To establish biomarkers associated with sequelae following RSV infection in infants	Sequelae of RSV will be assessed by parental questionnaire at 1,2 and 3 years of age.
	2	To characterise the viral load and genetic sequence of RSV associated with mild and severe disease	RSV viral load (copies/ml) and genetic sequence will be assessed from respiratory swabs. RSV disease severity will be assessed using the ReSVinet score and assessing the need for hospitalisation or not.
	3	To characterise the immune response to mild and severe RSV disease	RSV disease severity will be assessed using the ReSVinet score and assessing the need for hospitalisation or not. Immune responses will be assessed by flow cytometric cell phenotyping, intracellular cytokine staining, systems serology and neutralising antibody.
	4	To describe the transcriptomic, proteomic, metabolomic and epigenetic signatures associated with	RSV disease severity will be assessed using the ReSVinet score and assessing the need for hospitalisation or not. Sequelae of

		RSV disease severity and/or respiratory sequelae	RSV will be assessed by parental questionnaire at 1,2 and 3 years of age Transcriptomic, proteomic, metabolomic and epigenetic signatures of infants will be assessed from blood samples.
	5	To biobank specimens for later analysis of respiratory/stool microbiome, genetic associations with severe RSV disease and/or respiratory sequelae	Biological samples (respiratory, blood, stool and urine) will be stored in the Oxford biobank after completion of the study if parents consent to this
	6	To determine health care costs, health care resource use, interruption of normal activities, and Health Related Quality of Life (HRQoL) in RSV-associated ARTI patients and their families	Health care utilization and costs for RSV-associated ARTI will be assessed from visits to health providers (hospitals and GPs) and medication use.

Table 2.2.3.1 Objectives and outcome measures of the RESCEU case control study

2.2.4 Study participants

The RESCEU case control study was a prospective, observational clinical trial conducted at a number of sites across Europe. The trial aimed to understand the immunology and epidemiology of severe RSV disease in children below the age of one. Study participants were children below 12 months of age presenting in a clinic or hospital with a proven respiratory tract infection. Another group were children without a respiratory tract infecton recruited from the community as healthy controls.

The case control study recruited patients into 4 groups.

Group	Description	Number
1a	Children hospitalised with confirmed RSV infection hospitalised for at least 12 hours without comorbidities	
1b	Non-hospitalised children with confirmed RSV infection without comorbidities	
1c	Children hospitalised for at least 12 hours with confirmed RSV infection with comorbidities that would exclude them from 1a or 1b	
1d	Non-hospitalised children with confirmed RSV infection with comorbidities that would exclude them from 1a or 1b	
2	Healthy controls	

Table 2.2.4.1 Patient groups within RESCEU case control study

2.2.5 Inclusion and exclusion criteria

Children may only be recruited to the study if all the following criteria are met and are excluded if any of the exclusion criteria are met.

Inclusion criteria
Parent or carer of infant is willing and able to give consent to take part in the study
Patient must be below the age of 12 months at the time of recruitment
Parent of infant must have a telephone
Inclusion criteria for recruits to group 1 only
Patients in hospital must have been hospitalised for <48 hours at the time of recruitment
Patients in community must have been experiencing symptoms for <96 hours
Patients must live close enough to study centre for home visit 6-8 weeks later to be possible
Exclusion criteria
Patient must not have a history of receipt for any medication for RSV (e.g ribavirin)
Patient must not have any previous exposure to investigational RSV vaccines or medications
Patient must not have received immunoglobulins or monoclonal antibodies (including palivizumab)
Patients must have no history of receipt of steroids or montelukast within 7 days of enrolment in the study

Table 2.2.5.1 Inclusion and exclusion criteria for disease cases in the RESCEU case control study

Exclusion criteria for healthy controls (group 2)
Patient must not have history of concurrent clinically significant medical illness (not directly attributable to RSV infection) including but not limited to, cardiovascular, respiratory, renal, gastrointestinal, haematology, neurology, endocrinology, immunology, musculoskeletal, oncological or congenital disorders, as judged by the investigator. Specifically excluded examples include, but are not limited to Immunosuppressed states, Bronchopulmonary dysplasia/chronic lung disease of infancy, Congenital heart disease or Down's syndrome
Patients must not be premature, defined by gestational age <37 weeks at birth

Table 2.2.5.2 Exclusion criteria for healthy controls in the RESCEU case control study

2.2.6 Recruitment sites

Participants were recruited to the infant case control study at hospital sites and outpatient clinics in the United Kingdom (Oxford University, Imperial college London), the Netherlands (University medical centre Utrecht) and Spain (SERGAS).

2.2.7 Recruitment

Potential participants were identified by clinical research staff in participating hospital sites or in hospital settings (outpatient clinics, paediatric wards, Paediatric Intensive Care Units, emergency departments) or by responding to a poster/advert/online advert placed in a public area depending on the group.

In the hospital setting: A member of the paediatric clinical research team identified potentially eligible participants and provided an explanation of the study. Written information was also provided to parents expressing an interest. If permission was received from the parents, the child's medical records were reviewed to confirm eligibility. Participants were also be identified by posters, website based advertising, email distribution lists, media, social media and other forms of advertising, all of which text was submitted to an NHS REC for approval and with appropriate local permission to display/send advertising material. Contact details of the site research staff were be displayed for potential participants to discuss the study further and ask any questions.

In the community setting participants were approached by the same methods as hospitalised patients, and all text was submitted also to an NHS REC for approval and with appropriate local permission to display/send advertising material. Contact details of the site research staff were displayed for potential participants to discuss the study further and ask any questions.

2.2.8 Screening and eligibility assessment

In hospital setting: medical records were screened to confirm eligibility for inclusion, either after full consent has been received or with verbal permission from the parent/guardian/legally authorised representative. Written consent was sought to perform a bedside test to assess for eligibility.

In community setting: potential participants were screened using an online screening tool or by telephone/email contact with research staff. An appointment was made to visit the potential participant at their home, consent was sought to perform a bedside test for

infection with RSV. If this test was negative the potential participant was not considered eligible for enrolment into the study.

2.2.9 Informed Consent

The parent/guardian/legally authorised representative of a patient being recruited personally signed and dated the latest approved version of the informed consent form before any study specific procedures (including determination of RSV status) were performed.

Written and verbal versions of the participant information and informed consent were presented to the parent/guardian/legally authorised representative detailing no less than: the exact nature of the study; what it will involve for the participant; the implications and constraints of the protocol; the known side effects and any risks involved in taking part. It will be clearly stated that the parent/guardian/legally authorised representative is free to withdraw the infant from the study at any time for any reason without prejudice to future care, without affecting their legal rights, and with no obligation to give the reason for withdrawal.

2.2.10 Assessments and sampling

To be performed either in the hospital setting or in the community by clinical research staff. Information will be collected on participant's medical history, clinical examination and any laboratory and radiological investigations that are undertaken as part of routine medical care and recorded on a CRF. A standardised respiratory clinical severity score grading the severity of the RSV disease will be calculated at enrolment. An ARTI diary will be completed for 2 weeks from enrolment (Group 1 only)

At enrolment sampling will be done as follows:

Group 1 only
Nasopharyngeal swab or nasopharyngeal aspirate [NPA] for bedside RSV test (one sample)
Respiratory viral sample (flocked swab or synthetic absorption matrix [SAM] swab) (one sample)
Respiratory viral sample (flocked swab or synthetic absorption matrix [SAM] swab) (one sample) daily until discharge for those infants admitted to hospital
If the child is intubated and ventilated and has a bronchoalveolar lavage/tracheal aspirate taken as part of routine clinical care, a sample will be taken for analysis (one sample)
Groups 1 and 2
Nasopharyngeal bacterial swab (one swab aliquoted into several samples)
Venous or capillary blood sample (maximum of ~0.8ml/kg)
Urine sample (3-5ml)
Stool sample (2ml)

Table 2.2.10.1 Samples taken from patients recruited to the study

Ideally, at least one hour should elapse between initial nasal sampling (for bedside RSV test or diagnostic RSV PCR) and any further respiratory samples being taken.

One week after enrolment (Group 2 only)

One week after enrolment parents of participants will be contacted to assess if their infant developed any illness since enrolment

Six weeks (convalescent time point) (Group 1 only)

At 7 weeks (+/- 1 week) after discharge from hospital/emergency department, or 7 weeks post onset of symptoms for participants not attending hospital, participants will have one follow up visit. This will be a home visit by the study staff. At this time the following assessments will be made:

Group 1 only
Questionnaire of recent respiratory health
Clinical respiratory examination based on ReSVinet score
Nasopharyngeal bacterial swab (one swab)
Respiratory viral sample (swab) (one sample)
Venous or capillary blood sample (maximum of ~0.8ml/kg)
Urine sample (3-5ml)
Stool sample (2ml)

Table 2.2.10.2 Samples taken at convalescent time point from patients recruited to the study

Infants in Groups 1 and 2 will be followed up at these further time points

This study is recruiting over a period of 2 RSV seasons, participants will be informed at the time of enrolment how long the follow up will be for the individual participants.

One year of age
Assessment of respiratory symptoms by parental questionnaire at age 1 year through telephone contact/email/online
Cost, resource use and HRQoL data will be collected using questionnaires for caregivers
Two years of age
Assessment of respiratory symptoms by parental questionnaire at age 2 years through telephone contact/email/online
Cost, resource use and HRQoL data will be collected using questionnaires for caregivers Three years of age
Three years of age
Assessment of respiratory symptoms by parental questionnaire at age 3 years through telephone contact/email/online
Cost, resource use and HRQoL data will be collected using questionnaires for caregivers

Table 2.2.10.3 Information taken on study participants after completion of the study

2.2.11 Patient testing with Alere I point of care test

After consent was taken for participation in the study, patients were swabbed with a respiratory swab, and the swab placed in the Alere I point of care test to determine if there is any measurable amount of viral RNA. For patients being recruited to group 1, if the point of care test returns a positive result the patient is recruited into the study. For patients in group 2 if the point of care test returns a negative result the patient is recruited into the study.

2.2.12 Withdrawal from study

Each participant had the right to withdraw from the study at any time. In addition, the Investigator could discontinue a participant from the study at any time if the Investigator considered it necessary for any reason including:

Reasons for withdrawal
Ineligibility (either arising during the study or retrospectively having been overlooked at screening)
Significant protocol deviation
Withdrawal of Consent
Loss to follow up

Table 2.2.12.1 Potential reasons for withdrawal from the study

If the parent/guardian/legally authorised representative withdraws consent, we will analyse any previously collected sample and include their data in further analyses unless they state otherwise.

Withdrawn participants will not be replaced but we have factored a 10% drop out rate in our sample size.

The reason for withdrawal will be recorded in the case report form (CRF).

2.2.13 Reporting and storage of clinical information

Patient information and clinical information is recorded and stored in a CRF. The CRF is a document that describes the state of the participant at the time of admission to the study and functions as the record of the participant state at the time of recruitment. The CRF is filled out upon agreement of the legal guardian of the participant to admission to the study. If, during the completion of the CRF it is revealed that the participant does not meet the criteria of the study the participant is not recruited. The CRF also functions as a formal record of the participants consent for biobank usage of their samples and patient data. Full description of information recorded in the CRF is described in figure 2.2.11.1

CRF field	Description
Date of visit	Date at which participant was admitted to the study or date of convalescent visit
Informed consent	A formal record of the details of consent given to participate in the study.
Inclusion and exclusion criteria	Checklist of
Biobank consent	Formal record of consent given for biobank storage of samples
Medical history	Record of hospital attendance within first year of life, record of co-morbidities and regular medication required by participant
Demographic data	Record of participant date of birth, sex, birth weight, current weight, gestational age, and delivery method
Sample collection	Record of all samples taken as part of the study, time and date of sampling, method of sampling and amount of tissue sampled (where relevant)
Symptoms	Record of all relevant symptoms experienced by the participant at the time of recruitment, including fever, shortness of breath, coughing/wheezing, feeding intolerance, apnea and general condition. Time of onset of

	symptoms and end of symptoms is also recorded.
Treatment received	Record of any treatments prescribed for the current illness.
Further respiratory symptoms after discharge	For convalescent section of CRF. Record of any further respiratory symptoms experienced by the participant are recorded, including any episodes of coughing, wheezing, breathing difficulty or runny nose in subsequent months.
Further medical advice sought after discharge	Record of any further medical advice sought since recruitment to the study for respiratory symptoms and the level of medical advice sought (i.e sought advice from GP, hospitalisation etc)

Table 2.2.13.1 Fields in study CRF and description of information recorded in each CRF field

2.3 Sample collection and processing

Only processing and preparation of sample types that contribute to data used in this study are described in this section.

2.3.1 Whole blood collection and storage

1ml of blood was collected in 3ml tubes containing 1.4ml of Paxgene Blood RNA solution for transcriptomic analysis. The tubes were inverted 10 times to ensure a thorough mix of PAXgene solution and blood. Tubes were incubated at room temperature for up to two hours before transfer to -80 for storage.

2.3.2 RNA extraction from whole blood

Paxgene RNA blood tubes were defrosted and incubated at room temperature for 2 hours to ensure full lysis of blood cells. Paxgene blood tubes were centrifuged at between 3000 and 5000 g_0 for 10 minutes. Supernatant was removed by decanting. 4 ml of RNAse free water was added to the pellet and the tube sealed. Tube was vortexed until pellet was visibly dissolved. Tube was vortexed again at between 3000 and 5000 $g_0$for 10 minutes and the supernatant was discarded.

350 µl of resuspension buffer BR1 was added to each tube and vortexed until pellet was visibly resolved. Samples were transferred to 1.5 ml microcentrifuge tubes. 300 µl of binding buffer BR2 and 40 µlof proteinase K were added to each sample. Samples were mixed by vortexing for 5 seconds, followed by incubation at 55^0C for 10 minutes in a shaker incubator at 400-1400 RPM.

Lysed sample was then pipetted into Paxgene shredder spin columns placed in a 2 ml collection tube and spun at 17000 g_0 in a desktop centrifuge. Flow through was transferred to a fresh 1.5 ml microcentrifuge tube. 350ul of pure ethanol (96-100%) was added to each sample, mixed by vortexing and briefly centrifuged at 500 g_0 to ensure all drops collected.

700 µl of sample was placed in a Paxgene RNA spin column in a 2 ml processing tube and centrifuged for 1 minute at 17000 g_0. After centrifuging, the processing tube and flow through were discarded and the spin column placed in a new 2 ml processing tube. Remaining sample was pipetted into the spin column and centrifuged again at 17000 g_0 for 1 minute. Processing tube containing flow through was discarded and spin column was placed in a new 2 ml processing tube. 350 µl of wash buffer BR3 was pipetted into the spin column and spin column was centrifuged for 1 minute at 17000 g_0. Processing tube containing the flow through was discarded and spin column was placed in a new 2 ml processing tube. To digest any remaining DNA in the sample a mixture of 10 µl of DNAse I was mixed with 70 µl of buffer RDA for each sample to ensure optimal digestion activity of DNAse I and gently mixed by flicking the side of the tube. 80 µl of this DNAse digestion mix was added to each spin column and incubated at room temperature for 15 minutes.

An additional 350 µl of wash buffer BR3 was pipetted into the spin column and spin column was centrifuged for 1 minute at 17000 g_0. Processing tube containing the flow through was discarded and spin column was placed in a new 2 ml processing tube. 500 µl of wash buffer BR4 was pipetted into the spin column and centrifuged for 1 minute at 17000 g_0. Old processing tube containing flow through was discarded, and an additional 500 µl of wash buffer BR4 was added to the spin column and centrifuged for 3 minutes at 17000 g_0. Old processing tube containing flow through was discarded and the spin column was added to a new processing tube. The empty spin column was centrifuged for 1 minute at 17000 g_0 to dry the membrane of the spin column. Processing tube containing flow through was discarded.

Spin column was added to a 1.5 ml microcentrifuge tube for collection. To elute RNA from the membrane of the spin column, 40 µl of elution buffer BR5 was added to the spin column, and care was taken to ensure an even distribution of buffer across the membrane. Spin columns were centrifuged at 17000 g_0 for 1 minute to elute RNA. Without removing spin column from the microcentrifuge tube, an additional 40 µl of elution buffer BR5 was added to the membrane of the spin column and centrifuged at 17000 g_0 for 1 minute. Spin column was discarded, and microcentrifuge tube containing the elutate was incubated at 65^0C for 5 minutes in a shaker incubator. RNA was then immediately stored at -80^0C for future analysis by microarray.

2.4 Microarray preparation

2.4.1 cDNA preparation

In order to prepare cDNA for microarray analysis, a single strand must be synthesised off of the template RNA in sample. First strand master mix was prepared by mixture of 4 µl of the pico first strand buffer with 1 µl of the pico first strand enzyme for each sample to be run on microarrays. Reagents were mixed thoroughly before transfer to reaction plate. 5 µl of total RNA was added to each well containing the first strand pico master mix and mixed thoroughly.

Mixtures were incubated at 25 ^{o}C for five minutes, 42 ^{o}C for 60 minutes and then for two minutes at 4 ^{o}C using a thermocycler. Samples were centrifuged to return any precipitate to the mixture. 2ul of Pico clean up reagent was added to each well containing to reaction mixture and mixed thoroughly in order to clean up the RNA remaining in the sample. Mixture was Incubated for 30 minutes at 37 °C, for 10 minutes at 80 °C, and then for 2 minutes at 4 °C. After incubation mixture was centrifuged to bring precipitate back into reaction mixture.

3’ adaptor mix was prepared to amplify the single stranded cDNA in the mixture. 7 µl of Pico adaptor buffer and 1 µl of 3’ adaptor enzyme for each sample were mixed together on ice and vortexed to ensure mixture. 8ul of 3’ adaptor mix was added to each reaction well containing cDNA. Mixture was vortexed and lightly centrifuged. Mixture was Incubated for 15 minutes at 15 °C, for 15 minutes at 35 °C, for 10 minutes at 70 °C, and then for 2 minutes

at 4 °C using a thermocycler. After thermocycling, samples were vortexed and centrifuged briefly.

After synthesis of adaptor cDNA, the cDNA is them used to synthesise double stranded cDNA out of single stranded cDNA. The amplification mix was pre-prepared using 29 µl of Pico PCR buffer with 1 µl of Pico PCR enzyme for each sample to be run on the arrays. Amplification mix was briefly vortexed to mix and added to the well containing the sample mixture and was vortexed and centrifuged to mix. Samples were incubated for 2 minutes at 95 °C, for 6, 9 or 12 cycles of 30 seconds at 94 °C and 5 minutes at 70 °C, and then for 2 minutes at 4 °C using a thermocycler. After thermocycling, samples were vortexed and centrifuged briefly.

Next, cDNA was used to make cRNA by the in vitro transcription (IVT) method. IVT master mix was created by adding 24 µl of Pico IVT buffer to 6ul of Pico IVT Enzyme for each sample to be ran on the array and vortexed to mix. 30 µl of the IVT master mix was added to the sample mixture and was Incubate for 14 hours at 40 °C, and then at 4 °C. Samples were centrifuged after incubation to bring condensation back into solution.

cRNA was then purified using purification beads to remove, enzymes, salts, inorganic phosphates, and unincorporated nucleotides from the mixture, to prepare the cRNA for 2nd-cycle double-stranded-cDNA synthesis. 140 µl of purification beads for each sample to be run on the array were added to fresh plate wells after vortexing to resuspend magnetic particles. 80 µl of cRNA mixture was added to each well and mixed with the purification beads by pipetting. cRNA and purification beads were incubated together for 10 minutes to ensure binding. Purification beads were captured using a magnetic stand for five minutes. Once purification beads are assembled on magnet, supernatant was aspirated and discarded. While on magnet stand, 80% ethanol was added to each well and incubated for 30 seconds to was the beads, aspirated and discarded three times. Samples were air dried on the magnetic beads for five minutes.

To elute the purified cRNA the plate was removed from the magnetic and 27 µl of a preheated (65 ^{o}C) nuclease free water was added to each sample and incubated for one minute and mixed by pipetting. Plate was moved to magnetic stand for 5 minutes to capture purification beads and supernatant (now containing the cRNA) was transferred to nuclease

free tubes. cRNA yields and quality were assessed after purification by bioanalyzer analyses to ensure no samples contained excessively degraded RNA.

cRNA was then used to make 2nd-cycle double-stranded-cDNA. Synthesis of 2nd-cycle double-stranded-cDNA is required to ensure that cDNA exists for both the sense and anti-sense strands. 20 µg of cRNA was prepared by dilution of cRNA to a concentration of 833 ng/µL, and 24 µl of this was removed for a total of 20 µg of cRNA. The 2nd-Cycle ds-cDNA Master Mix was prepared by adding 4ul of Pico 2nd-cycle primers, 8 µl of Pico 2nd-Cycle ds-cDNA Buffer and 4ul of Pico 2nd-Cycle ds-cDNA Enzyme per sample to be run on the array, mixed thoroughly and ran mixed with the 24 µl of cRNA. Samples were incubated for 10 minutes at 25 °C, 50 minutes at 42 °C, for 10 minutes at 70 °C, and then for 2 minutes at 4 °C using a thermocycler. Samples were centrifuged after incubation to bring condensation back into solution.

RNA was removed from the mixture by hydrolysis using the Pico Hydrolysis buffer. 7 µl of the Pico hydrolysis buffer was added to the sample mixture and incubated for 20 minutes at 65 °C, and then for at least 2 minutes at 4 °C. Samples were centrifuged to return condensation to reaction mixture and 10ul of Pico neutralization buffer was added to each sample and mixed thoroughly by vortexing.

cDNA was then purified using purification beads to remove, enzymes, salts, inorganic phosphates, and unincorporated nucleotides from the mixture, to prepare the cDNA for labelling and fragmentation. 140 µl of purification beads for each sample to be run on the array were added to fresh plate wells after vortexing to resuspend magnetic particles. 80 µl of cDNA mixture was added to each well and mixed with the purification beads by pipetting. CDNA and purification beads were incubated together for 10 minutes to ensure binding. Purification beads were captured using a magnetic stand for five minutes. Once purification beads are assembled on magnet, supernatant was aspirated and discarded. While on magnet stand, 80% ethanol was added to each well and incubated for 30 seconds to was the beads, aspirated and discarded three times. Samples were air dried on the magnetic beads for five minutes.

To elute the purified cDNA the plate was removed from the magnetic and 27 µl of a preheated (65 ^{o}C) nuclease free water was added to each sample and incubated for one

minute and mixed by pipetting. Plate was moved to magnetic stand for 5 minutes to capture purification beads and supernatant (now containing the cDNA) was transferred to nuclease free tubes. cDNA yields and quality were assessed after purification by bioanalyzer analyses.

Before running on a microarray DNA must be fragmented to ensure that DNA is appropriately sized to hybridise to array probes and is labelled to facilitate fluorescence. Fragmentation is performed using uracil-DNA glycosylase (UDG) and apurinic/apyrimidinic endonuclease 1 (APE 1). DNA was labelled with terminal deoxynucleotidyl transferase (TdT) using the proprietary DNA labelling reagent that is covalently linked to biotin.

Double stranded cDNA was made up to a concentration of 250 ng/µl in 22 µl of nuclease free water. 6 µl of the fragmentation and labelling buffer was added to 2 µl of the fragmentation and labelling enzyme for each sample to be run on the array and mixed with the cDNA in a nuclease free tube. Samples were incubated for 1 hour at 37 °C, for 2 minutes at 93 °C, and then for 2 minutes at 4 °C.

2.4.2 Array Hybridisation

Hybridisation master mix was prepared by mixing 2.5 µl of Control Oligonucleotide B2 (3 nM), 7.5ul of 20X Hybridization Controls, 2 µl of Control Oligo B2, 40 µl of 1.3X Hybridization Solution A, and 70 µl of 1.3X Hybridization Solution B for each array. 120 µl of Pre-Hybridization Buffer was pipetted into the appropriate wells of the hybridization tray. The array strip was placed in the hybridization tray. The hybridisation tray was placed in the Hybridisation station for 10 minutes at 45 ^{o}C. The hybridisation cocktail of the sample and the reaction mix was prepared by mixing 100 µl of Hybridization Master Mix, and 20 µl of Fragmented and labelled ds-cDNA for each sample. Plate was sealed, vortexed and centrifuged to ensure the mixture collected at the bottom of the wells. The hybridisation mix was then incubated for 5 minutes at 95 °C, then for 5 minutes at 45 °C in a thermal cycler, then centrifuged to collect condensation.

After pre-hybridisation the array strips were removed from the hybridisation station and immersed in pre-hybridisation solution. 90 µl of pre-heated hybridization cocktail was added to the appropriate wells on a new hybridisation tray. Array strip was removed from the hybridisation tray and placed in the hybridisation tray containing the cocktail mixture. This tray was then placed in the GeneTitan hybridisation station. Samples were registered on the

GeneTitan system and array was scanned into the system. The hybridisation was run on the array.

2.5 Assessment of disease severity

In order to assess the relationship between gene expression and disease severity, metrics of disease severity had to be defined. As severe disease can be defined in many ways, it is useful for exploratory analyses to consider many different ways to define severe disease to allow for the development of robust biomarkers.

2.5.1 ReSVinet score

Clinical severity scores are a common method of quantifying disease severity in clinical studies. The ReSVinet clinical severity score is one such score that has been developed for assessing RSV disease(291). The ReSVinet score assigns a numerical value ranking the severity of a number of symptoms (including apnea, respiratory rate, temperature, requirement for supplemental oxygen) on a numerical scale ranging from 0 to 3. The score ranges from zero to 20, with a score of 14 or higher indicating “severe” RSV disease.

Item	0 points	1 point	2 points	3 points
1. Feeding intolerance	None	Mild	Partial	Total
2. Medical intervention	None	Basic	Intermediate	High
3. Respiratory difficulty	None	Mild	Moderate	Severe
4. Respiratory frequency	Normal	Mild/occasional tachypnea	Prolonged/recurrent tachypnea	Severely altered
5. Apnea	No			Yes
6. General condition	Normal	Mild	Moderate	Severe
7. Fever	None	Mild: <38.5^{o}C	Moderate: >38.5^{o}C	

Table 2.5.1.1 ReSVinet scoring system for RSV disease severity, points are allocated for most severe form symptom being experienced by the patient. ReSVinet score is calculated as the sum of the score of the patient in each category. A score of 14 is considered severe disease.

2.5.2 Requirement for respiratory support

Another way of measuring RSV disease severity that may be meaningful for biomarker development is the is the requirement for respiratory support. Respiratory support entails some addition of supplementary oxygen, and ranges from lowly invasive methods like low flow canula, up to mechanical ventilation. Data on degree of respiratory support were recorded in the trial and ranked from least severe to most severe and given a numeric score for analyses (table 2.4.2). A score of above 3 or above was considered as a threshold for severe disease for analyses and was described as requirement for invasive respiratory support. This scoring method was used as the primary scoring method in biomarker discovery analyses, due to its use as part of the GRSS score(286) and its frequent use in studies of RSV severity(292)(276)(278)(283)(284)(238)(277)(279)(285), allowing for the highest degree of comparability with other studies.

Respiratory support type	No respiratory support	O2 > 21%	Low flow nasal cannula	CPAP/BiPAP	High flow nasal cannula	Invasive ventilation
Score	0	1	2	3	4	5

Table 2.5.2.1 Scoring system of respiratory support methods used in treatment of RSV used for analysis

2.5.3 Requirement for hospitalization

Another meaningful health outcome for biomarker design is hospitalization. For the purpose of biomarker analysis hospitalized patients were considered to have severe disease, whereas patients recruited in community visits were considered to have mild disease.

Chapter 3: Transcriptomic response to RSV infection in infants

3.1 Introduction

This chapter, as well as chapters 4 and 5 will present an analysis of the host transcriptomic response to RSV infection.

More severe RSV disease is associated with specific host immune responses and secretion of inflammatory molecules in the respiratory tract(146)(60)(125). These host responses are signalled and regulated by changes in the host transcriptomic response(285). Changes in host transcriptome can also be measured precisely and comprehensively for a given tissue using modern gene expression technologies such as microarrays and RNA-seq(333)(334). This capacity to observe changes in tens of thousands of variables simultaneously makes transcriptomic analysis an extremely valuable tool both in understanding disease and in the development of biomarkers. In this study microarrays are used to analyse changes in gene expression. A DNA microarray is a chip lined with probes corresponding to the sequences of human genes(335). These probes are capable of hybridising when presented with a complementary DNA sequence and cause the cDNA to fluoresce when hybridised. The probes correspond to genes or to different isoforms of genes. As multiple copies of each probe are contained on the array, it is possible to compare the relative expression of genes by the intensity of their fluorescence on the plate. Within the current study this technology is used to analyse the differences in gene expression between children infected with RSV and those same children in convalescence.

In this chapter I aim to identify the gene expression changes that accompany acute RSV infection and link those gene expression changes to the immune response against RSV.

To accomplish these aims I will first perform quality control of the data to control for the effects technical errors may have on gene expression. Following this, I will demonstrate the function of genes associated with RSV infection and show how these changes in gene expression are likely reflected in changes in blood cellular composition.

3.2 Methods

RNA from 102 of children with a confirmed RSV infection recruited to the infant case control study (chapter 2.2.1.2) was extracted (chapter 2.2.2), along with RNA from 101 blood samples from the same children in convalescence and RNA from ten healthy control patients. RNA was extracted from 2ml of whole blood stored in PAXgene solution. RNA quality and quantity was assessed using a bioanalyzer, and made up to a concentration of 100ng/µl in nuclease free water. RNA was used to synthesise cDNA, which was subsequently hybridised to an Affymetrix microarray (chapter 2.4) and probe fluorescence assessed by visualisation using a GeneTitan machine.

3.2.1 Demographic information for samples used in transcriptomic analysis

Full description of all participants included in the study, and power calculations used to determine appropriate sample sizes for analyses are discussed in chapter 2.2.2. Samples available for transcriptomic analysis at the time the samples were run, alongside patient characteristics before and after QC are shown in table 3.2.1. Breakdown of full recruitment to the case control study for each site is shown in table 3.2.2.

	Oxford	Utrecht	Healthy control	Acute RSV	Convalescent	Average age	Percentage hospitalised	Premature	Req supplemental Oxygen
Pre QC	77	136	10	102	101	132 days	70%	.05%	60%
Post QC	73	130	7	98	98	132 days	70%	.05%	60%

Table 3.2.1 Breakdown of demographic information for all patients included in microarray analysis, with sites of recruitment, age, percentages of patients hospitalised, required supplemental oxygen and prematurity shown.

Site	RSV cases (healthy, n=250)	RSV Controls (healthy, n=250)	RSV cases (comorbid n=25)	RSV Controls (comorbid n=25)	Healthy controls (n=80)
Utrecht	35/63 (56%)	13/63 (21%)	2/6 (33%)	1/6 (13%)	0/20 (0%)
Oxford	3/63 (5%)	0/63 (0%)	3/6 (50%)	0/6 (0%)	0/20 (0%)
Imperial	0/63 (0%)	0/63 (0%)	0/6 (0%)	0/6 (0%)	0/20 (0%)
Spain	0/63 (0%)	0/63 (0%)	0/6 (0%)	0/6 (0%)	0/20 (0%)
Total	38/250 (15%)	13/250 (5%)	5/25 (20%)	1/25 (4%)	0/80 (0%)

Table 3.2.2 Breakdown of participants recruited to the RESCEU case control study for each participating study site, with the total planned numbers for each site given, alongside achieved numbers and the proportion of the projected total numbers accounted for by achieved numbers.

3.2.2 Removal of outliers

The distributions of probe intensities for each were assessed using a density plot in R, and the samples that had a significantly different distribution of probe intensities according to Fishers F test(336) were removed from the analysis before progressing to further quality control steps. Outliers were detected using the arrayQualitymetrics package (version 3.1.4) in R(337).

3.2.3 Gene filtering

Probes with zero variation in a large proportion of samples can create statistical noise that makes it more difficult to understand real sources of biological variation in the data set. To control this noise, probes with zero variation in over half of all samples were removed from further analyses. The package Genefilter in R was used to filter low variation probes(338).

Expression of genes found on the sex chromosomes X and Y act differently than those found on other chromosomes, as their expression is dependent on conditions that are independent of the experimental conditions. Because of this expression of genes on the sex chromosomes do not fit the assumptions of models used to determine fold change of gene expression between states. These genes were removed using probe annotation data for the microarray chip.

3.2.4 Normalization

To ensure comparability between samples, normalization is employed to ensure that probe intensities are centred around a common mean and fit a similar scale(339). Normalization was performed in R, two methods were explored to ensure robustness of the final chosen method. Background correction of probe intensities was applied before normalization using the rma function in R to correct for background fluorescence on the plate from neighbouring probes(340). Quantile normalization using the robust multi-array average (RMA) method (340) in the oligo package (version 1.14.0) (341) and robust splines normalization (rsn) using the lumi package (version 2.4.0)(342) were applied and quality control output metrics measured.

3.2.5 Identification of differentially expressed probes on infection

Linear regression was performed with an empirical Bayes model on expression values of probes of healthy controls and patients with confirmed RSV infection, and those same patients 6-8 weeks later using the “lmfit” and “eBayes” functions of the R package limma (version 3.44.1)(343). A Benjamini and Hochberg false discovery rate (FDR) adjusted p-value of less than .05 was used as a threshold for differential gene expression between patients with mild RSV disease, healthy controls and patients with more severe RSV disease.

3.2.6 Gene set enrichment analysis with GSEA

Gene set enrichment analysis is a method to interpret biological function from lists of genes. Gene enrichment uses annotation from publicly available databases (specifically msigdb(344)), and prepares a ranked list of the terms that are most highly enriched within a given gene set, alongside various statistics by which these enrichment terms may be ranked, including an expression score, p-values and number of genes per annotation term. GSEA is a gene enrichment analysis algorithm that can be used for these analyses. The pre-ranked GSEA algorithm (version 4.0)(345) was run on all genes on the microarray. The genes were ranked according to the t-statistic, which is a statistic that combines both statistical significance and change in expression. GSEA algorithm was run for biological process (GOBP) terms, cellular component (GOCC) terms, molecular function (GOMF) terms hallmark signature terms and immune database terms. Annotation terms were ranked by expression score and filtered by FDR, where a cut-off of .05 was applied.

3.2.7 Principal component analysis

Principal component analysis (PCA)(346) is a method for understanding the sources of variation in a data set. PCA uses an orthogonal transformation algorithm to reduce a large number of variables with a high degree of covariance into a smaller set of linearly uncorrelated variables. These components are ranked according to the amount of the total variation within the data set that they account for, and transformed onto a scalar projection upon which it is possible to visualise the extent to which individual variables contribute towards or are associated with a principal component axis. Samples can also be projected on these axes, based on their expression of a given component. This functionality allows for the visualisation of sources of variation within sample groups. PCA was run on all samples used in analysis after each quality control adjustment performed on the data using the mixOmics algorithm (version 6.12.1)(347) and visualisation software. Individuals were plotted upon principal components to determine batch effects and identify sources of variation within the data.

3.2.8 CIBERSORTx deconvolution of transcriptomics data

CIBERSORTx is an algorithm that deconvolutes transcriptomics data by comparison of the gene expression profile to gene expression profiles for selected cell types(348), in order to give an estimate of cell populations in the samples. The CIBERSORTx algorithm (version 1.05) was given a matrix of all genes contained on the microarray with gene symbols given in Human Genome Organisation (HUGO) format alongside the residuals of these genes after data were fit to a linear model to control for the effects of confounding variables and estimate the effect of disease status on gene expression. A reference file (LM22) provided to use with the CIBERSORTx algorithm was used as a cell type reference. The algorithm was run over 500 iterations and results plotted using the ggplot2 package (version 3.3.1)(349) in R.

3.2.9 ComBat batch correction

Batch correction was performed on data using the ComBat method(350). ComBat first standardises the mean and variance of each gene to be similar, then estimates the batch effects for the batches provided to the algorithm using a linear shrinkage model and corrects for these by adjusting expression value of every probe. Microarray batches were recorded during microarray preparation for use in batch correction.

3.2.10 PVCA (Principal Variance Component Analysis)

PVCA is a method that allows for quantification of proportion of sources of variance in a data set as a percentage of the total variation within the data set. PVCA analysis was performed using the pvca package in R, version 1.28.0(351).

3.3 Quality control of microarray data reveals signal associated with RSV infection

3.3.1 Low quality samples removed from analysis

All available transcriptomics samples at the time of running (described in table 3.2.1) were ran on microarray and visualised using the GeneTitan machine as described in chapter 2.4. 213 samples in total were run, including 10 healthy controls, 102 samples from participants at the time of acute infection and 101 from participants in convalescence. Distributions of probe intensities for each sample were plotted in R to determine the existence of outliers with incomparable distributions of probe intensities within the data set (fig 3.3.1.1). A total of ten samples were identified that did not fit the distribution of probe intensities shared by the majority of the samples, and these samples were removed from further analyses. Outliers were defined as having a significantly different distribution of probe intensities according to in the arrayqualitymetrics package version 3.4.1 in R(337).

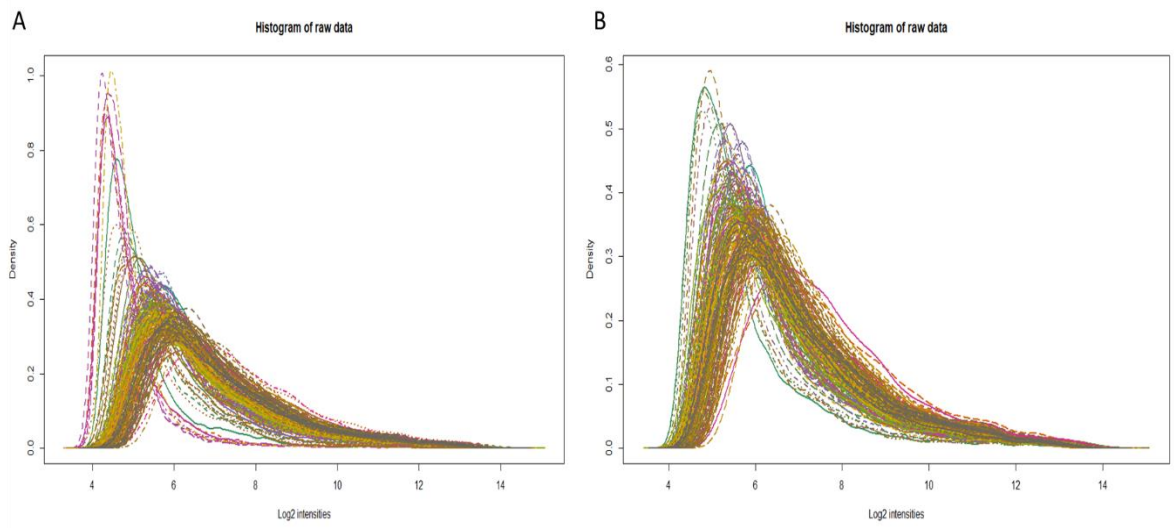


Figure 3.3.1.1 Distribution of array probe intensities before and after removal of outliers with significantly different distribution of probe intensities across the whole array. The average distribution of log2 probe intensities observed across each sample on each array is shown for before and after removal of outliers for all array probes in all participants included in microarray analyses

3.3.2 The majority of variation between samples prior to quality control is attributable to technical variation

After removal of outlier samples, 203 samples remained, of which 7 were from healthy control participants, 98 were from participants at the time of acute infection, and 98 were from participants during convalescence. Probe intensities of each sample across each array were visualised (fig 3.3.2.1). Visualisation of probe intensities aids in the detection of outliers and indicators of noise in the data. Principal component analysis was also performed on the raw data prior to any further quality control steps to determine the potential systematic effects introduced due to experimental design prior to correcting for these sources of bias. Principal component and probe intensity analyses after implementation of quality control steps is a useful tool in determining the effects of quality control measures on data.

Principal component analysis (as described in 3.2.7) of raw data prior to normalization and filtration found the first principal component to account for 60% of the variance in the data set (fig 3.3.2.2), with the second principal component account for 5% of the variance in the data set, while the third principal component accounted for 3%. Principal component analyses were run using only the 98 samples from participants at the time of infection, and the 98 samples from participants in convalescence. Healthy controls were excluded. Samples formed two distinct clusters on the first and second principal components (fig 3.3.2.3). This clustering is attributable to batch effects from microarray plates (fig 3.3.2.3 A) and cDNA yields (fig 3.3.2.3 B). Samples separate based on RSV infection status on the third principal component (fig 3.3.2.3 C). PCA mappings on the second and third components for batch and cDNA and the first and second components for disease status are shown in appendix 1.1.1.

PVCA analysis (described in 3.2.10) of results (fig 3.3.2.4) indicated that prior to any quality control the largest sources of variation in the data set were cDNA quantity (25%), microarray batch (17.5) and age of participant (20%). As large sources of technical variation exist within the data set normalization and batch correction are required.

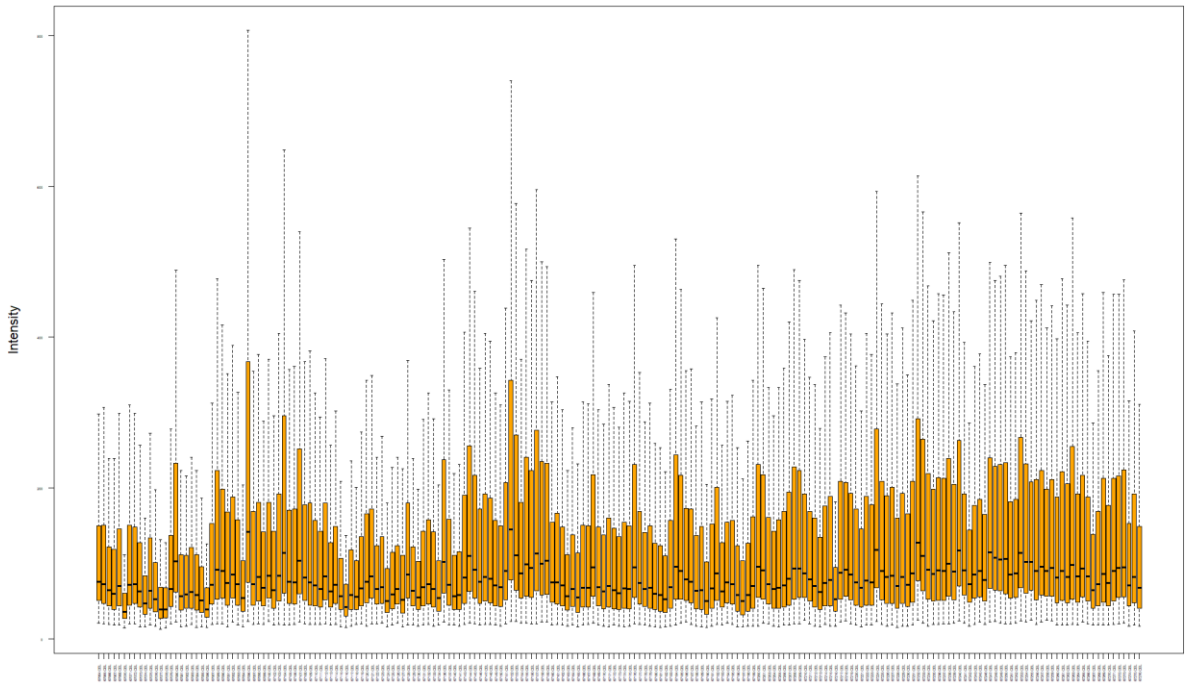


Figure 3.3.2.1 Boxplot of raw array probe intensities for each sample included in microarray analysis after removal of outliers and prior to any other QC with mean of all probe intensities for each sample, the distribution of average probe intensities for each replicated probe and the distribution of probe intensities for all probes shown. Range of probe intensities for each sample are shown. Black bar indicates the mean of all probe intensities for that sample, orange bar indicates range of standardised (average across all replicate probes) probe intensities, error bars indicate standard error of probe intensity values, across all replicates for each probe.

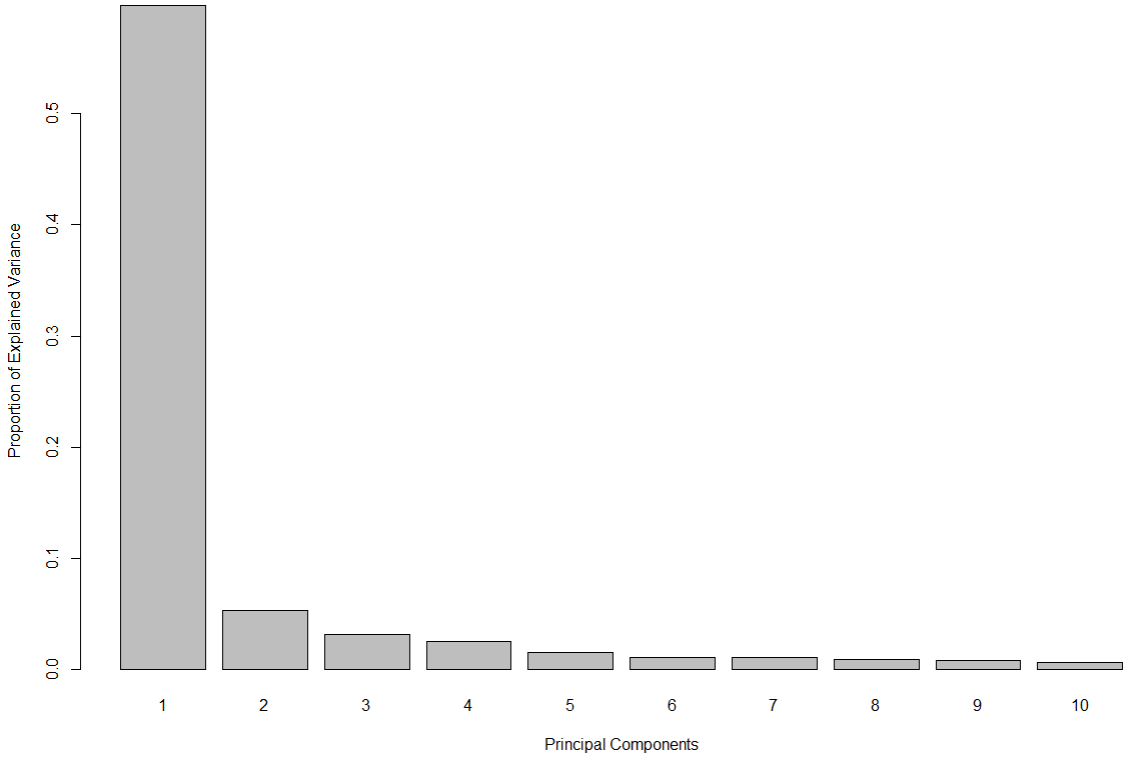
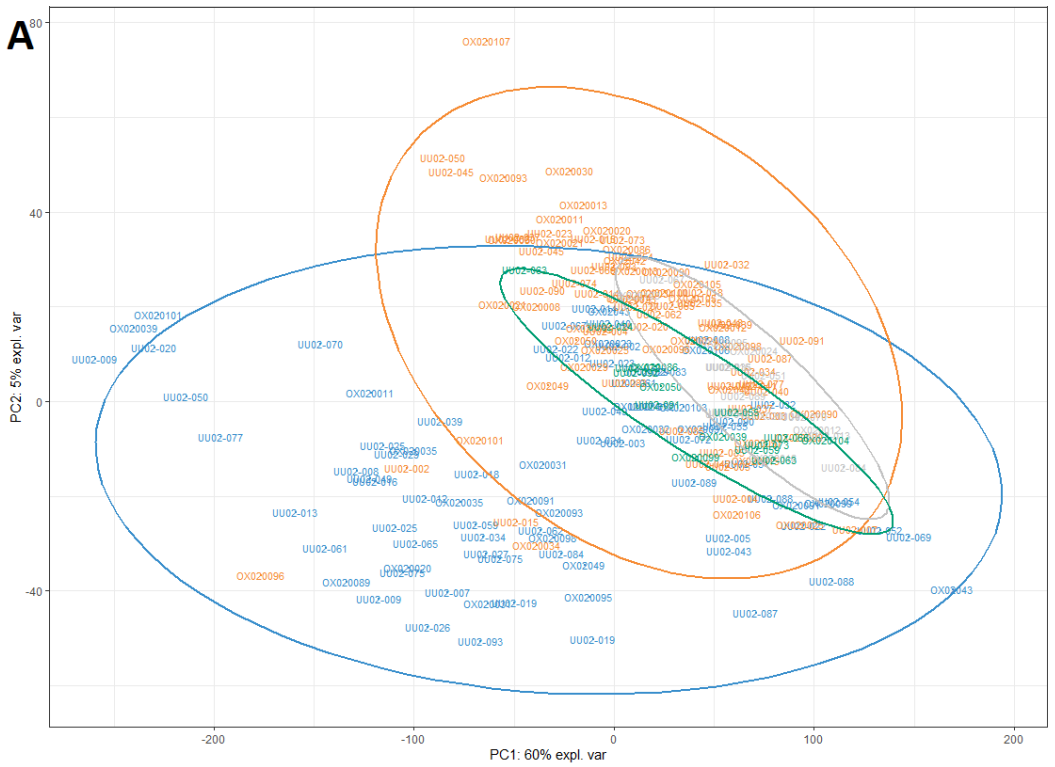
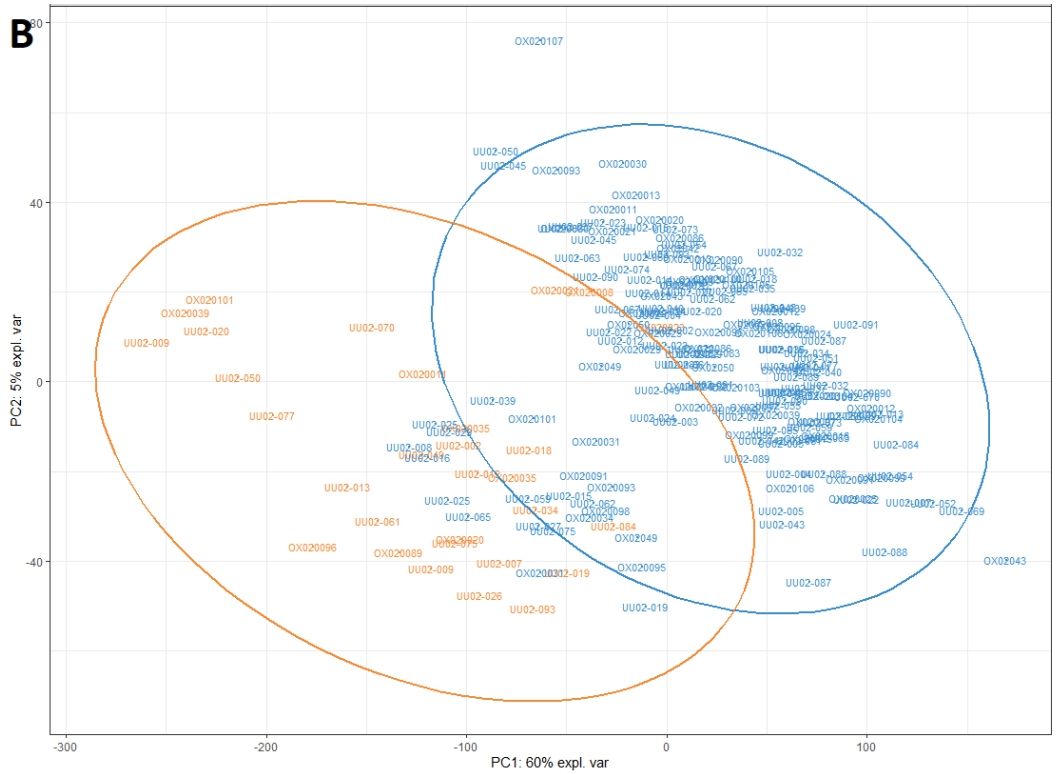


Figure 3.3.2.2 Scree plot of loadings of principal components in raw data showing proportion of variance attributable to each principal component

The percentage of the total variance in the data set represented by each principal component is shown for the first ten principal components. Each principal component represents a distinct grouping of co-correlated probes that do not correlate with other groupings co-correlated probes.


A
PC2: 5% expl. var
PC1: 60% expl. var
Batch
1
2
3
4


B
PC2: 5% expl. var
PC1: 60% expl. var
cDNA yield
High
Low

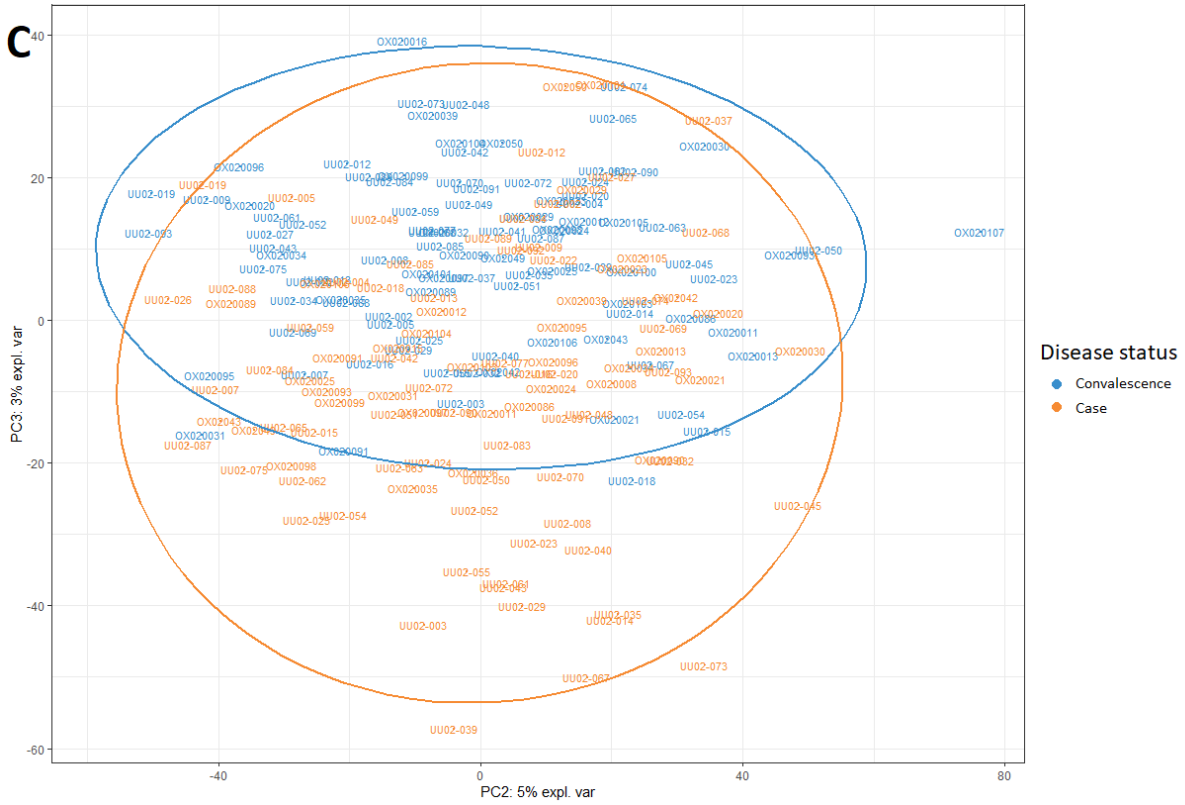


Figure 3.3.2.3 Principal component analysis of samples before quality control Grouping of samples according to expression of probes in the A: first and second principal components coloured by microarray batch, B: first and second principal components coloured by cDNA yield and C: second and third principal components coloured by disease status is shown. Expression of the first principal component is mapped on the x axis and expression of the second principal component is mapped on the y axis. Samples originating from patients recruited in Oxford are labelled with the study title "OX02", while the samples from patients recruited in Utrecht are labelled as "UU02". Samples are coloured according to infection status of the patient they were taken from.

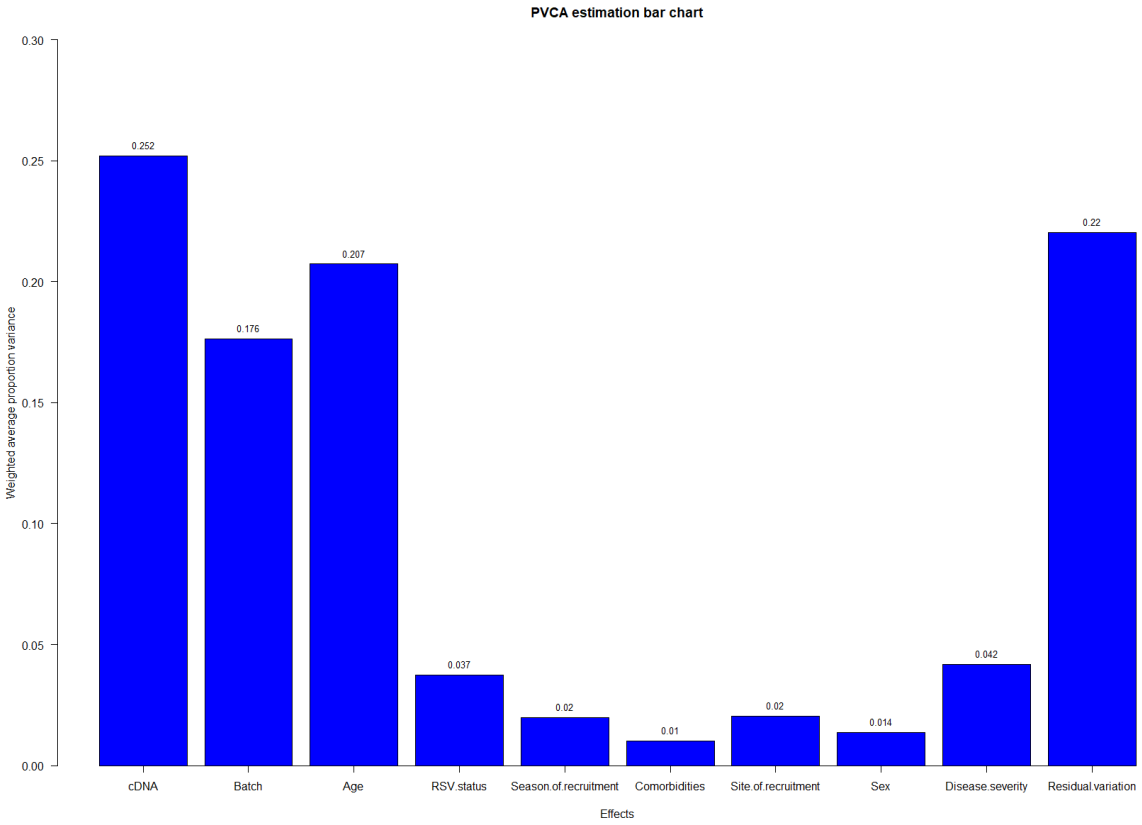


Figure 3.3.2.4 PVCA of raw data of all samples after removal of outliers showing the proportion of variation in the data set attributable to patient characteristics and technical effects, such as Batch and cDNA yields

3.3.3 Normalization of array data reveals distinct signal of RSV infection

To ensure comparability between samples quantile normalization(340) was used on the probe intensities as described in chapter 3.2.4, Quantile normalization aims to make the distributions of intensities for probes as similar to each other as possible by assigning each value in the distribution a ranking. The average of each probe of each discrete rank in each distribution and assigning that average value to every value of the same rank, ensuring the distributions have an identical scale and are centred around the same mean value, meaning any differences between the distribution are less likely to be attributable to systematic batch effects that may skew all the values in a distribution upwards or downwards. Identical values are assigned the mean of the assigned rank value for that value and the value of the

rank directly above the rank of the identical values. In addition to quantile normalization, genes found on sex chromosomes were filtered from the data set.

Principal component analyses were run using only the 98 samples from participants at the time of infection, and the 98 samples from participants in convalescence. Healthy controls were excluded. Mappings of samples on PCA according to expression of these genes on the sex chromosomes was performed to screen for mislabelled samples. Two distinct clusters formed based on expression of sex chromosome genes (fig 3.3.3.1), with no participants of one sex clustering with the others, suggesting that sex labelling for all samples was correct.

Probes with no variation between samples were also filtered from the data set. Filtration of both sex chromosome probes and probes with no variation between samples as described in chapter 3.2.3. This filtration resulted in a total of 929 probes being removed from analysis, resulting in a remaining total of 17706 probes out of an original 18635 gene coding probes.

A box plot of probe intensities and a principal component analysis were performed on the samples after the application of quantile normalization, sex filtration (described in chapter 3.2.3) and probe filtration (fig 3.3.3.2). Mean expression value of all samples was shifted to a to a shared mean, with a few samples displaying small deviations from the new mean.

PCA was also performed on all samples after normalization to show effects that normalization may have had on sources of variation that may exist within the data. After quantile normalization the first principal component accounted for 28% of the variance in the data set, the second principal component accounted for 6% of the variance in the data set (fig 3.3.3.3),, showing a large change from the pre normalization variance (fig 3.3.2.3), where the first principal component accounted for 60% of the variance in the data set.

Mapping of samples on the first three principal components was visualised for all samples after normalization (fig 3.3.3.4). Mappings on the first component after normalization shows remaining clustering according to microarray batch and cDNA yield (fig 3.3.3.4 A, B). However, when the second and third principal components are visualised (fig 3.3.3.4 C) there is separation between patient with disease and patients in convalescence, particularly along the second principal component axis, indicating that the variance observable on the second and third principal components may be attributable to RSV infection. Batch

correction will be employed in order to remove effects of microarray batch and cDNA yield from the data and ensure any remaining variation is meaningful biological variation. PCA mappings on the second and third components for batch and cDNA and the first and second components for disease status are shown in appendix 1.1.2.

PVCA analysis of results (fig 3.3.3.5) indicated that after quantile normalization the largest sources of variation in the data set were cDNA quantity (26%), microarray batch (17%) and age of participant (21%). As large sources of technical variation still exist after normalization, batch correction is required to reduce the effect of these on the data set.

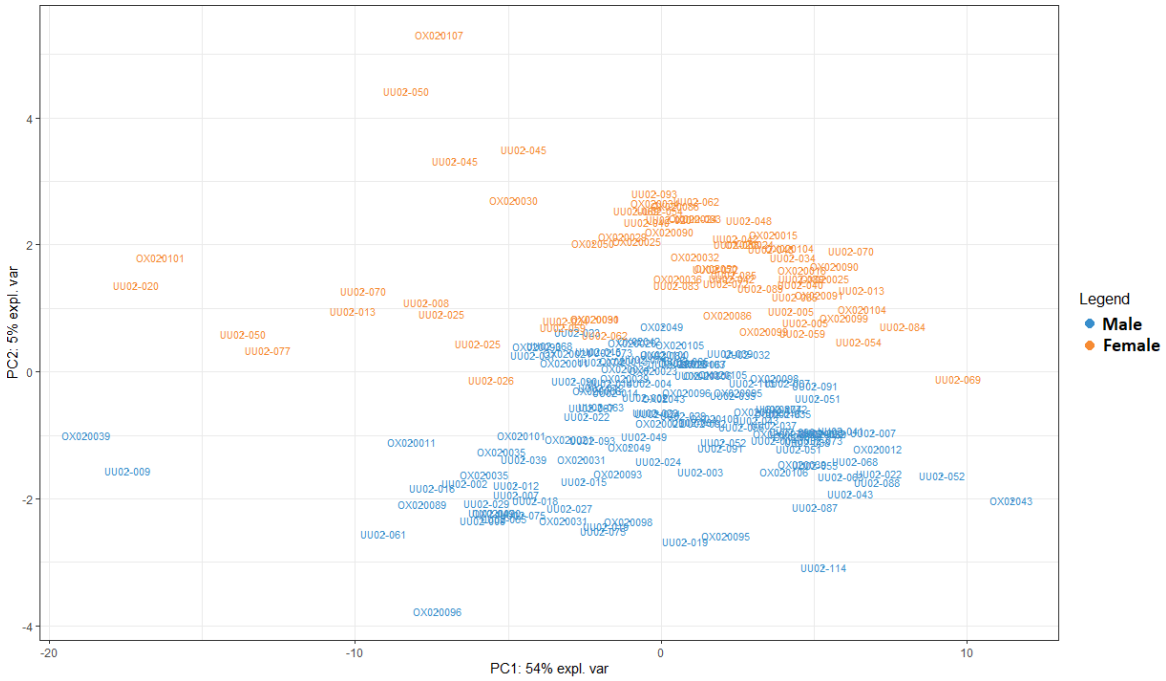


Figure 3.3.3.1 PCA mapping of samples from all patients based on expression of sex chromosome genes on the first and second principal components labelled for sex of participant and recruitment site. Grouping of samples according to expression of sex chromosome probes in the first and second principal components is shown. Expression of the first principal component is mapped on the x axis and expression of the second principal component is mapped on the y axis. Samples originating from patients recruited in Oxford are labelled with the study title "OX02", while the samples from patients recruited in Utrecht are labelled as "UU02". Samples are coloured according to infection status of the patient they were taken from. Samples taken from healthy patients are labelled in orange, samples taken from patients in convalescence are coloured blue, and samples taken from patients at the time of RSV infection are coloured grey.

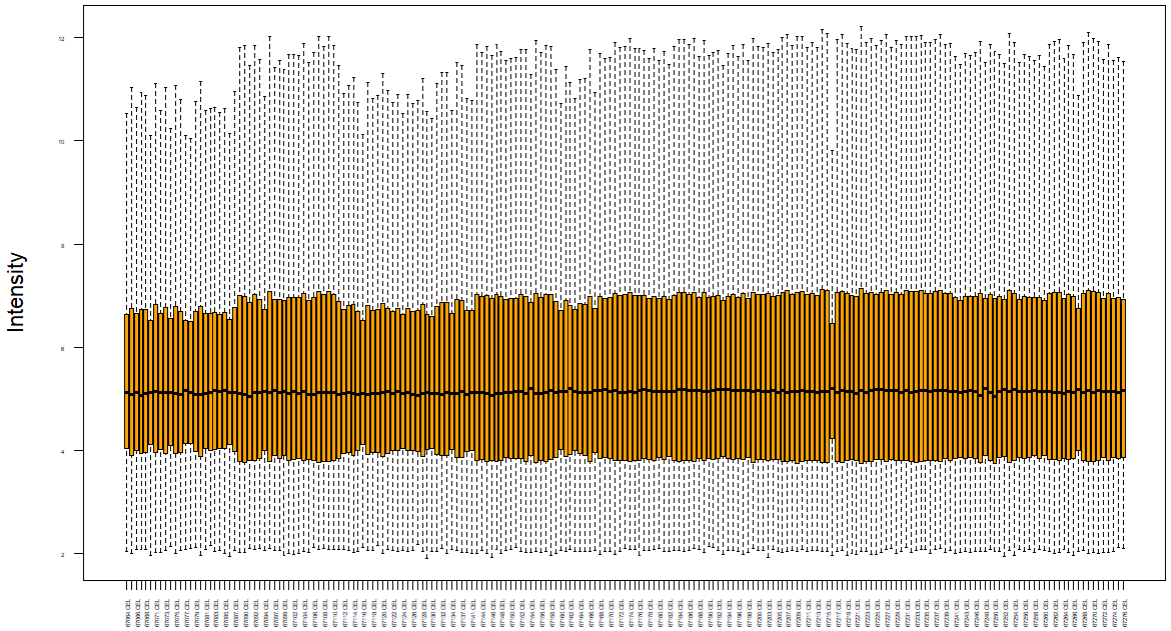


Figure 3.3.3.2 Boxplot of normalized array probe intensities for each sample included in microarray analysis after removal of outliers with mean of all probe intensities for each sample, the distribution of average probe intensities for each replicated probe and the distribution of probe intensities for all probes shown. Range of probe intensities for each sample are shown. Black bar indicates the mean of all probe intensities for that sample , orange bar indicates range of standardised (average across all replicate probes) probe intensities, error bars indicate standard error of probe intensity values, across all replicates for each probe.

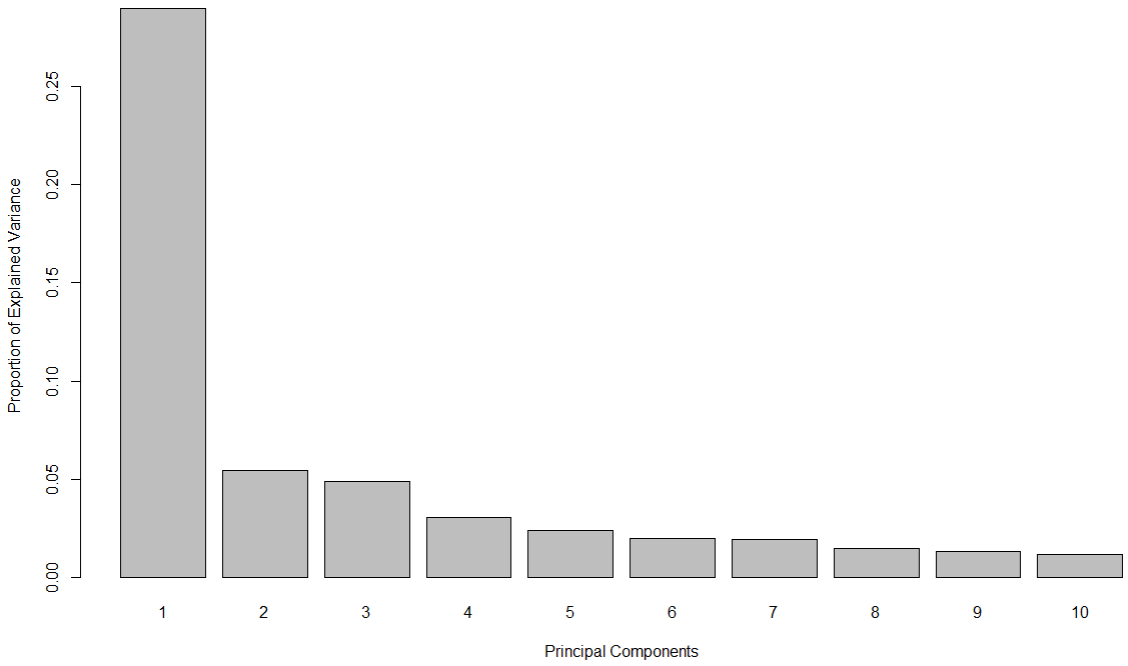
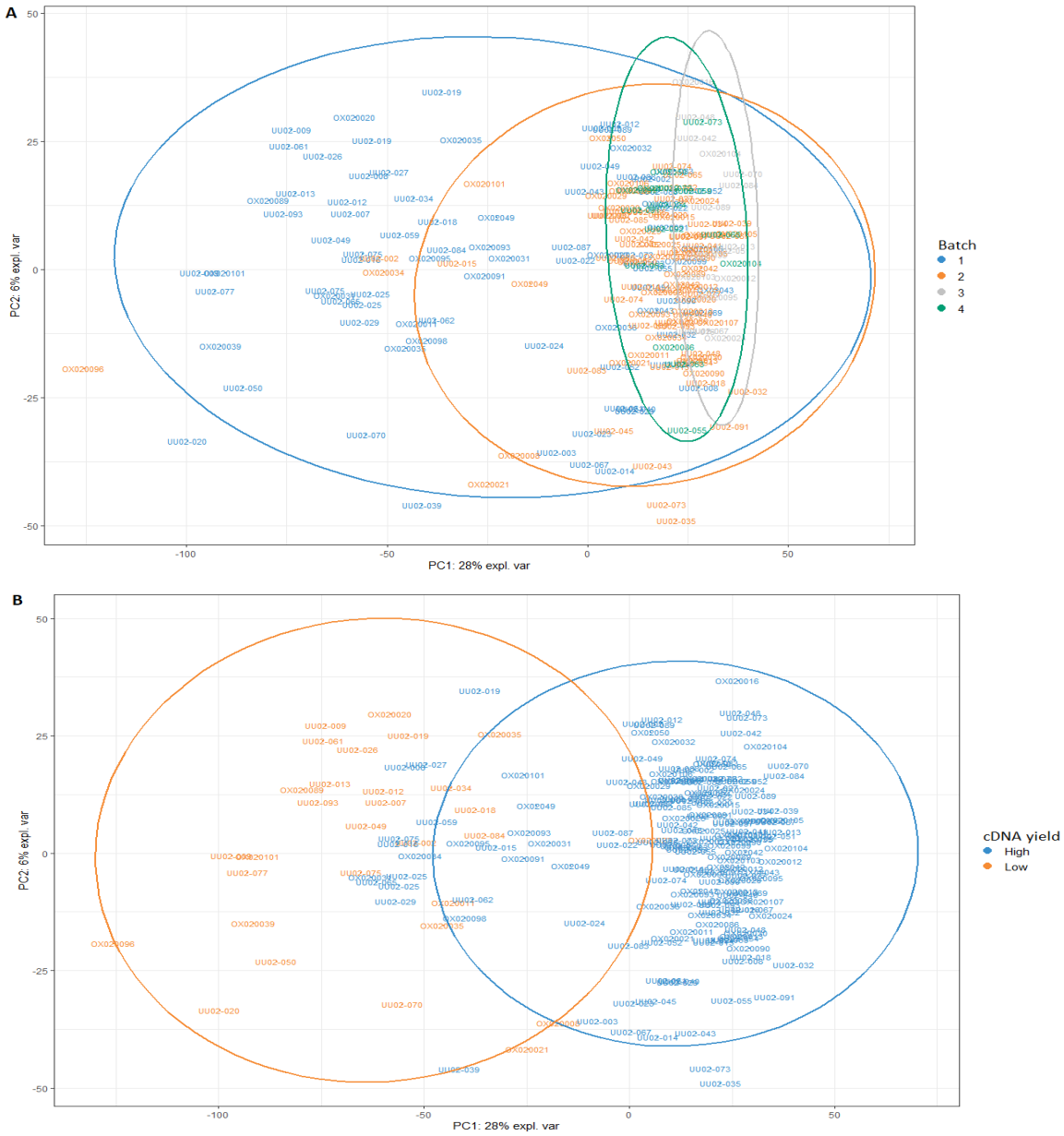


Figure 3.3.3.3 Scree plot of loadings of principal components in quantile normalized data showing proportion of variance attributable to each principal component The percentage of the total variance in the data set represented by each principal component is shown for the first ten principal components. Each principal component represents a distinct grouping of co-correlated probes that do not correlate with other groupings co-correlated probes.


A
PC2: 6% expl. var
PC1: 28% expl. var
Batch
1
2
3
4
B
PC2: 6% expl. var
PC1: 28% expl. var
cDNA yield
High
Low

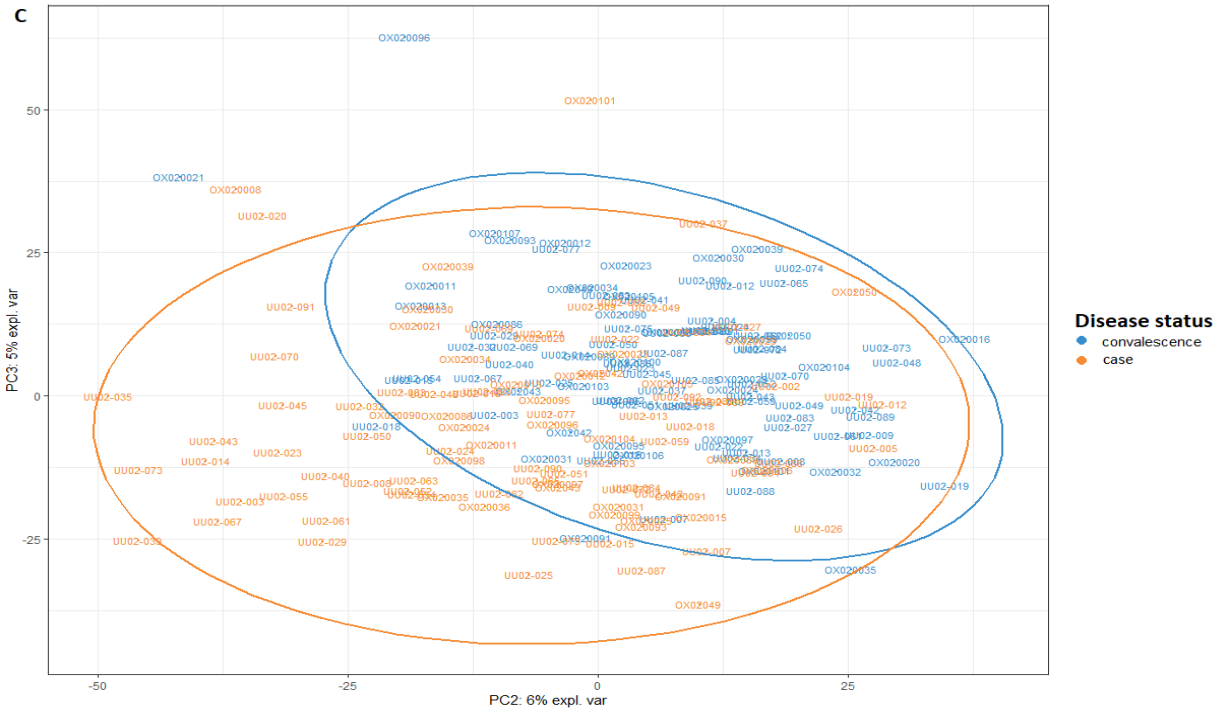


Figure 3.3.3.4 PCA mappings of samples after RMA normalization, sex chromosome removal and probe filtration Grouping of samples according to expression of probes in the A: first and second principal components coloured by microarray batch, B: first and second principal components coloured by cDNA yield and C: second and third principal components coloured by disease status is shown. Expression of the first principal component is mapped on the x axis and expression of the second principal component is mapped on the y axis. Samples originating from patients recruited in Oxford are labelled with the study title “OX02”, while the samples from patients recruited in Utrecht are labelled as “UU02”. Samples are coloured according to infection status of the patient they were taken from.

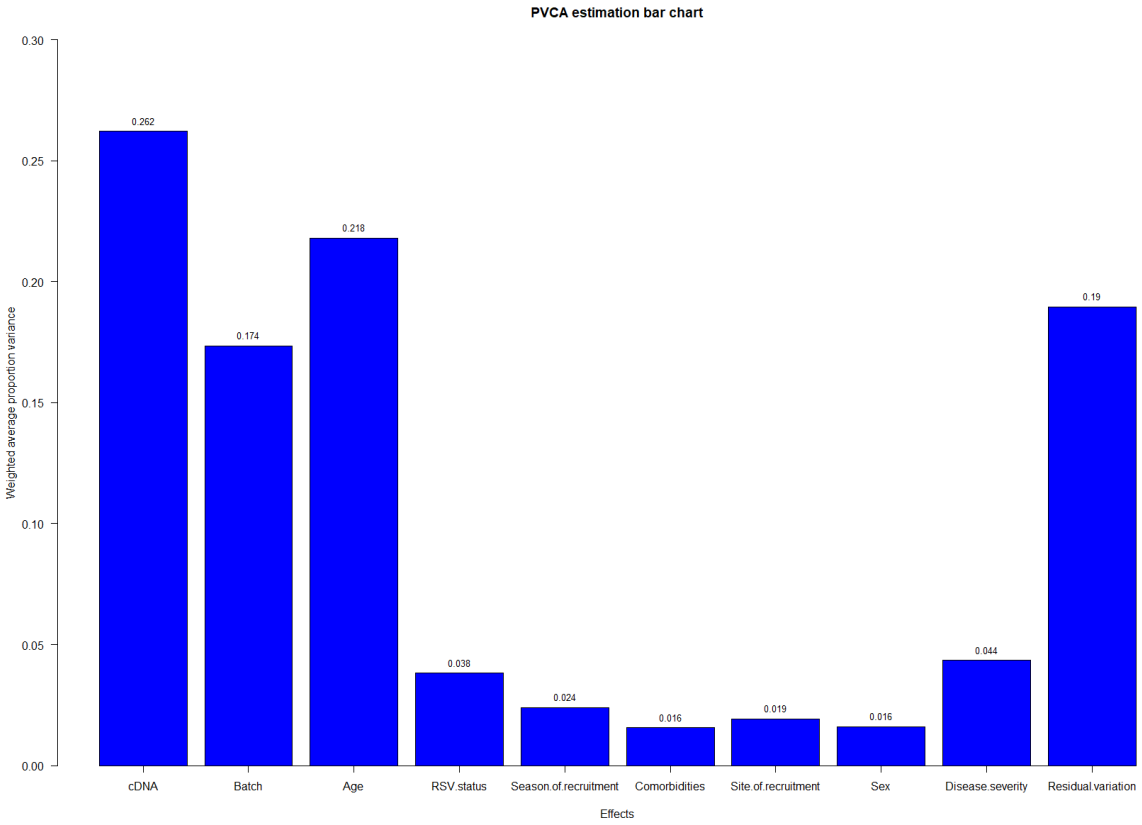


Figure 3.3.3.5 PVCA of quantile normalized data of all samples showing the proportion of variation in the data set attributable to patient characteristics and technical effects, such as Batch and cDNA yields

3.3.4 Batch correction with ComBat removes effects of microarray batches and cDNA yields

Batch correction was performed on the data set using the ComBat algorithm as described in chapter 3.2.9, which was provided with the microarray batch number for each sample, as well as whether that sample had a lower cDNA yield than 10ng. Principal component analyses were run using only the 98 samples from participants at the time of infection, and the 98 samples from participants in convalescence. After batch correction the first principal component accounted for 13% of all the variation in the data set, the second accounted for 9% and the third accounted for 5% (fig 3.3.4.1). Clustering according to batch or cDNA yield was no longer visible in the first or second components (fig 3.3.4.2 A, fig 3.3.4.2 B), and separation of samples based on RSV infection status becomes visible on both the first and

second components. PCA mappings on the second and third components for batch and cDNA and the first and second components for disease status are shown in appendix 1.1.3.

Analysis of sources of variation in the dataset using PVCA (fig 3.3.4.3) indicates that the effect of microarray batch is now minimal, while cDNA yield still accounts for a large amount of variation (27%). disease severity, infection with RSV and age are the largest remaining sources of variation in the data set.

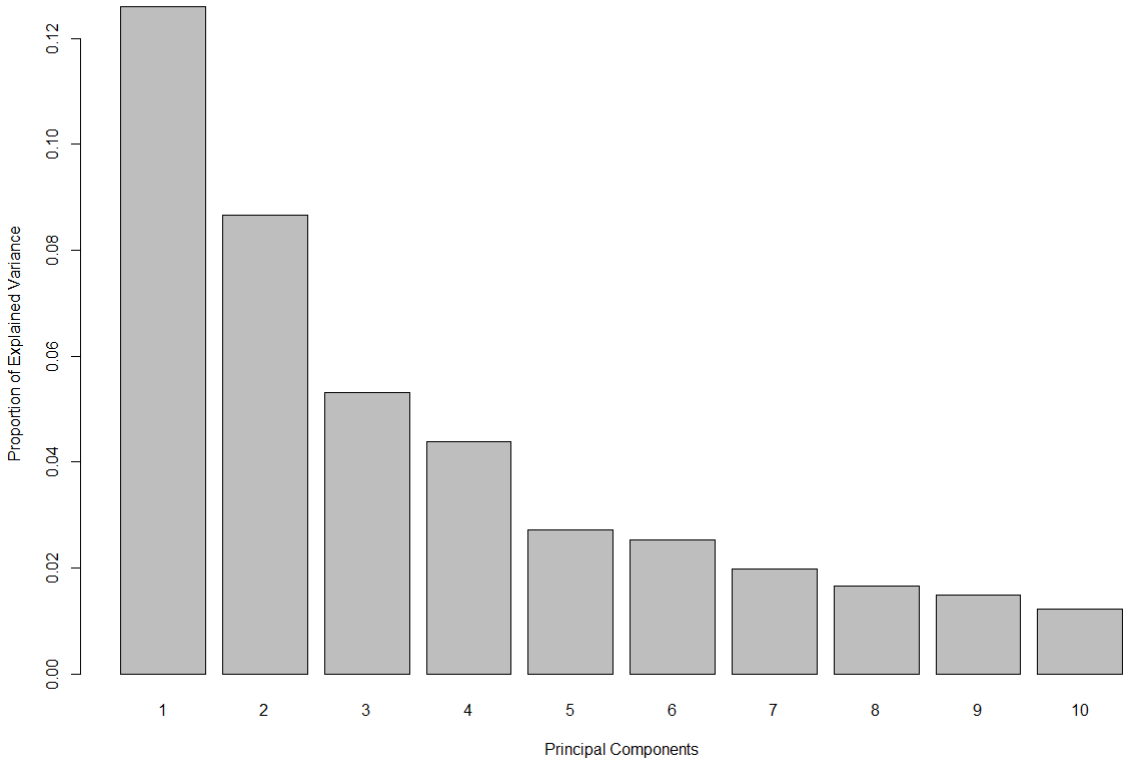
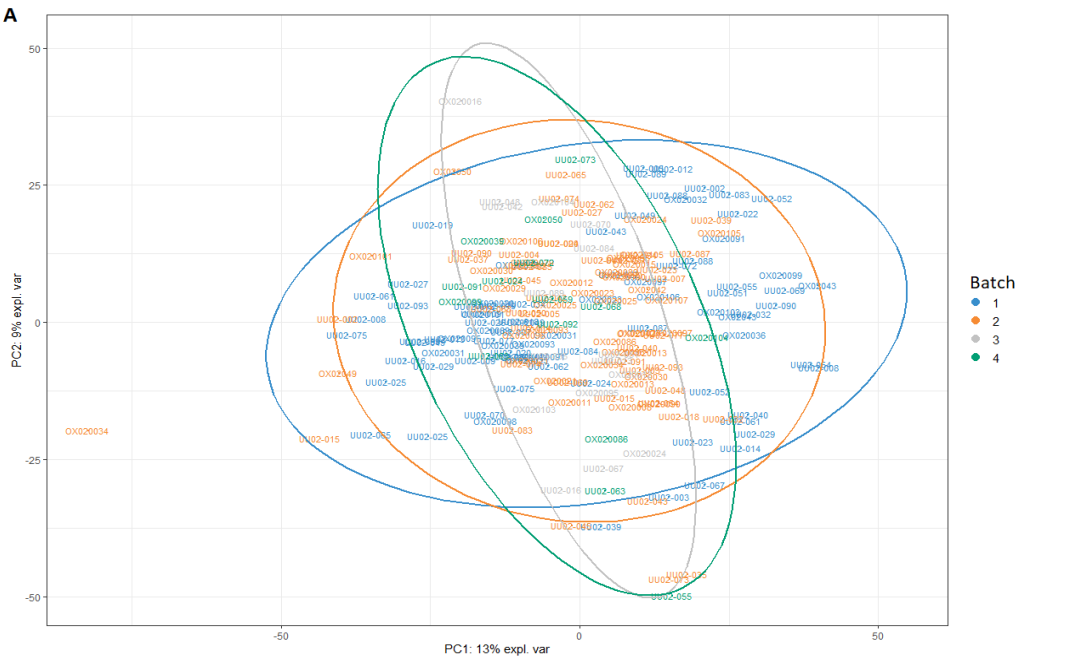
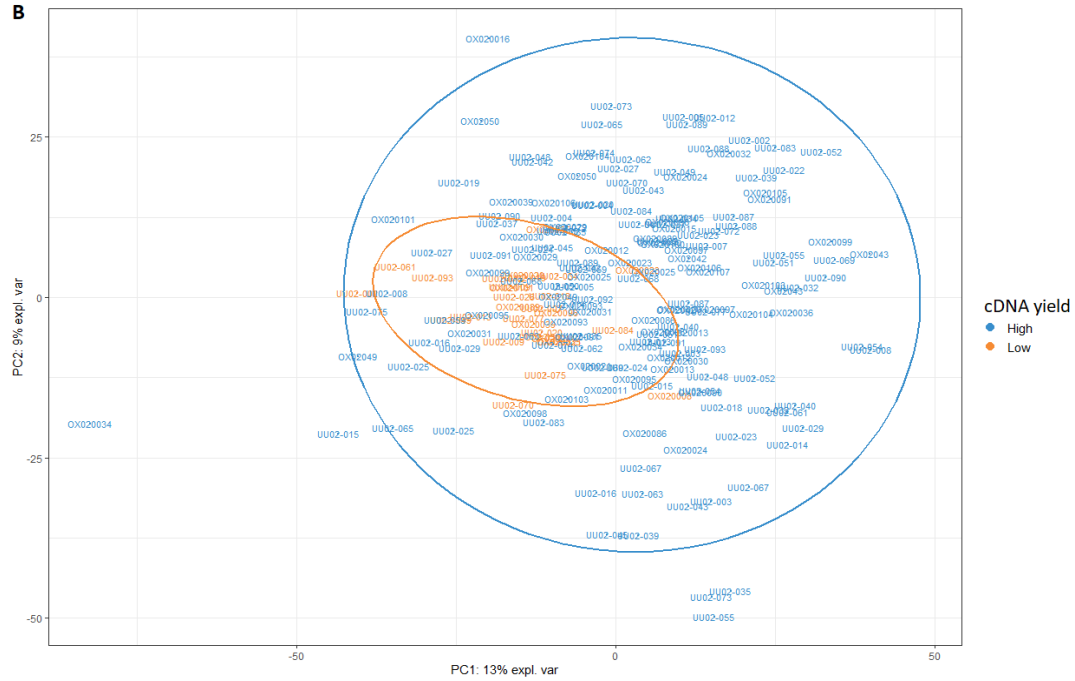


Figure 3.3.4.1 Scree plot of loadings of principal components in quantile normalized and batch corrected data showing proportion of variance attributable to each principal component The percentage of the total variance in the data set represented by each principal component is shown for the first ten principal components. Each principal component represents a distinct grouping of co-correlated probes that do not correlate with other groupings co-correlated probes.

A


PC2: 9% expl. var
PC1: 13% expl. var
Batch
1
2
3
4

B


PC2: 9% expl. var
PC1: 13% expl. var
cDNA yield
High
Low

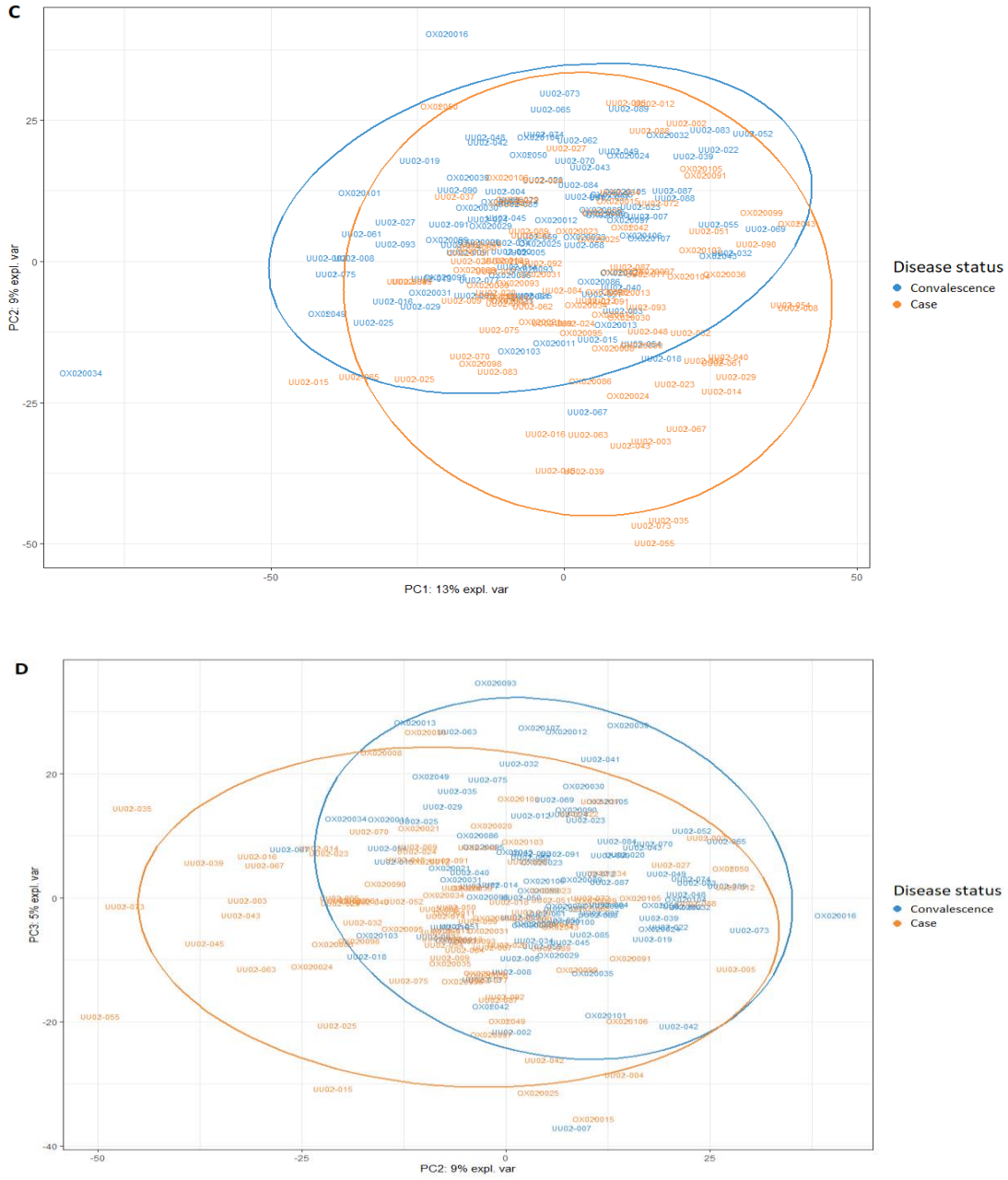


Figure 3.3.4.2 PCA mappings of samples after normalization and background correction Grouping of samples according to expression of probes in the A: first and second principal components coloured by microarray batch, B: first and second principal components coloured by cDNA yield, C: first and second principal components coloured by disease status and D: second and third principal components coloured by disease status is shown. Expression of the first principal component is mapped on the x axis and expression of the second principal component is mapped on the y axis. Samples originating from patients recruited in Oxford are labelled with the study title “OX02”, while the samples from patients recruited in Utrecht are labelled as “UU02”. Samples are coloured according to infection status of the patient they were taken from.

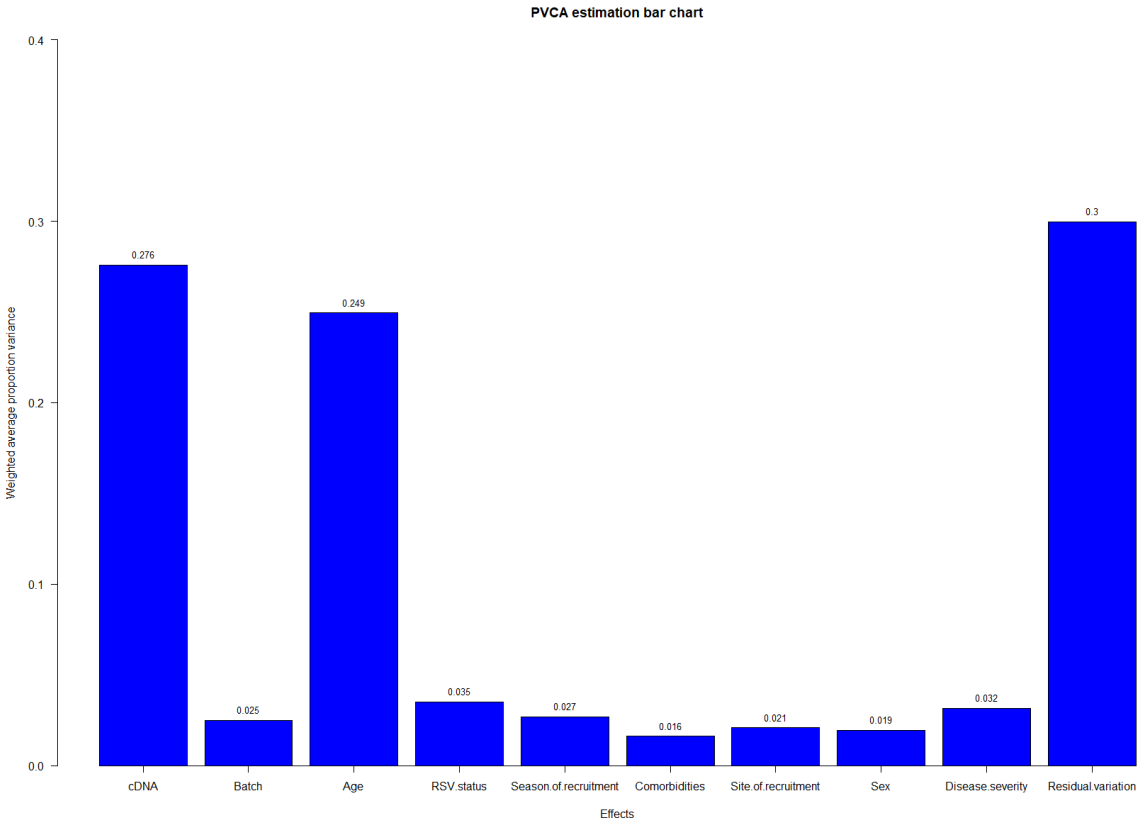


Figure 3.3.4.3 PVCA of batch corrected data of all samples after removal of outliers showing the proportion of variation in the data set attributable to patient characteristics and technical effects, such as Batch and cDNA yields

3.4 RSV infection is associated with neutrophil and interferon signalling

3.4.1 Infection with RSV induces significant differential expression of 1707 genes

In order to determine which genes are significantly differentially expressed between experimental conditions, fitting of normalized, background corrected and expression filtered gene expression to a linear model in the limma package (version 3.44.1) in R(343) was performed as described in 3.2.5, and differential expression was determined by comparing the expression of a gene between acute infection and convalescence while accounting for the effects of age, site of recruitment and sex. A false discovery rate adjusted p.value (FDR) of .05 was considered significant.

Differentially expressed genes during an RSV infection were identified by comparison of whole blood transcriptome of 98 patients at the time of acute infection with the whole blood transcriptome of those same patients in convalescence 7 weeks later (fig. 3.4.1.1).

Additionally, differentially expressed genes were identified between seven healthy control patients and the 98 patients with an RSV infection, at both time of infection (sfig. 1.2.1) and in convalescence (sfig 1.2.2) are described in the supplemental materials.

1707 genes were found to be differentially expressed after application of a false discovery rate (FDR) adjusted p.value cut off of .05 of which 951 were upregulated and 756 were downregulated. The most highly differentially expressed gene between acute infection and convalescence is *IFI27*, a gene involved in the response to IFN-γ signalling, followed by *MMP8* and *OLFM4*.

Differential gene expression analysis between the healthy controls and RSV disease is shown in appendix 1.2

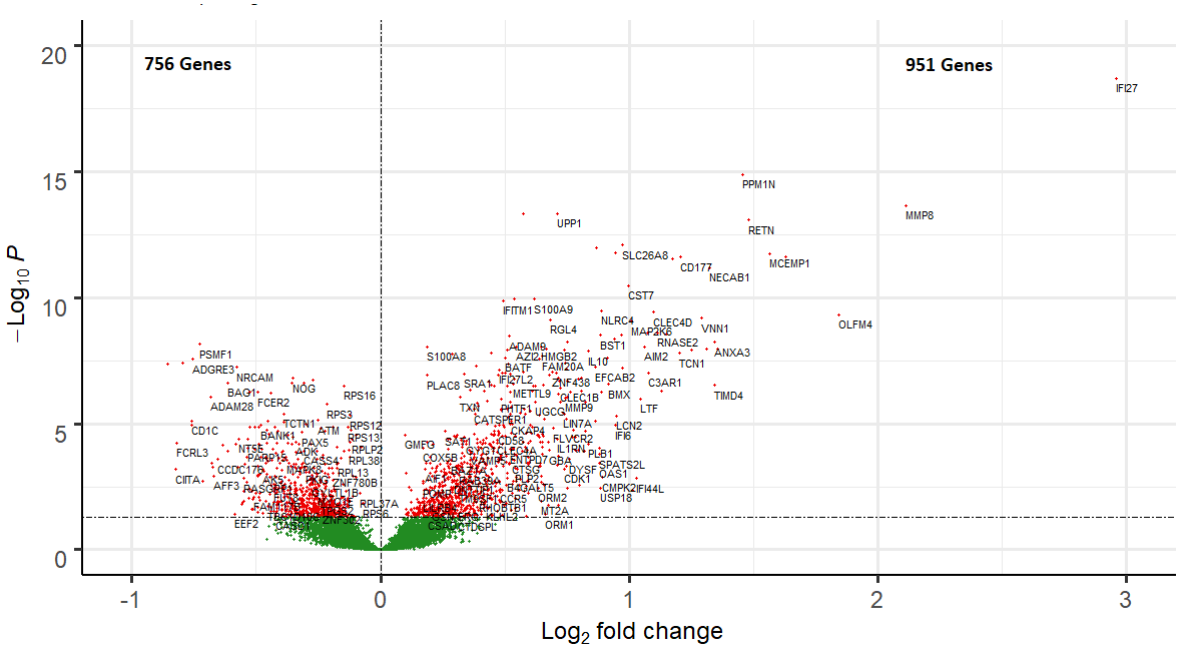


Figure 3.4.1.1 Volcano plot illustration of transcriptional response to RSV infection for each gene when comparing RSV infected children to convalescent children. Fold change of genes between acute infection and convalescence is plotted against p value of change in that gene. N=98 patients at the time of RSV infection compared with the same patients 7 weeks later in convalescence. Red dots denote significantly differentially expressed probes (FDR <.05), green dots indicate genes that are not significantly differentially expressed. Horizontal dotted lines represent FDR cut-offs, vertical dotted lines represent log2 fold change cut-offs.

3.4.2 Gene set enrichment analysis reveals neutrophil degranulation and decreased ribosomal activity to be associated with RSV infection

The broad institutes' Preranked GSEA algorithm (version 4.0)(345) was used to search for enrichment of terms from biological process (C5.1), cellular component (C5.2), molecular function (C5.3) immunological (C7) and hallmark (H) gene expression databases to determine function of changes in gene expression in whole blood in response to RSV infection as described in chapter 3.2.6.

Gene set enrichment analyses were run using only the 98 samples from participants at the time of infection, and the 98 samples from participants in convalescence. All genes included in the analysis were used with the GSEA algorithm in order to determine enrichment of functional terms for those genes. Genes were ranked according to the t-statistic, which takes into account both significance of the change in expression and the direction of regulation. Terms were considered significantly enriched in the gene set if they passed an FDR cut off of .05. The Preranked GSEA algorithm produces an enrichment score that can be used to rank terms according to their degree of enrichment in the gene set. Terms were plotted by enrichment score for the top 20 highest scoring terms for the provided gene set. The number of genes that are annotated with a given enrichment term was also plotted as dot size.

The most highly upregulated biological process term (fig 3.4.2.1a) in these analyses is antimicrobial humoral response, which refers to defence response against microbes through the medium of a bodily fluid. Multiple terms relating to IL-1 regulation and production are enriched, as well as multiple terms relating to reactive oxygen species (possibly attributable to a high degree of neutrophil proliferation). Terms relating to chemokine induction were also induced, indicating that virally induced signals are drawing cells to the lung tissue. Another enriched term is neutrophil migration, which in conjunction with other terms related to neutrophil activity (specifically the enrichment of reactive oxygen species terms and antimicrobial immunity terms) suggests that the transcriptomic response to RSV may be dominated by neutrophils.

The most highly downregulated biological process terms (fig 3.4.2.1b) involved viral gene expression, suggesting that means by which the virus replicates within the host are

downregulated during infection, ribosomal proteins have been previously associated with viral infection as viruses have been demonstrated to utilise the human ribosomal complexes to facilitate replication(352). It is possible that downregulation of ribosomal proteins is one of the means by which the host downregulates RSV replication. Another common term among the most highly enriched downregulated terms related to translation of mRNA to protein, which is likely related to the downregulation of ribosomal terms, as ribosomes are one of the primary means of protein translation. B cell related terms are also significantly downregulated in acute infection, which is similar to what has been observed previously in transcriptomic studies in acute RSV infection(285).

The hallmark GSEA database is a highly curated set of few distinct and not co-correlating terms that aims to represent specific biological states. Interferon alpha responses are the most highly enriched hallmark term in this dataset (fig 3.4.2.3), with interferon gamma responses also being significantly enriched. Other significantly enriched terms include xenobiotic metabolism, mTORc1 signalling (indicating a high degree of cellular replication), coagulation and complement. No downregulated hallmark terms were identified in this analysis.

The most highly upregulated cellular components (fig 3.4.2.2a) in this dataset all related to granules and cell membranes primarily associated with granulocytic cells and neutrophils. The top five most highly enriched terms all related to the specific granule and tertiary granule, which are membrane contained molecules primarily associated with neutrophils, and are one of the primary effector mechanisms of neutrophils(353). The most highly downregulated cellular components(fig 3.4.2.2b) primarily related to ribosomal terms, indicating downregulation of production of ribosomes in response to viral infection, and therefore a downregulation of translation.

The immune cell database referenced in GSEA shows the similarity of a ranked expression values of a set of genes to that observed in other published microarray studies. The algorithm compares the similarity between the ranked set of genes it is presented with, to the top and bottom 200 genes (ranked according to t score) for that study. Enrichment terms from this database show a simplified title of the study that the genes are related to, describing briefly the comparison being made in the experiment the gene set is derived from. The most highly positively enriched term from this database (3.4.2.4a) in the present

dataset is from an experiment where PBMCs were treated with G-CSF, a molecule that induces proliferation of neutrophils(354). Other highly enriched terms are from influenza and RSV studies, along with terms from bacterial studies. The fourth most highly enriched term for this database relates to a study involving *pneumococcus*, proliferation of which has been observed in acute RSV infection previously. In light of the highly enriched neutrophil related terms in the cellular components analyses, it is possible that the observed bacterial terms may be attributable to the strong neutrophil response induced by RSV infection. Most highly downregulated terms from the immune database (fig 3.4.2.4b) primarily come from studies involving influenza and RSV, as well as studies of neutrophil and CD4 T cells, meaning that the downregulated terms from the current study have some overlap with the downregulated terms from these studies.

The molecular function database shows upregulation of pattern recognition receptors, lipopolysaccharide and various kinases in acute RSV (fig 3.4.2.5 A) and downregulation of MHC class II molecules, rRNA and ribosomes (fig 3.4.2.5 B).

Overall enrichment analyses suggest a strong upregulation of neutrophil migration and activity, alongside a downregulation of host ribosomes, possibly to control viral replication, with simultaneous upregulation of type I and type III interferon related terms, and a downregulation of B cell related terms.

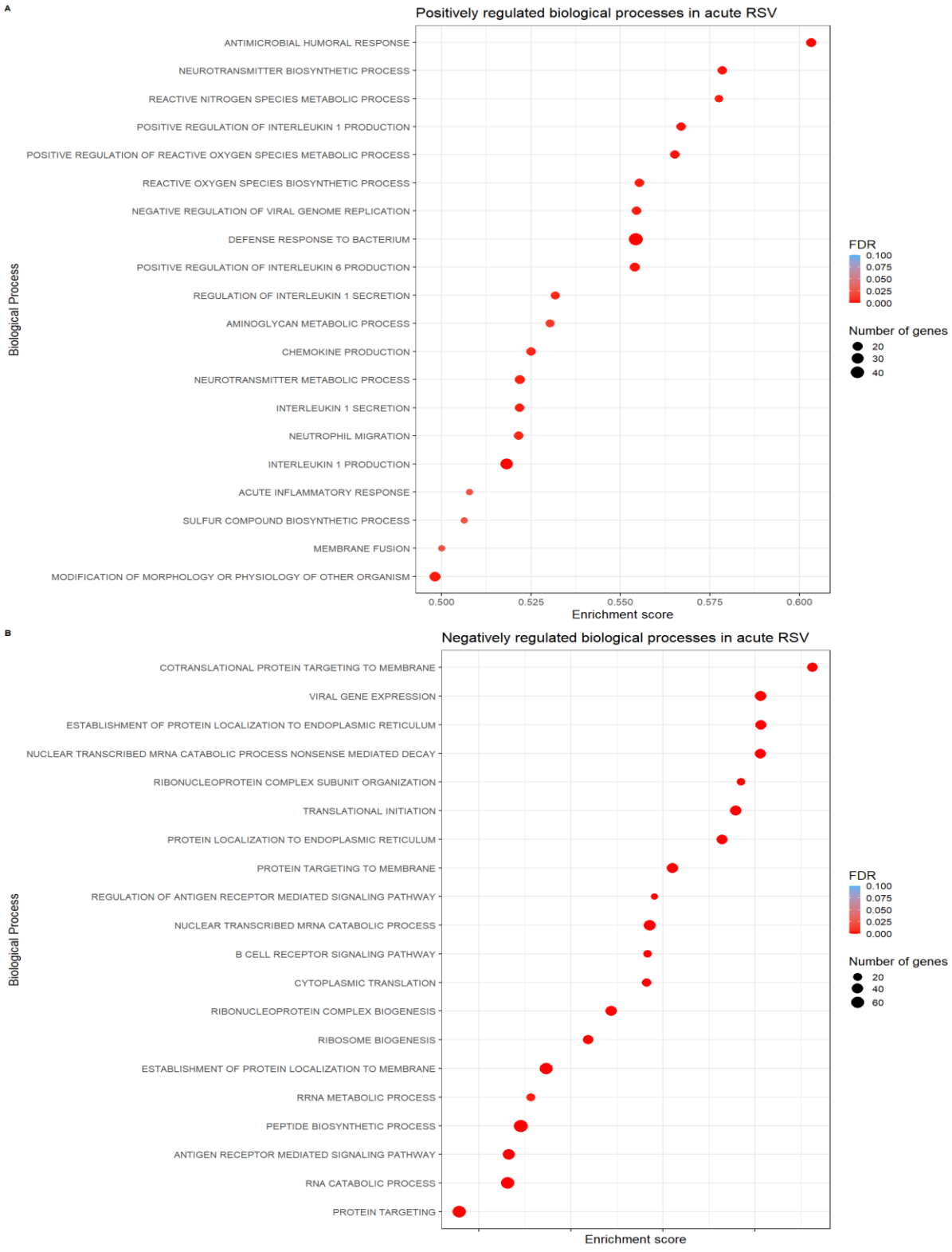


Figure 3.4.2.1 Gene set enrichment analysis of A Positively regulated and B negatively regulated biological processes from the set of all genes in analysis ranked by statistic between patients at the time of RSV infection and the same patients in convalescence. Terms are ranked according to enrichment score for the term. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

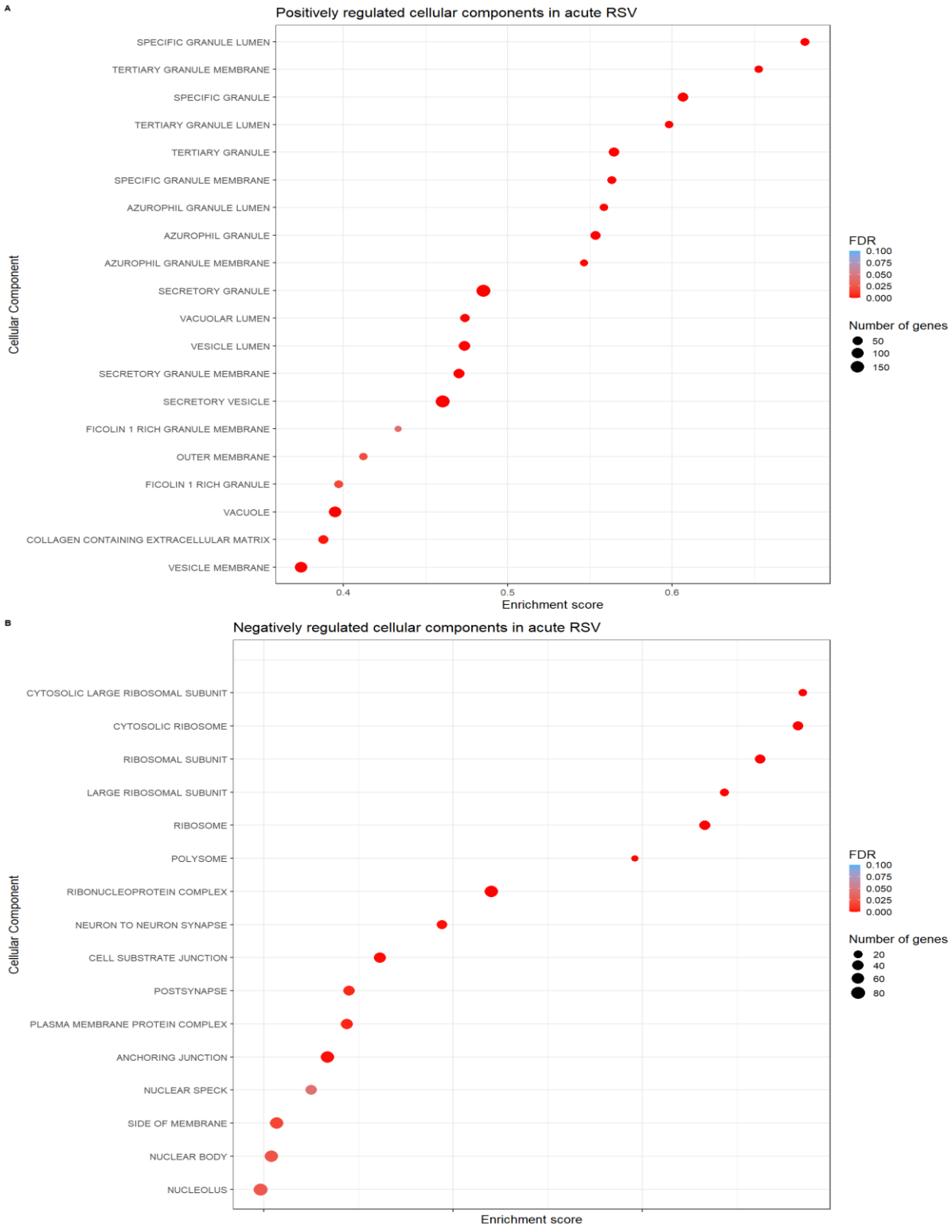


Figure 3.4.2.2 Gene set enrichment analysis of A Positively regulated and B negatively regulated cellular components from the set of all genes in analysis ranked by statistic between patients at the time of RSV infection and the same patients in convalescence. Terms are ranked according to enrichment score for the term. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

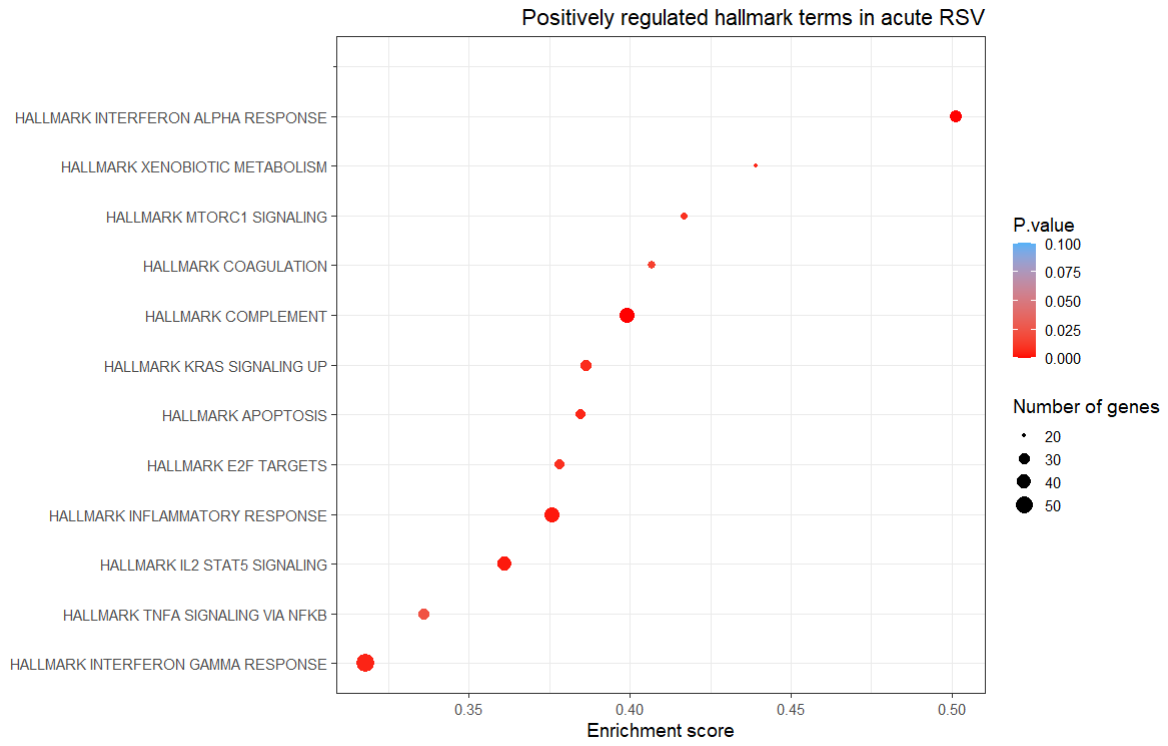


Figure 3.4.2.3 Gene set enrichment analysis of Positively regulated hallmark gene expression terms from the set of all genes in analysis ranked by statistic between patients at the time of RSV infection and the same patients in convalescence. Terms are ranked according to enrichment score for the term. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

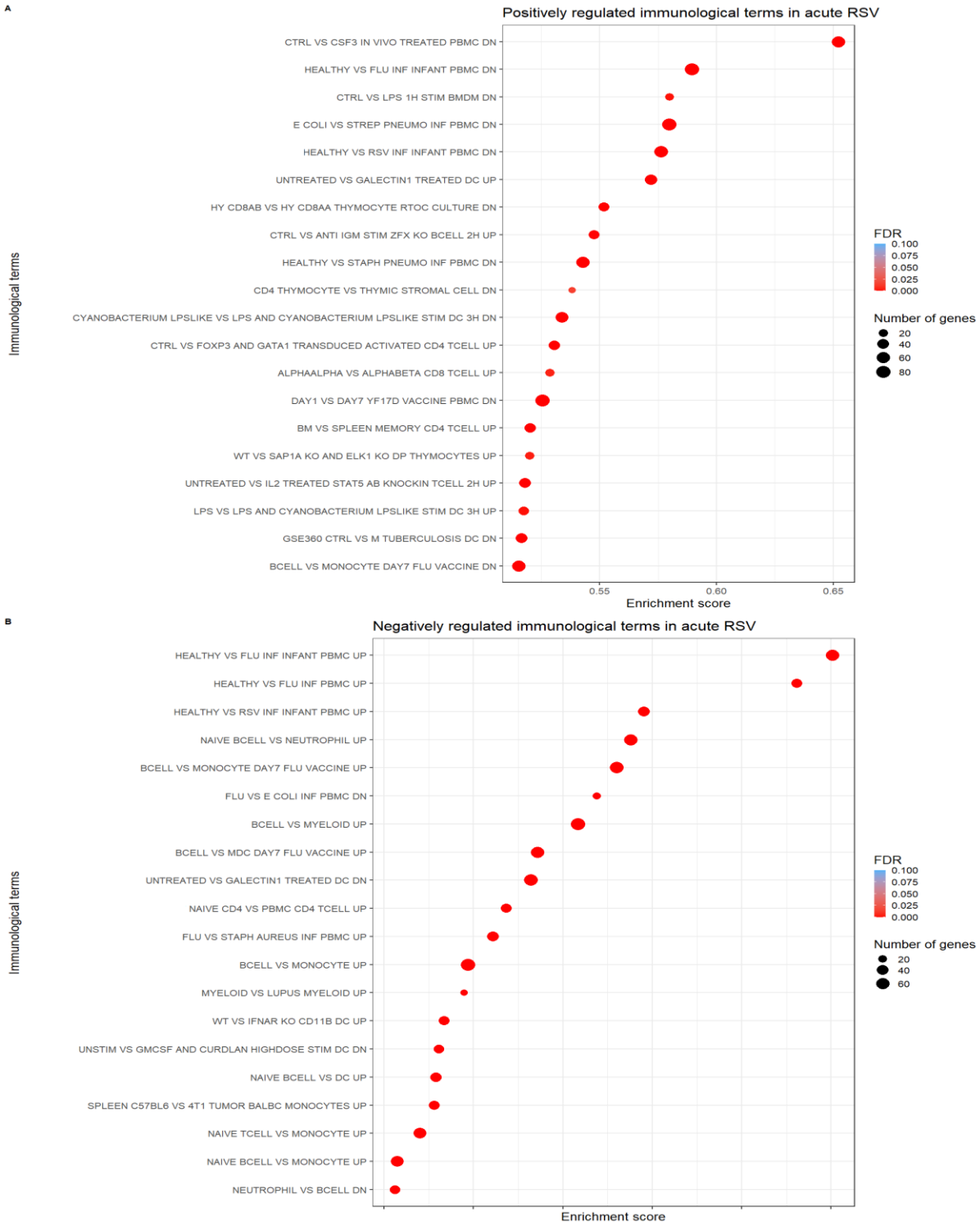


Figure 3.4.2.4 Gene set enrichment analysis of A Positively regulated and B negatively regulated immunological database terms from the set of all genes in analysis ranked by statistic between patients at the time of RSV infection and the same patients in convalescence. Terms are ranked according to enrichment score for the term. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

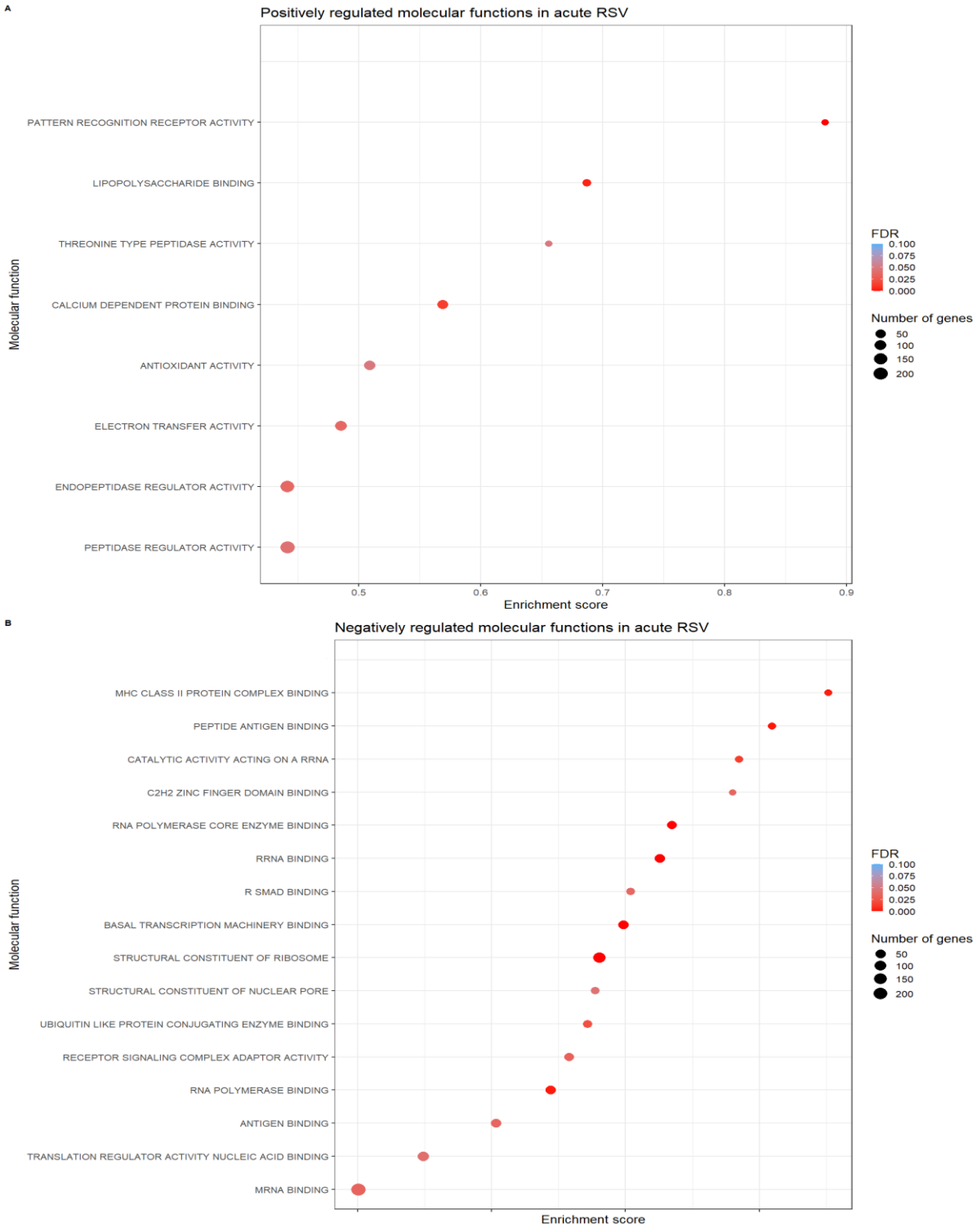


Figure 3.4.2.5 Gene set enrichment analysis of A Positively regulated and B negatively regulated molecular functions from the set of all genes in analysis ranked by statistic between patients at the time of RSV infection and the same patients in convalescence. Terms are ranked according to enrichment score for the term. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

3.4.3 Increased populations of neutrophils, eosinophils and monocytes associate with acute RSV infection, while increased proliferation of B cells and CD8+ T cells occurs in convalescence

In order to estimate the abundances of cells in whole blood during an acute RSV infection, and how they change with infection the Cibersortx deconvolution tool (version 1.05) was used(348) as described in chapter 3.2.8. Cibersortx uses gene expression profiles from a given tissue to estimate the abundance of cell types in that tissue. Cibersortx compares the expression profile of a gene expression set provided to it with a database of expression signatures associated with selected cell types. The immune cell database (LM22) was used for this analysis. Normalized and filtered gene expression values of differentially expressed genes were used to calculate cell abundances in paired samples (during ongoing RSV infection and in convalescence 7 weeks later) from 98 participants. Healthy controls were also included in analyses to determine if any differences in cell proportions existed between healthy controls and participants in convalescence.

A significantly higher abundance of monocytes in whole blood is inferred by deconvolution analysis in patients with an ongoing RSV infection compared with those in convalescence or healthy controls (fig 3.4.3.1 F). Resting NK cells also show a significant increase in acute RSV disease when compared to convalescence (fig 3.4.3.1 L). Significantly higher abundances of memory B cells (fig 3.4.3.1 H), and resting memory $CD4^+$ T cells (fig 3.4.3.1 B) were seen in patients in convalescence compared with patients with an acute RSV infection. Many cell types showed no significant increases between RSV infection and convalescence, including M2 macrophages (fig 3.4.3.1 I), dendritic cells (both resting and activated) (fig. 3.4.3.1, J,K), activated natural killer (NK) cells (fig. 3.4.3.1 M), naïve $CD4^+$ T cells (fig 3.4.3.1 A) and $CD8^+$ T cells (fig 3.4.3.1 N)

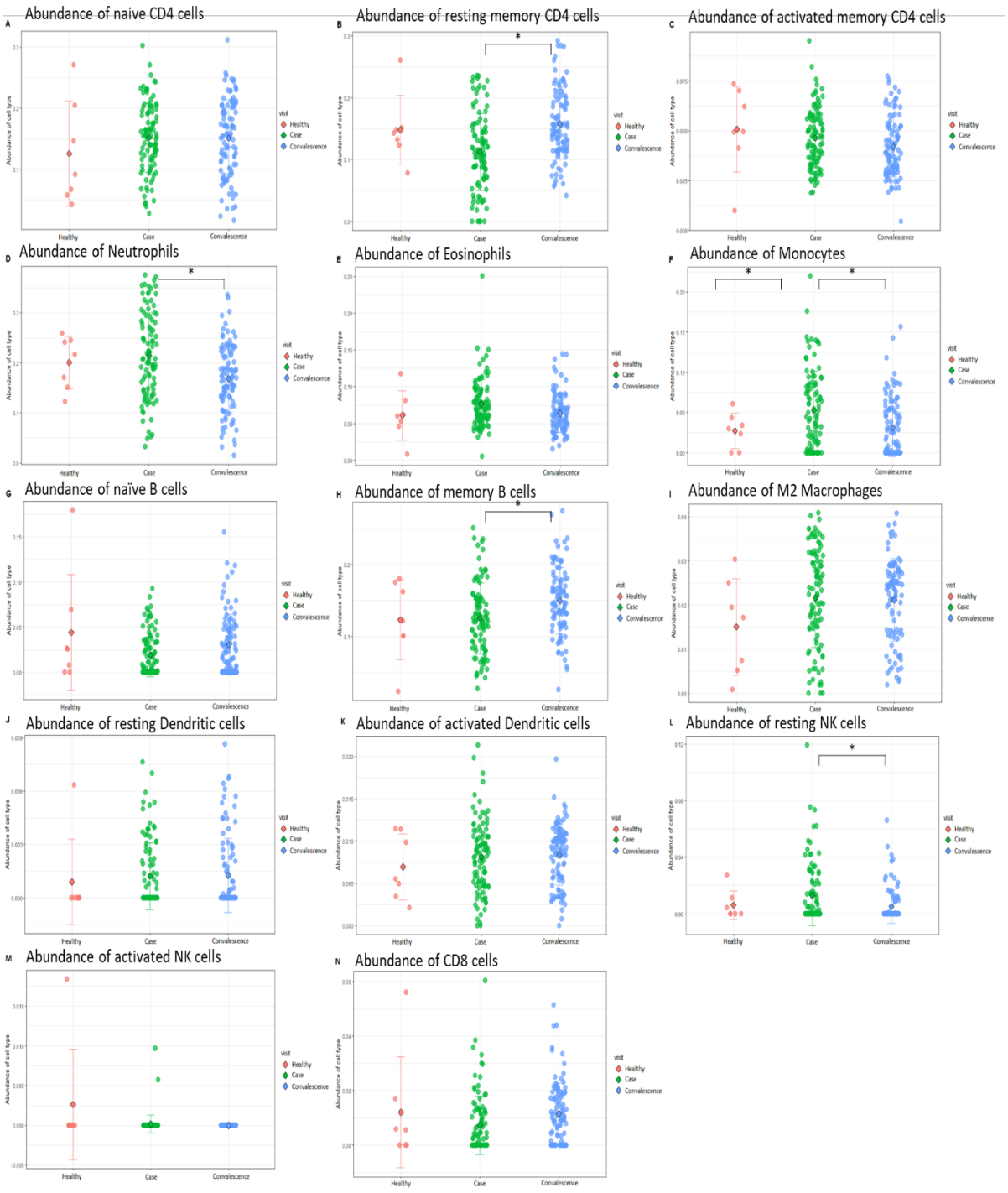


Figure 3.4.3.1 Relative abundances of cell types derived from transcriptomics data using CIBERSORTx A – naïve CD4 T cells, B – resting memory CD4 T cells, C – activated memory CD4 T cells, D - neutrophils, E - Eosinophils, F - Monocytes, G – naïve B cells, H – memory B cells, I – M2 macrophages, J – resting Dendritic cells, K – activated Dendritic cells, L – resting NK cells, M – activated NK cells , N – CD8 T cells detectable in transcriptomic data from whole blood between patients with ongoing RSV infection (cases) and the same patients in convalescence 6-8wks after initial RSV infection. * = FDR < .05 according to FDR adjusted t test

3.5 Discussion

3.5.1 Summary of main findings

RSV infection induces changes in gene expression that are detectable in whole blood. These initial changes in gene expression brought about by infection are characterised by increased expression of genes related to Th1 type signalling pathways, including type-I interferons and genes downstream of them (in particular the IFI and IFIT families of genes). Additionally, neutrophil degranulation and neutrophil migration related gene expression are strongly induced in RSV infection, and increased neutrophil counts are detectable in the whole blood transcriptome, indicating an increase in neutrophil traffic towards the lungs. Increased monocyte counts are also observed during acute infection as well as significant downregulation of ribosomal gene expression, indicating possible downregulation of translation to control viral replication.

A shift towards a memory response is seen in convalescence, with significant increases in both resting and active memory $CD4^+$T cells as well as memory B cells. Increases in $CD8^+$ T cells and plasma cells are also observed in convalescence. However, these increases are not statistically significant.

3.5.2 Quality control of microarray analysis

Quality control of samples identified ten samples to be removed from the analyses. Quantile normalization was selected as the most appropriate normalization method for these data, and successfully made samples more comparable to each other. Batch correction was required due to the large amount of variation in the data set attributable to microarray batches and cDNA yield, and correction for this using ComBat resulted in variation attributable to infection with RSV to become the largest source of variation in the data set.

Principal component analyses of samples indicate that variance attributable to RSV infection associates with the variation on the first and second components after batch correction and is the largest source of variation and clustering in the data set.

3.5.3 Differential expression and gene set enrichment

The most highly differentially expressed genes induced by RSV are *IFI27*, *MMP8* and *OLFM4*, which are characteristic of interferon-γ signalling (IFI27)(355), induction of neutrophil responses (MMP8, OLFM4)(356)(357) and NF-kB responses (OLFM4)(292), as is observable from gene set enrichment analyses performed on this data set. This is a similar pattern of expression to what has been previously in gene expression studies of RSV(281)(276), as well as what has been seen in Influenza virus infection(358)(359).

Consistent upregulation of terms related to antibacterial responses and humoral responses are observed between participants infected with RSV and participants in convalescence. It is possible that these are an artifact of the neutrophil response that occurs in RSV infection, as neutrophils have well characterised roles in responding to bacteria(360), and gene expression associated with secreted antibacterial neutrophil granules such as azurophilic granules is significantly upregulated in participants with acute RSV infection in this study. A strong expansion of certain commensal bacteria located in the respiratory tract has also previously been observed in RSV infection(239), which may also explain the upregulation in bacterial response terms. This hypothesis is supported by the enrichment results from the cellular component database, where the top nine most highly upregulated cellular components are all related to neutrophils, specifically to the neutrophil granules and membranes. Induction of these suggests an increase in neutrophil degranulation activity. This increase in neutrophil activity is characteristic of RSV, and has been observed in other comparable respiratory viruses, such as influenza(359) where it is also a key aspect of pathogenesis, and is associated with worsened disease outcomes and higher risk of bacterial co-infection. It is possible that neutrophil proliferation is not entirely harmful in RSV infection, as some evidence exists for influenza infected neutrophils acting to present antigen to CD8 T cells(361), and NETs have been shown to potentially have some antiviral activity(215). It is possible therefore that neutrophils have some role in controlling viral replication despite causing significant immunopathogenesis.

The immune database that was used in GSEA in this study showed enrichment of terms primarily from studies in RSV and influenza, as well as terms from experiments where cells were treated with granulocyte stimulating factors such as G-CSF. Bacterial response terms

are also highly enriched in the immune database analyses, likely due to the large amount of neutrophil related terms as discussed above.

Molecular function database shows that pattern recognition receptors are strongly enriched in the gene set at the time of acute infection, while MHC class II molecules are more highly enriched in convalescence, indicating a large myeloid response at the time of acute infection, followed by a shift towards lymphocytic immunity in convalescence. This is to be expected, considering the induction and proliferation of neutrophils and monocytes observed in this study. These dynamics are consistent with the observations previously made on accumulations of neutrophils and monocytes early in respiratory viral infections(61) followed by proliferation of lymphocytes in convalescence(362). Monocytes have demonstrated an increase in expression of pattern recognition receptors during infection(363). A consistent lymphopenia is observed early in influenza infection(364).

3.5.4 Deconvolution of cell types in whole blood

Deconvolution analysis of the normalized expression values for differentially expressed genes in this data set allowed for the calculation of relative abundances of various cell types in the blood at both the time of infection and in convalescence, which can provide useful information about the timing of responses of cell types in response to infection. Based on what was revealed in these analyses, RSV infection can be characterised by a sudden induction of neutrophil and eosinophil responses during acute infection, as well as a proliferation of monocytes and memory $CD4^+$ T cells, with a switch to B cell and memory responses in convalescence.

During convalescence, the fall in blood neutrophil and eosinophil abundances (although not statistically significant), simultaneous with significant increases in populations of naïve B cells and memory B cells indicates a shift away from acute inflammatory responses and towards a memory response. This is consistent with what has been observed previously, where neutrophil responses decline over time post infection(146), while RSV specific B and T cells (particularly $CD8^+$ T cells) increase in abundance with the transition into convalescence(362). Similar patterns have been observed in other viral infections. In Dengue virus infections T cell populations typically peak 7-10 days post onset of symptoms(365), while in Influenza they peak at 8-10 days post onset of symptoms(366).

There is no observable difference in the transcriptomic signature associated with plasma cells between patients during RSV disease and during convalescence. While it is possible that RSV is weakly inducing of plasma cells, it is also possible that there has been an increase in RSV specific plasma cell populations but not overall plasma cell populations, which is not detectable without assays to determine antigen specificity of plasma cell populations. This also applies to other populations that develop specificity to a pathogen, such as $CD4^+$ and $CD8^+$ T cells. RSV specific plasma cell populations have previously been observed to significantly increase within the first 11 days of infection(367). Similarly within B cell populations, significant increases only in activated subsets were seen in acute RSV infection(97). Significant increases in memory B cell populations have been observed weeks following infection in Influenza are interesting, as one might expect only small populations of RSV specific memory B cells to remain. This is possibly attributable to natural increases in memory B cell populations that are naturally seen in children below six months of age(368). Further data generated through CyTOF or flow cytometry may aid in explaining this phenomenon.

3.5.5 General limitations

A lack of statistical power is consistently seen when making comparisons involving the healthy control patient population, due to the relatively small numbers of these patients recruited to the study. Because of this, the convalescent samples taken from the patients included in the acute disease analysis are used as an imperfect control. As some aspects of the inflammatory response (in particular $CD8^+$ T cells) remain perturbed (and indeed even more highly perturbed than during acute RSV infection) even seven weeks after infection(145), however many aspects of the host transcriptome will have returned to baseline, making this population suitable to study the disease in these patients. The use of whole blood as the tissue to be studied creates some questions about the applicability of these results to the lower respiratory tract, where the infection is primarily located, however as many cells must traffic to the lung through the blood, any changes in cell recruitment to the lung will be reflected in the whole blood transcriptome. Using transcriptomic data, it is possible to capture changes in populations and proportions of cell types, however it is not possible with the methods employed in this study to estimate populations of antigen specific cell types. Therefore, the fact that no significant changes in

the populations of $CD8^+$ T cells are observed does not necessarily mean that a small population of antigen specific cells was not induced by infection, as the non-specific CD-8 T cell response typically peaks 9-12 days post onset of symptoms(362). Such a population may exist, and not create any statistically significant change in the total numbers of cells, and thus not be detectable by these methods.

Many of the inferences made in this analysis about cell populations will require supportive data using technologies capable of robustly assessing cell proportions, such as flow cytometry, or mass cytometry (CyTOF). These data were not available at the time of analysis but will become available in the future.

There is a lack of significance in comparisons between deconvoluted cell proportions in healthy controls and RSV cases, likely due to the small numbers of healthy control samples available for analysis (7) when compared to the much larger number of RSV infected samples (98).

3.5.6 Conclusions

These data capture the change in the inflammatory gene expression response between acute infection and convalescence, giving a snapshot of RSV disease as observable in the blood, and how it changes as the infection transitions into convalescence.

RSV is strongly inductive of type I interferon and neutrophil responses in the whole blood of infected children during acute RSV infection, as evidenced by gene set enrichment analyses. In this dataset early RSV infection seems dominated by certain innate inflammatory cell types, such as neutrophils and monocytes, which is followed by a transition into increased expression of lymphocytes and memory cells in recovery.

Chapter 4: Associations of changes in gene expression with RSV disease

4.1 Introduction

In the previous chapter (chapter 3) I showed the gene expression signature associated with acute RSV infection and showed how this is reflective of the changes in cell populations and of the immune response during acute RSV infection.

After gene expression analysis, it can be useful to associate those changes in gene expression with some meaningful clinical outcome measures. This can allow us to understand how changes in gene expression in response to infection are associated with disease severity. This can be used to develop more dynamic ideas of patients who are at risk for disease, define the patterns of gene expression that predispose to severe disease, and potentially apply therapeutic or preventative measures.

Previously patterns of gene expression have been associated with more severe disease. In particular expression of genes relating to neutrophil recruitment and activity(278), as well as down-regulation of genes associated with CD4, CD8, NK cell and B cell responses(276). We have used measurement of invasiveness of respiratory support as the primary method of defining severe RSV in this study, due to both its extensive use as a metric of disease severity in other transcriptomics based studies(292)(276)(278)(283)(284), and its incorporation into the GRSS(286), which has also been extensively used in studies of the effects of gene expression on disease severity(238)(277)(279)(285). Usage of this disease severity scoring metric ensures comparability with other studies and allows for the use of these studies in validation of the results of this study. Analysis of other metrics of disease severity (such as ReSVinet score and hospitalisation) are shown in appendices 2.2, 2.3 and 2.4.

In this chapter, I aim to determine the genes and patterns of gene expression with the greatest association with more severe RSV infection, and also to determine the association between these patterns of gene expression and the inflammatory response to disease.

To accomplish these aims, I will first present the disease severity data from patients recruited to the RESCEU study and discuss the different means by which severe disease can be defined, followed by presenting the statistical associations between individual genes and disease severity. I will then use partial least squares discriminant analysis to identify families of highly co-expressed genes that associate with disease severity, and use gene set enrichment analysis to determine the association between these families of genes and inflammatory responses.

4.2 Methods

4.2.1 Disease severity assessment

For purpose of analysis multiple different methods of determining disease severity were considered to ensure applicability of signatures to a broad range of potential disease phenotypes. Methods used were described in chapter 2.5. More severe disease was also defined for our analyses as requirement for receipt of any supplemental oxygen more invasive than high flow nasal cannula (specifically requirement of CPAP/BiPAP or mechanical ventilation. Invasiveness was graded on a scale of 0 to 5 as described in chapter 2.5 for correlative analyses. Other methods of determining disease severity were also used and are included in appendix 2.2.

4.2.2 Correlative analyses

Genes were correlated with measures of severity using Pearson correlation, which measures the covariance of two variables, divided by the product of the standard deviations of those variables. P. values for all Pearson correlations were adjusted using the Benjamini-Hochberg procedure to control the false-discovery rate (FDR) (369). FDR adjusted p.value cut offs of .05 were used to determine significance of correlations.

4.2.3 Venn diagrams

Lists of genes that were significantly correlated with each severity metric were provided to the Venn diagram generation tool developed by Ghent university(370).

4.2.4 PLS-DA

Partial least squares discriminant analysis (PLS-DA) is supervised machine learning method that locates the combination of variables that are maximally different between two or more groups of samples(371). To do this, highly co-correlated variables are linearly combined into single variables that do not correlate with any other combined variables. Samples can then be mapped according to these combined variables, which are similar to components in principal component analyses. A sparse version of this method is also employed(372), that penalises the variables that minimally contribute to differences between groups and allows for selection of genes that have the highest contribution towards difference between groups. PLS-DAs were generated using the package “mixOmics” in R(347).

4.3 Distribution of RSV disease severity among participants

4.3.1 Proportions of participants were hospitalised and required supplemental oxygen

Disease severity of all participants used in analysis was determined based upon invasiveness of respiratory intervention. A total of 30 of the participants recruited to the infant case control study whose samples were used in this mcicroarray analysis did not require any supplemental oxygen, while 62 required some form of supplemental oxygen, ranging from requirement of an oxygen face mask to invasive mechanical ventilation (fig 4.3.1.2). Criteria for invasive respiratory ventilation are described in 2.5.2. A full description of the entire cohort used in microarray analysis is given in figure 3.2.1. After removal of samples that did not pass quality control (as described in chapter 3.3.1). 6 samples in the case control study that were run on the microarray did not have available clinical meta data at the time of analysis, and were excluded from severity analyses, resulting in a final 92 samples used in the analysis in this chapter and the following chapter. Only one participant required a face mask alone, while 21 participants required low flow nasal cannula, none required high flow nasal cannula, 13 required CPAP/BiPAP, and 26 required mechanical ventilation. For statistical associations with gene requirement of supplemental oxygen by methods more invasive than low flow nasal cannula were considered as severe disease as described in methods 2.5. Using the former metric 30 participants were classed as having mild disease

and 62 were classed as having severe, while using the latter 53 were classed as having mild disease while 39 were classed as having severe.

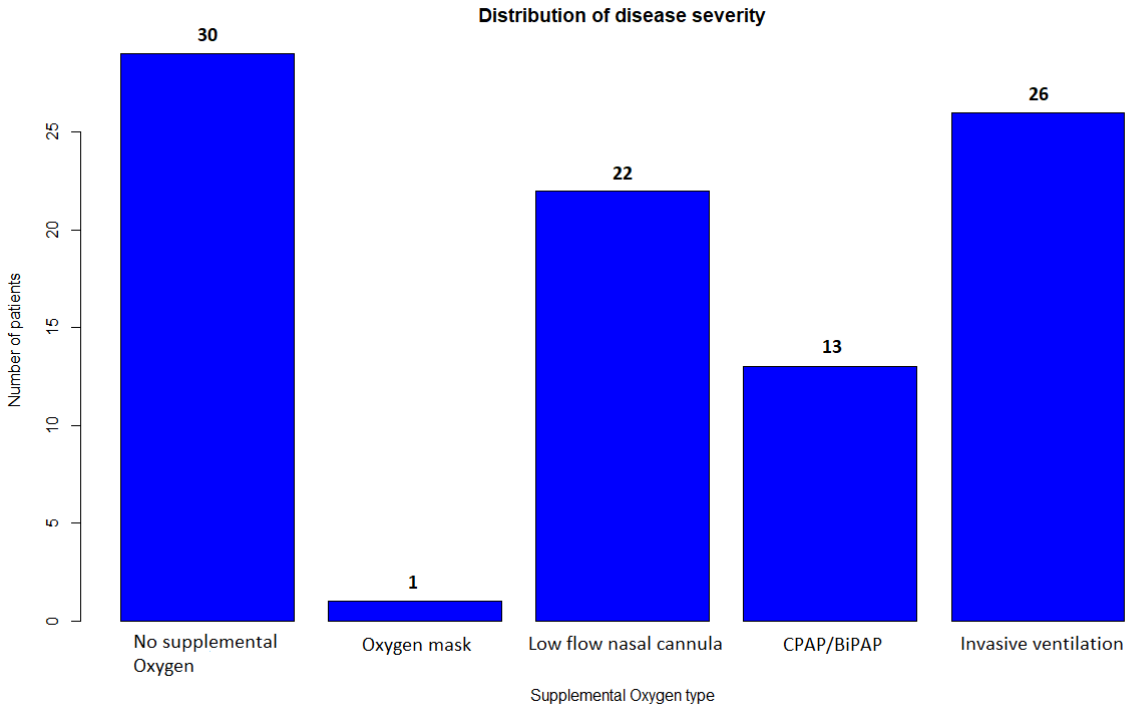


Figure 4.3.1.2 Distribution of participants recruited to the RESCEU case control trial according to methods of oxygen supplementation required. Different methods of oxygen supplementation were arranged in order of invasiveness from lowest to highest, and number of patients receiving supplemental oxygen by each method were given.

4.4 Correlations of gene expression with clinical disease

4.4.1 Neutrophil and Monocyte gene signatures correlate significantly with invasiveness of respiratory support

Ranked gene set enrichment analysis was performed (described in 3.2.6) on all 1707 genes found to be differentially expressed between acute disease and convalescence in 3.4.1. The genes were ranked by their correlation with disease severity according to the invasiveness of respiratory intervention to determine the most common biological functions among

genes that correlate with disease severity. By using the Pearson correlation coefficient value as the ranking value for gene set enrichment analysis, genes and processes that positively correlate with more severe disease and genes and processes that were associated with milder disease can be observed separately. Terms found to be enriched in more severe disease therefore be considered to be reflective of processes that are altered in more severe disease, while terms found to be enriched in milder disease are reflective of processes that are either protective against or just negatively regulated during more severe disease.

A total of 360 genes were significantly correlated with RSV disease severity defined according to the invasiveness of respiratory support, of which 228 were significantly positively correlated and 132 were significantly negatively correlated. The genes with the highest positive correlation with the invasiveness of supplemental oxygen treatment were *BPI*, *PGLYRP1* and *TDRD9*. The genes with the highest correlation with milder disease were *HLA-DPB1*, *HLA-DMB* and *HLA-DPA1* (fig 4.4.1.1).

The most highly enriched biological process terms in more severe disease were related to antibacterial responses, humoral responses, terms relating to modification of foreign organisms, and terms related reactive oxygen species and neutrophils, among general inflammatory responses, such as IL-1 production. The most highly enriched terms for milder disease were related to B cells, signalling downstream of antigen receptors, expression of viral genes and Ribosomes (fig 4.4.1.2).

The most highly enriched cellular component terms in more severe disease, were all related to neutrophil and granulocyte activity, such as specific granule lumen, tertiary granule membrane and tertiary granule, while the most highly enriched terms that correlated with milder disease were all related to ribosomes (fig 4.4.1.3).

The most highly enriched molecular function terms that correlated with disease severity were hydrolase activity, heparin binding and glycosaminoglycan binding, with sulphur binding and oxidoreductase binding also being significantly enriched. The terms that were significantly enriched in the milder disease related to binding of transcription factors to DNA and to ribosomes (fig 4.4.1.4).

The most highly enriched term positively correlating more severe disease related to *Streptococcus pneumoniae* infection. Other immune terms positively correlating with severe disease included terms related to a low ratio of B cells, NK cells $CD8^+$ and naive $CD4^+$ T cells to neutrophils, as well as terms related to viral infection. Immune terms correlating with milder disease included an increased B cell phenotype relative to monocytes and neutrophils, and increased expression of lymphocyte related genes relative to NK cell related genes, as well as lower expression of neutrophils relative to B cells and lymphocytes (fig 4.4.1.5).

Positively correlated genes			Negatively correlated genes		
Pearson correlation	FDR	Gene	Pearson correlation	FDR	Gene
0.532	4.63E-05	BPI	-0.439	0.000723	HLA.DPA1
0.525	4.63E-05	PGLYRP1	-0.438	0.000723	HLA.DMB
0.522	4.63E-05	TDRD9	-0.425	0.001032	HLA.DPB1
0.511	6.93E-05	DEFA4	-0.418	0.001199	ADGRG5
0.505	8.24E-05	OLFM4	-0.417	0.00123	HLA.DQA1
0.487	0.000173	CLEC5A	-0.4	0.001977	SLC30A4
0.486	0.000173	S100A8	-0.395	0.002278	CRIP1
0.485	0.000173	NSUN7	-0.393	0.002424	ZNF831
0.47	0.000334	MGAM2	-0.392	0.002424	HLA.DRA
0.469	0.000334	IL1R2	-0.369	0.005264	NUP88
0.469	0.000334	ARG1	-0.366	0.005738	FCMR
0.465	0.000349	LCN2	-0.363	0.00605	IKZF2
0.465	0.000349	MMP8	-0.362	0.00612	EVL
0.462	0.000349	CA4	-0.358	0.007119	IL32
0.461	0.000349	SMPDL3A	-0.355	0.007735	FUCA1
0.461	0.000349	MSRB3	-0.355	0.007792	ID3
0.459	0.000363	UPP1	-0.354	0.008071	YPEL1
0.453	0.000484	MANSC1	-0.352	0.008113	ADK
0.449	0.000553	RETN	-0.352	0.008113	CD74
0.448	0.000575	RAB27A	-0.352	0.008113	FAM214A
0.446	0.000591	RGL4	-0.351	0.008231	UGT2B11
0.444	0.000651	ADAM9	-0.351	0.008231	IMP3
0.439	0.000723	OMG	-0.351	0.008231	ATM
0.438	0.000723	PPM1N	-0.347	0.009105	KLRG1
0.438	0.000723	GLIPR2	-0.346	0.009138	TSPYL1
0.435	0.000784	SLC26A8	-0.346	0.009138	PLEKHA1
0.432	0.000903	FAM160B1	-0.342	0.010039	E2F5
0.43	0.000951	DHRS9	-0.342	0.010039	CD27
0.429	0.000951	CEACAM4	-0.342	0.010039	MALT1
0.428	0.000953	CEACAM8	-0.341	0.010039	CD2

Table 4.4.1.1 Table of the top 30 genes that were most positively and negatively correlated with graded requirements for supplemental oxygen, where participants requiring no supplemental oxygen were graded as a 0, receipt of oxygen via face mask as 1, low flow nasal cannula as 2, CPAP/BiPAP as 3 and mechanical ventilation as 4. Pearson correlations between gene expression and disease severity were shown along with FDR adjusted p values for each correlation.

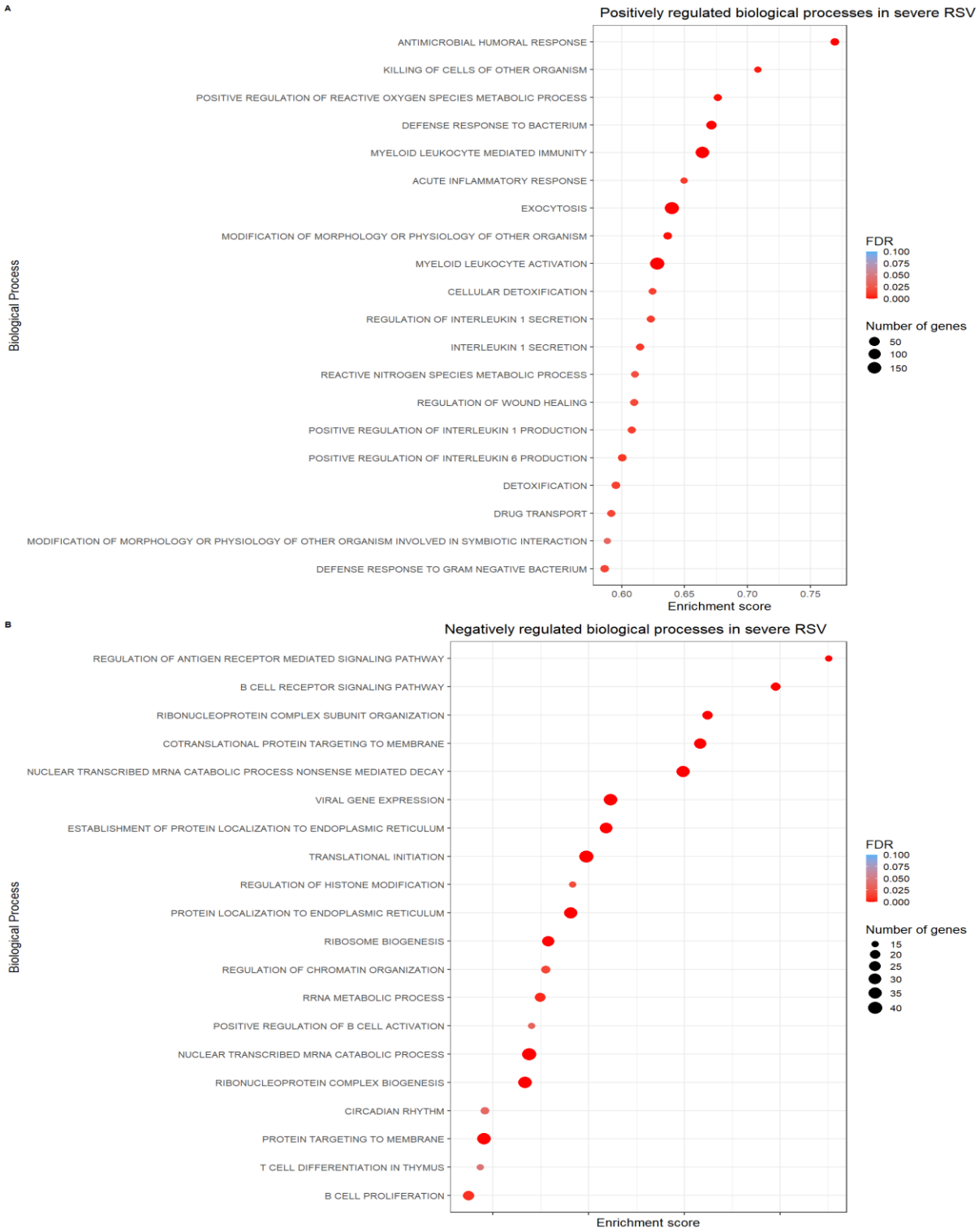


Figure 4.4.1.2 Gene set enrichment analysis of A – biological process terms correlated with more severe and B – biological process terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with invasiveness of supplemental oxygen treatment, ranked on a scale of 0-4- by invasiveness of oxygen supplementation required. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

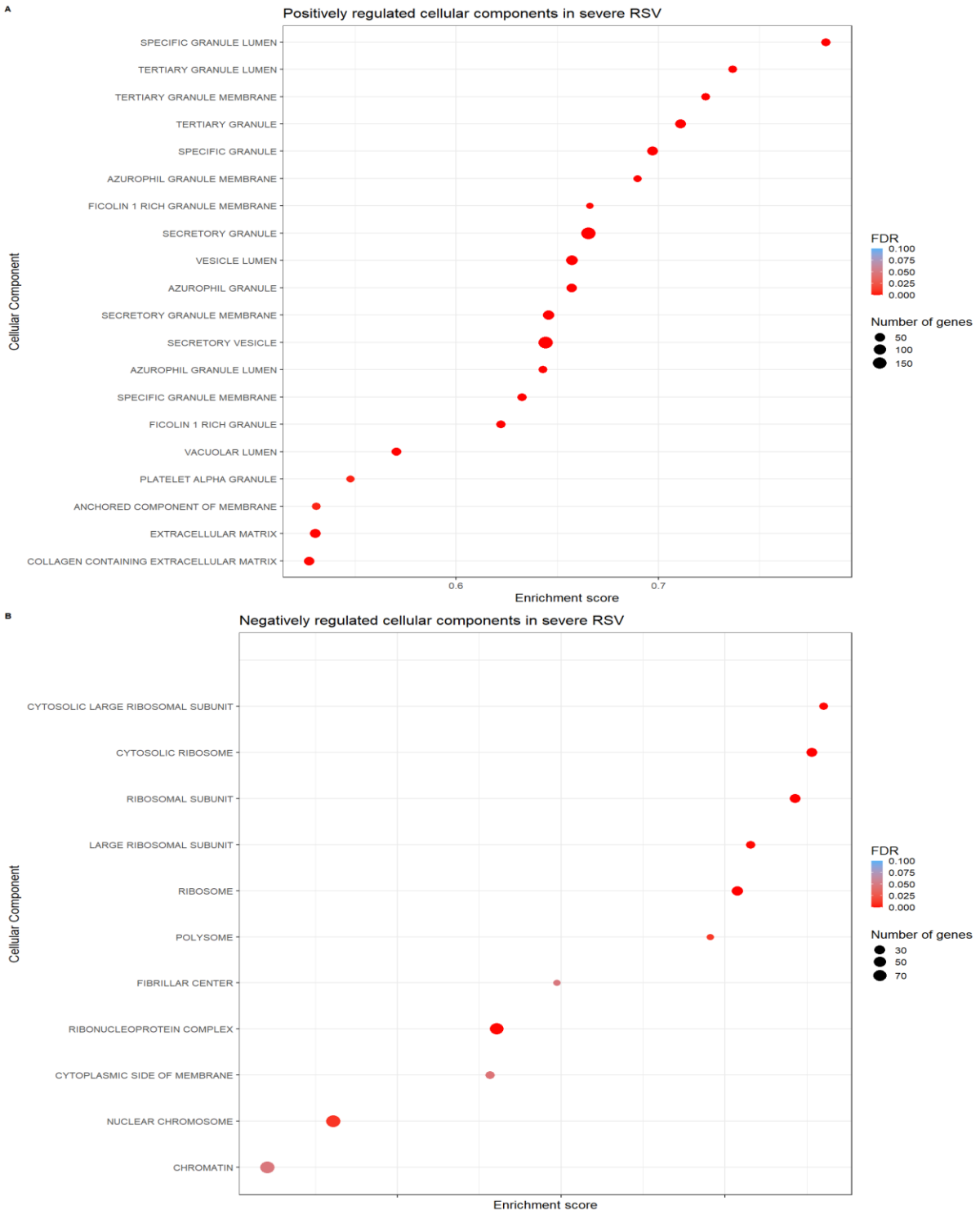


Figure 4.4.1.3 Gene set enrichment analysis of A – cellular component terms correlated with more severe and B – cellular component terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with invasiveness of supplemental oxygen treatment, ranked on a scale of 0-4 by invasiveness of oxygen supplementation required. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

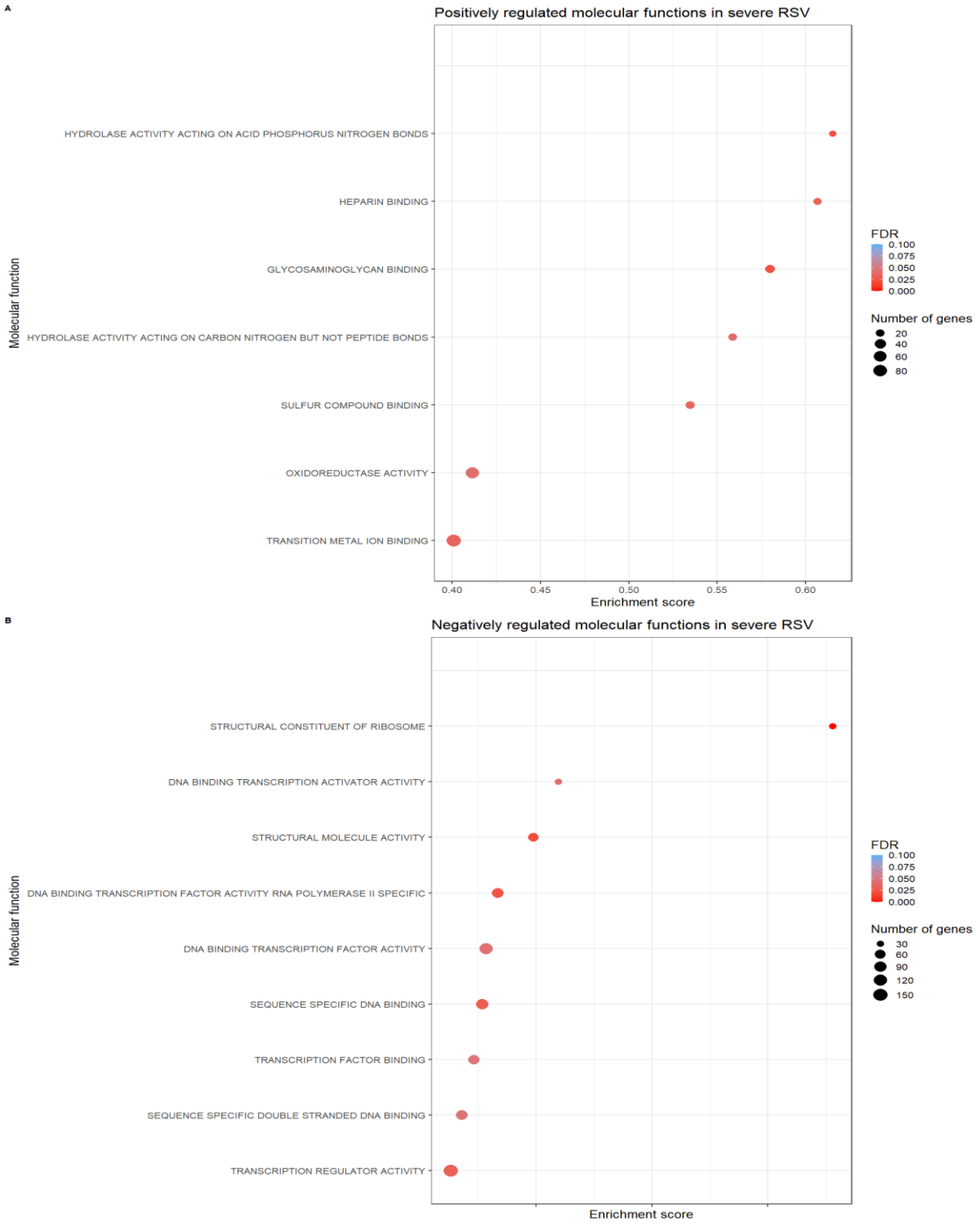


Figure 4.4.1.4 Gene set enrichment analysis of A – molecular function terms correlated with more severe and B – molecular function terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with invasiveness of supplemental oxygen treatment, ranked on a scale of 0-4 by invasiveness of oxygen supplementation required. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

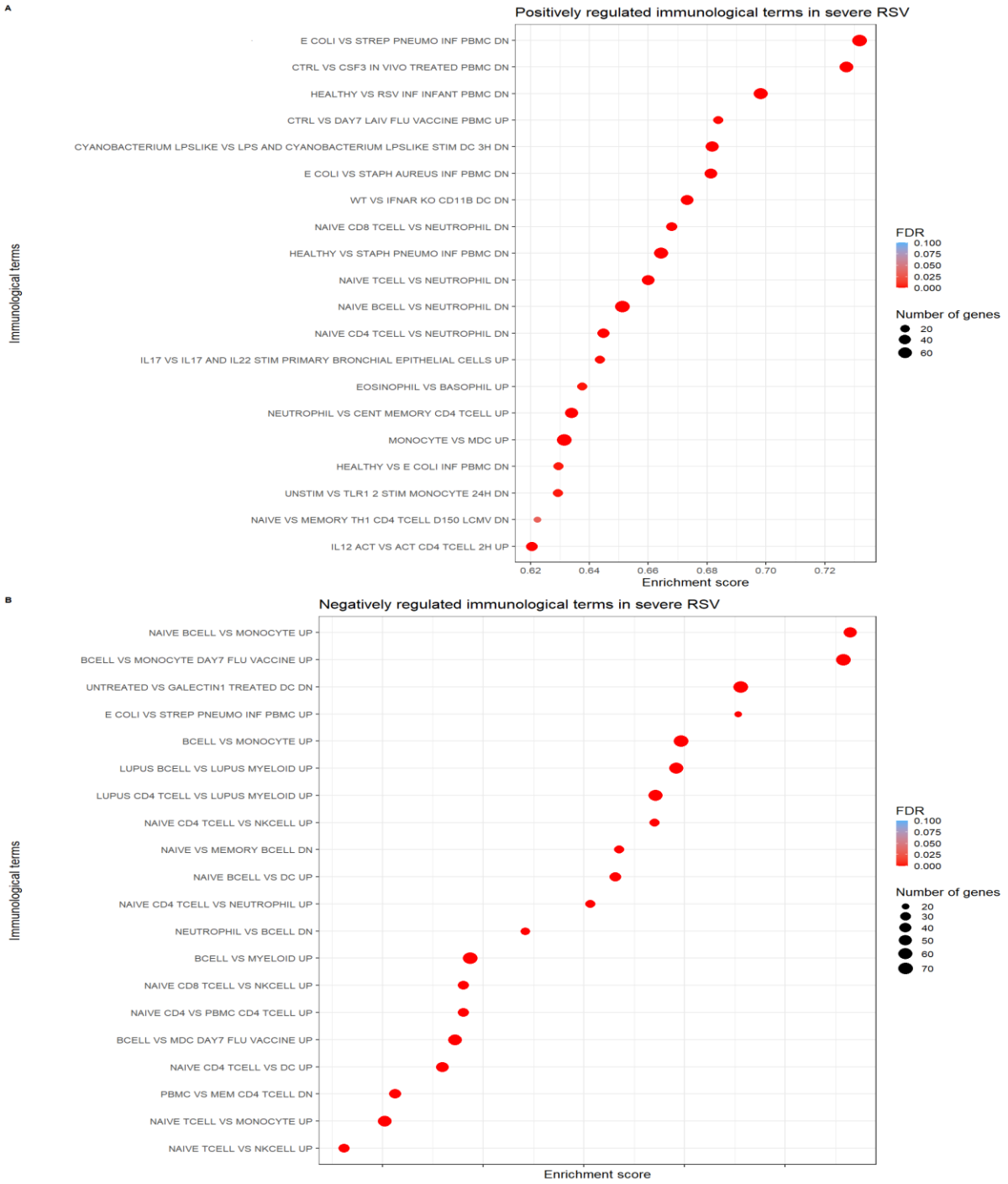


Figure 4.4.1.5 Gene set enrichment analysis of A – immunological terms correlated with more severe and B – immunological terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with invasiveness of supplemental oxygen treatment, ranked on a scale of 0-4. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

4.5 Patterns of gene expression are distinct between participants with severe and mild disease

To determine patterns of gene expression that associate most highly with disease severity, PLS-DA was used on the set of 1707 differentially expressed genes found in acute RSV infection. PLS-DA was described in 4.2.4. As PLS-DA finds the set of variables that were maximally different between different groups, participants had to be categorised as mild or severe prior to application of the algorithm. Because PLS-DA summarises highly correlated patterns of gene expression in components, it is possible to interpret the different arms of the immune system that associate with disease severity by using the loadings of the genes in different components of the PLS-DA. A severity threshold was defined using invasiveness of respiratory intervention (criteria for which are described in 2.5.2, where any participants receiving respiratory intervention more invasive than low flow nasal cannula (CPAP/BiBAP or mechanical ventilation) were considered to have severe disease, placing 39 (41%) of participants in the severe category. Maps of groupings of participants according to PLS-DA groupings were generated, and functional analysis was performed on genes contributing to the first two components of each PLS-DA.

This analysis was also performed using ReSVinet score as the metric for disease severity, described in appendix 2.4.

4.5.1 Genes on first component of invasiveness of respiratory intervention PLS-DA associate with upregulated neutrophils and monocytes, and downregulated lymphocytes

The loadings of all 1707 differentially expressed genes used in PLS-DA analysis were used to rank genes for gene set enrichment analyses. The first component identified by PLS-DA using invasive respiratory intervention as the determinant of severity comprised of 12% of the variation in the set of differentially expressed genes. An FDR adjusted p.value cut-off for significance of .05 was used as a threshold for significance of an enriched term in these analyses.

The most highly enriched biological processes in more severe disease (Fig 4.5.4.1 A) were related to antimicrobial humoral responses, antibacterial responses, wound healing and

neutrophil related terms , while the most highly enriched biological processes in mild disease (Fig 4.5.4.1 B) were related to ribosomes, B cells, CD4 and CD8 T cells, viral expression and translation.

The most highly enriched cellular components in severe disease (Fig 4.5.1.2 A) were all related to neutrophils, while the most highly enriched cellular components in mild disease (Fig 4.5.1.2 B) were related to ribosomes.

The most highly enriched molecular functions in severe disease (Fig 4.5.1.3 A) were hydrolase activity on acid phosphorous nitrogen bonds, while the most highly enriched molecular functions in mild disease (Fig 4.5.1.3 B) were related to RNA binding, ribosomes, and structural molecules.

The most highly enriched immunological terms in severe disease (Fig 4.5.1.4 A) were related to low ratios of lymphocytes to neutrophils and monocytes, responses to *Streptococcus pneumoniae* and *Staphylococcus aureus*, while the most highly enriched immunological terms in mild disease (Fig 4.5.1.4 B) were related to a high ratio of memory lymphocytes to naïve, a high ratio of lymphocytes to myeloid cells and similarities to healthy controls in viral infection.

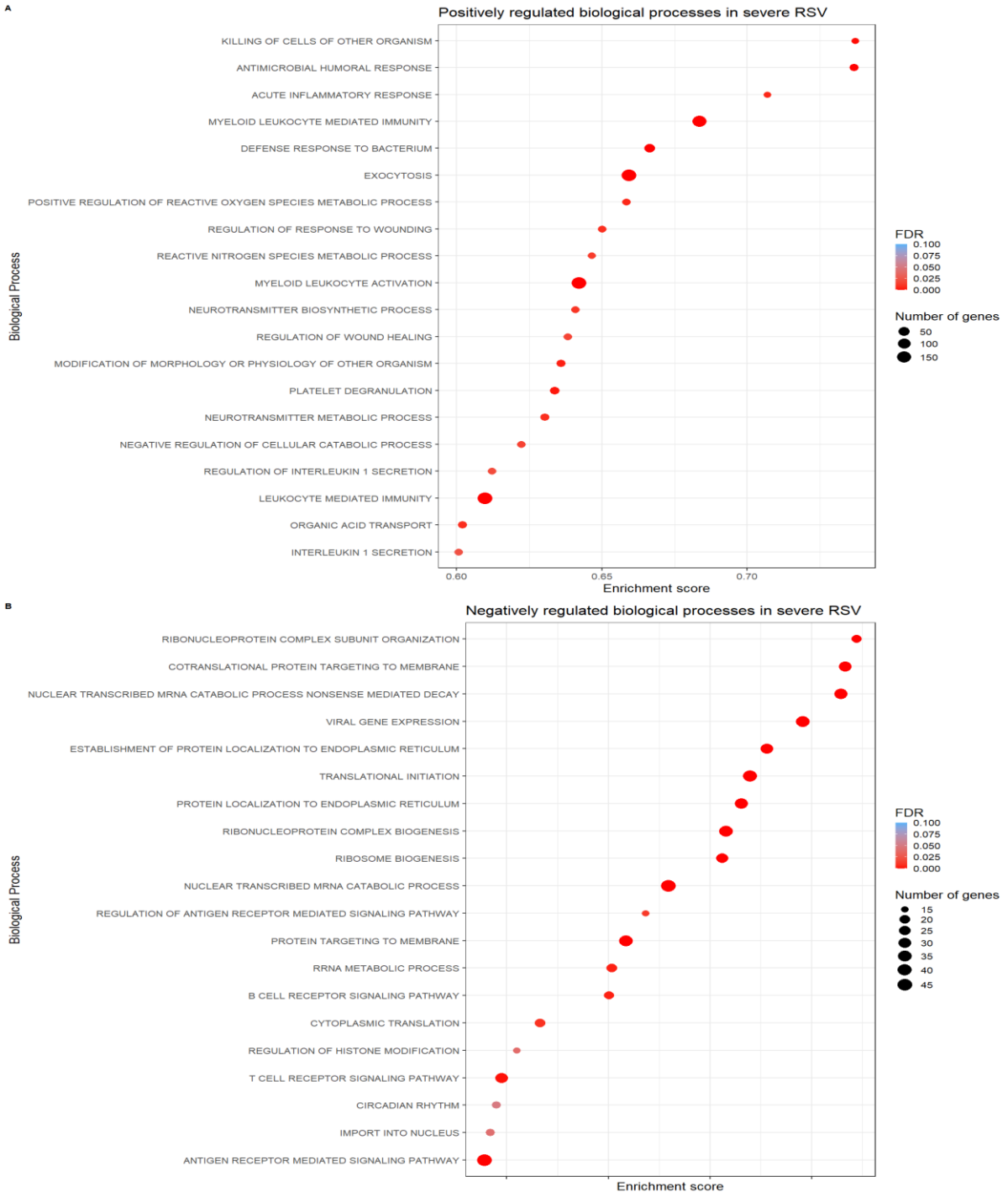


Figure 4.5.1.1 Gene set enrichment analysis of A – biological process terms correlated with more severe and B – biological process terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the first component of the invasiveness of respiratory support disease severity PLS-DA. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one

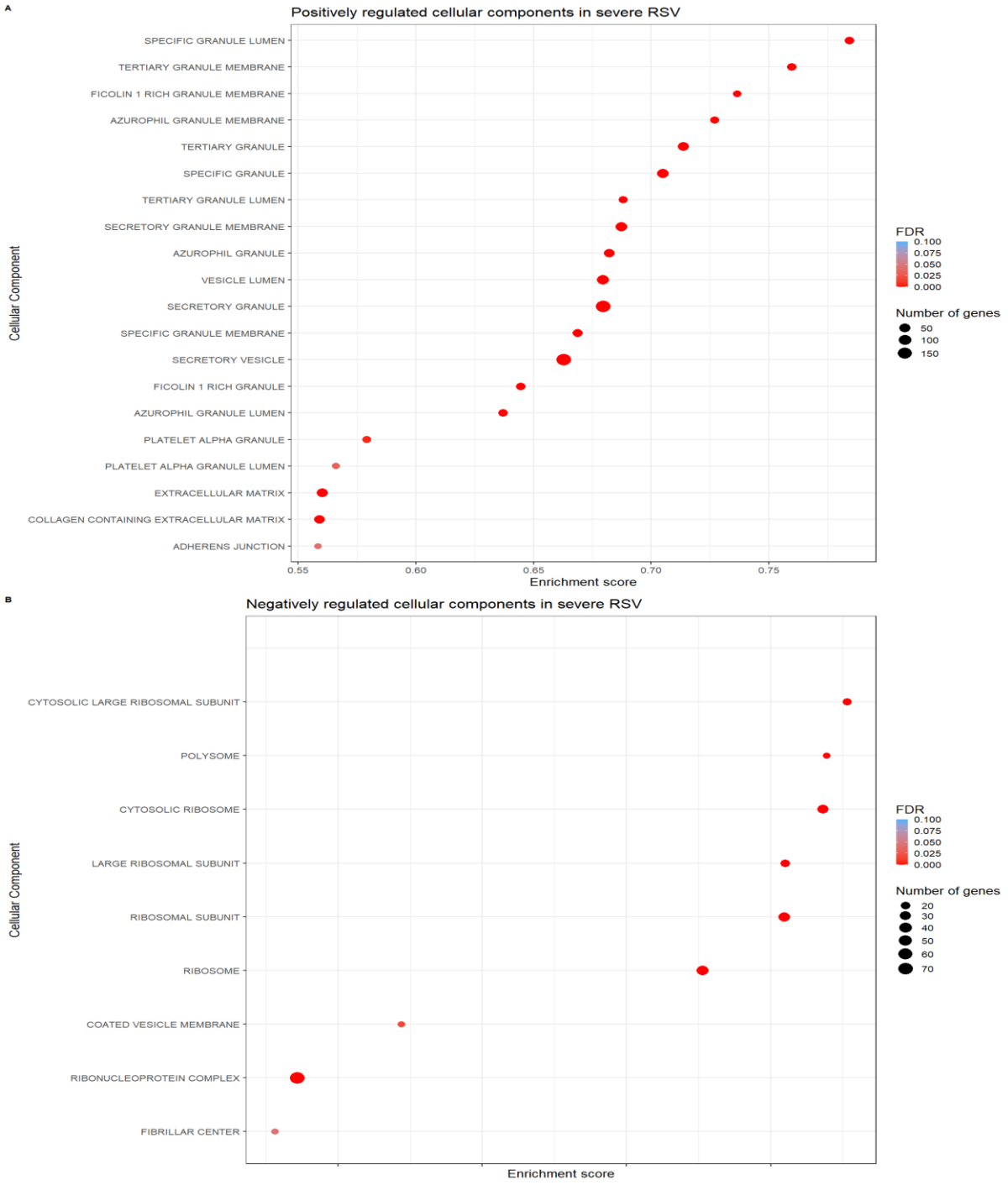


Figure 4.5.1.2 Gene set enrichment analysis of A – cellular component terms correlated with more severe and B – cellular component terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the first component of the invasiveness of respiratory support disease severity PLS-DA. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one

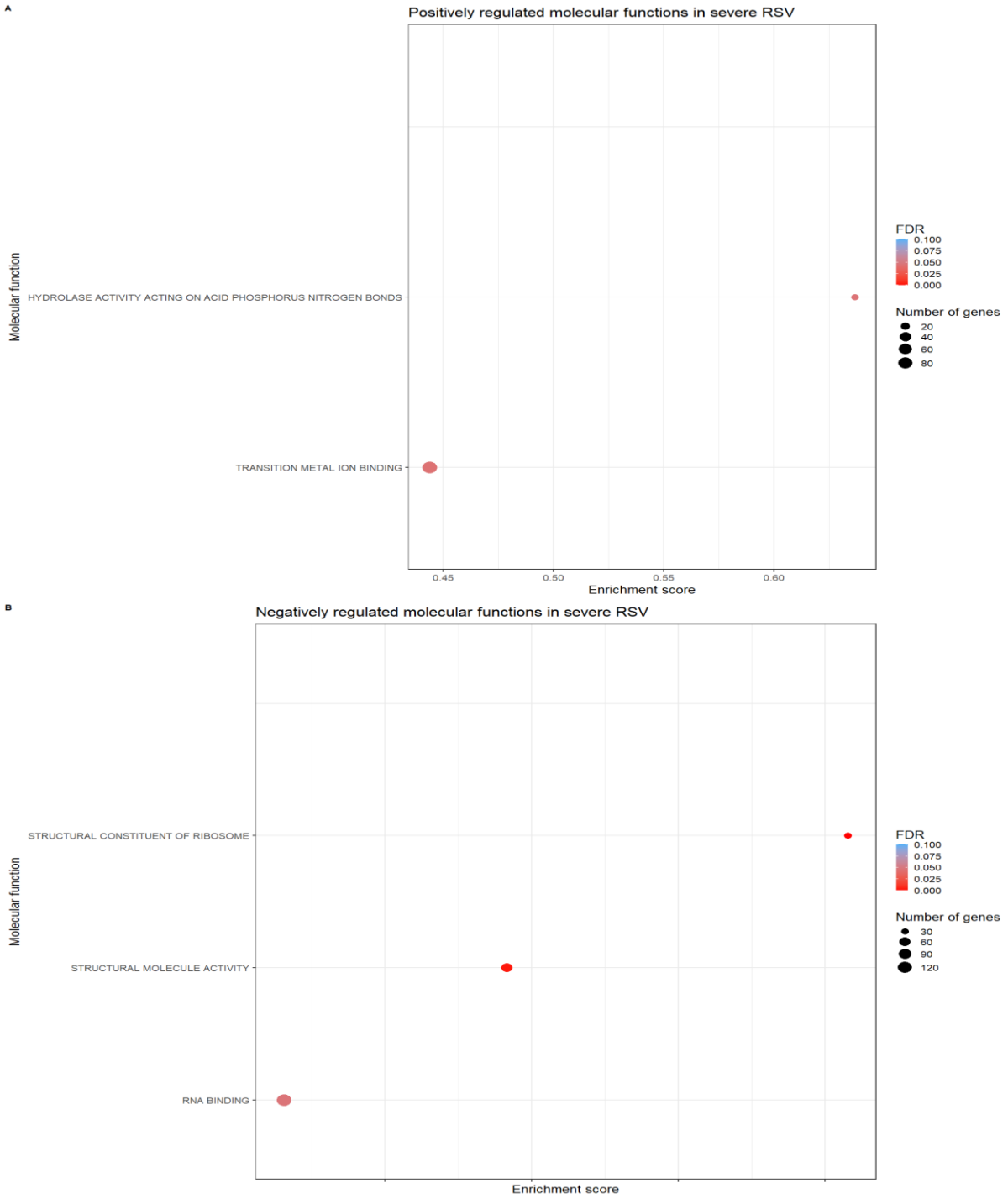


Figure 4.5.1.3 Gene set enrichment analysis of A – molecular function terms correlated with more severe and B – molecular function terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the first component of the invasiveness of respiratory support disease severity PLS-DA. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one

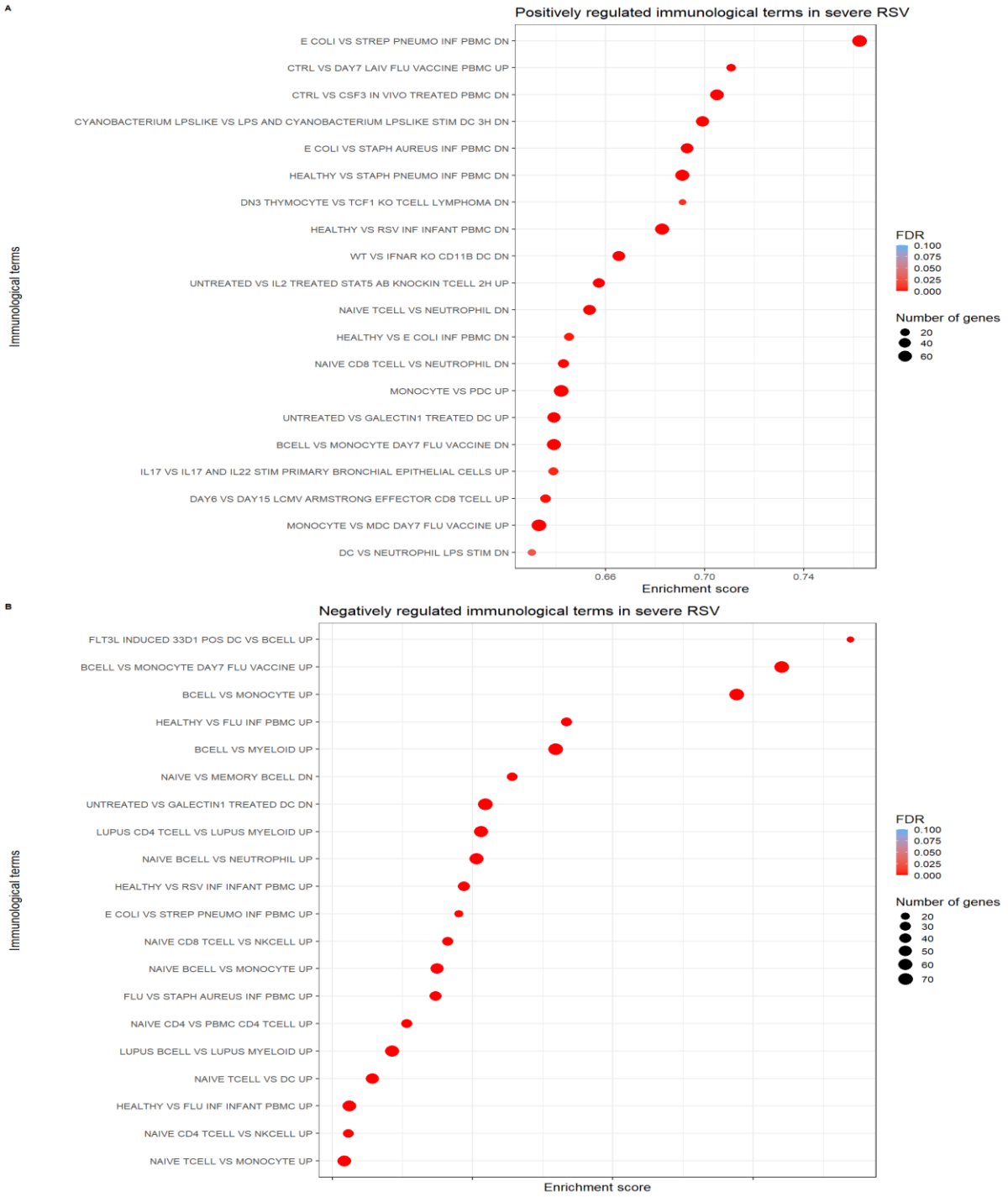


Figure 4.5.1.4 Gene set enrichment analysis of A – immunological terms correlated with more severe and B – immunological terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the first component of the respiratory support invasiveness disease severity PLS-DA. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

4.5.2 Genes on second component of invasiveness of respiratory intervention PLS-DA associate with upregulated neutrophils and monocytes, and downregulated lymphocytes

The loadings of all 1707 differentially expressed genes (described in 3.4.1) used in PLS-DA analysis were used to rank genes for gene set enrichment analyses. The second component identified by PLS-DA using invasive respiratory intervention as the determinant of severity comprised of 4% of the variation in the set of differentially expressed genes.

There were no significantly enriched biological process terms in the second component.

There were also no significantly enriched cellular component terms for mild disease in the second component, and only a single significantly enriched term for severe disease (Fig 4.5.2.1), which was coated vesicle membrane.

There were also no significantly enriched molecular functions related to mild disease in the second component, and only a single significantly enriched term related to severe disease (Fig 4.5.2.2), which was phosphatase binding.

Immunological terms that were associated with more severe disease (Fig 4.5.2.3 A) were also related to upregulation of terms related to days 3 and 7 post YF17D vaccination, and downregulation of genes related to day 1 post vaccination. Other significantly enriched terms that correlated with severe disease included a low ratio of lymphocytes to neutrophils and monocytes, and a low ratio of T-follicular helper cells to other CD4 T cells, and signals relating to IFNγ signalling. Immunological terms that were associated with less severe disease (Fig 4.5.2.4 B) related to a higher ratio of baseline lymphocyte to activated lymphocytes, as well as a high ratio of lymphocytes to monocytes, and significantly increased B cell expression compared to baseline.

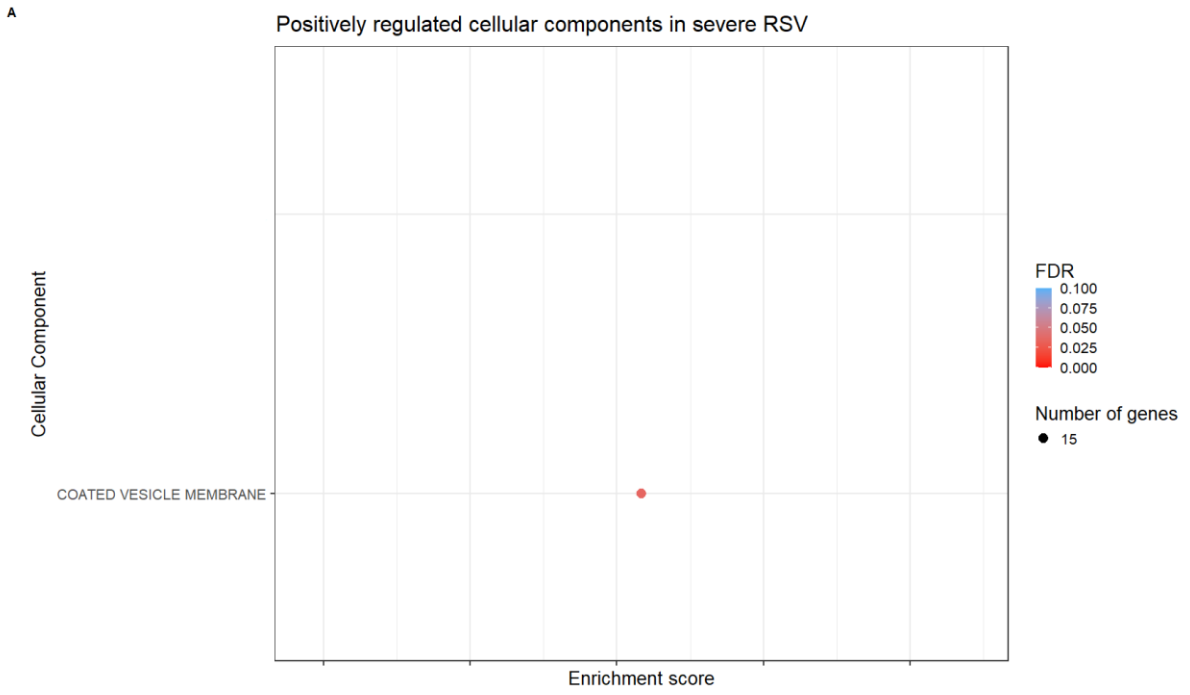


Figure 4.5.2.1 Gene set enrichment analysis of A - cellular components associated with more severe from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the second component of the invasiveness of respiratory support disease severity PLS-DA. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one

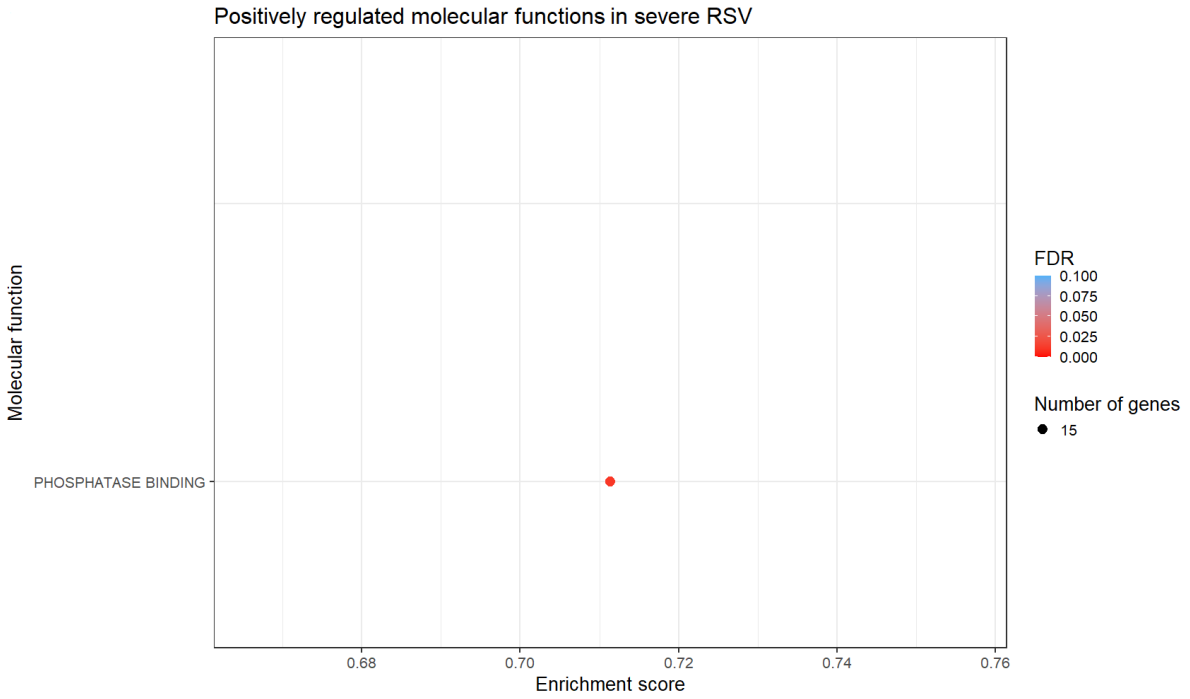


Figure 4.5.2.2 Gene set enrichment analysis of A - molecular function terms correlating with more severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the second component of the invasiveness of respiratory support disease severity PLS-DA. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one

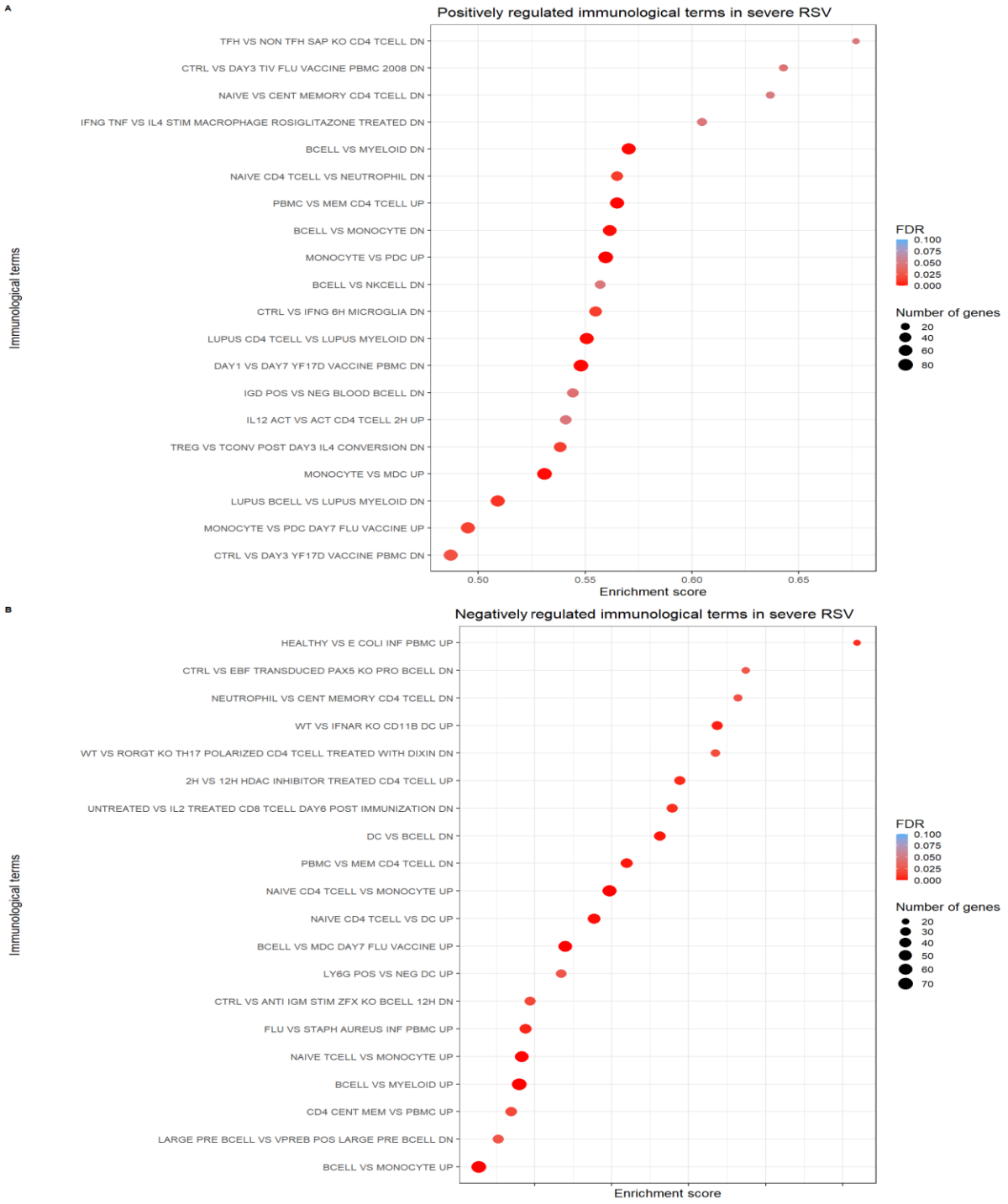


Figure 4.5.2.3 Gene set enrichment analysis of A – immunological terms correlated with more severe and B – immunological terms correlated with less severe disease from set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the second component of the respiratory support invasiveness disease severity PLS-DA. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

4.6 Discussion

4.6.1 Summary of main findings

Disease severity of RSV was statistically associated with higher expression of neutrophil and monocyte related genes, and lower expression of ribosomal and lymphocyte related genes (and of B cell related genes in particular). Association between gene transcription and less severe disease may be attributable to this ribosomal response, although a causative relationship is not clear. It is possible that lower expression of ribosomal genes relates to less severe viral infection as the virus has a lower capacity to replicate where there is a lower expression of ribosomal genes(352). Genes relating to humoral immunity were also significantly associated with more severe RSV disease, likely due to the association observed between the secretory products of neutrophils and disease severity. Additionally, the genes that associate most highly with disease severity also associate significantly with patterns of gene expression induced in response to some bacterial infections, specifically *Streptococcus pneumoniae*, proliferation of which has previously been shown to associate with more severe RSV disease(239) and Influenza(373).

Patterns of gene expression that differed most highly between participants with more severe disease and those with less severe disease were quantified using PLS-DA. The ranked loadings of the that were genes used to distinguish severe and mild disease using this model that showed an association between low lymphocyte associated gene expression relative to neutrophil and monocyte associated gene expression was associated with more severe disease, along with skews towards humoral and antibacterial responses.

4.6.2 Correlations between gene expression and disease severity

The gene most highly enriched terms in genes that statistically associate with more severe disease related to increased expression of neutrophil (reactive oxygen and nitrogen species related terms), eosinophil (granulocyte related terms), monocyte (similarities to monocyte expression profiles) and humoral immunity (antibacterial response terms, neutrophil granule terms) related genes, and decreased expression of T cell and B cell related genes. These terms were similar to those that were found in significantly differentially expressed genes between acute RSV infection and convalescence. These associations are consistent with what has been previously observed, where neutrophils(61) and monocytes(62) have

been linked to more severe disease, and lymphocytic responses(145) have been linked to less severe disease. Reactive oxygen species (ROS) released by neutrophils have been shown to damage lung tissues(374). Various granulocyte associated proteins such as MPO have been shown to encourage the release of ROS(375), suggesting one potential role for humoral neutrophil products in contributing towards severe disease, however the effects of individual secreted neutrophil products on viral disease severity remain poorly characterised. The results shown here demonstrate that it is a potential area of interest for further study.

The antibacterial response terms observed in this study are also likely a product of the strong association between neutrophil granule related gene expression and disease. There is evidence that commensal bacteria, particularly *Streptococcus pneumoniae* undergo extensive replication during RSV infection(239), and that this replication correlates highly with neutrophil influx into the lungs and increased expression of anti-bacterial factors by neutrophils such as AZU1 and MPO(239). It is possible that this pathogenic anti-bacterial response (and possibly the dominant neutrophil responses seen in RSV and Influenza in general) primarily act to control this extensive bacterial replication, which has been shown to be pathogenic in the aftermath of both severe RSV and Influenza infections(376).

4.6.3 Patterns of gene expression that have significant association with RSV disease severity

PLS-DA analysis allows for the discovery of the most maximally different patterns of gene expression between groups of participants. Because PLS-DA maximises the difference between groups, it has a tendency to overfit and create gene signatures that are overly specific to the population of the study(377)(378), and which are not necessarily generalizable to a broader population. Because of this, it is important that other algorithms are employed in biomarker discovery, to ensure robustness and generalisability of gene signatures. In this analysis PLS-DA is used in exploration of patterns of highly co-correlated associated with disease severity, and thus issues of overfitting are not relevant. Identification of a robust predictive signature will be the subject of the following chapter, chapter 5.

In this chapter PLS-DA was used for exploratory analyses, to determine the patterns of gene expression that differ most highly between participants with milder and more severe disease. Results were presented from the first two components, which represent two sets of highly co-correlated genes (each of which does not correlate with any other component) that are maximally different between mild and severe disease.

The patterns of gene expression associated disease severity observed when ranking gene set enrichment analyses by PLS-DA loadings were similar to those using correlation with disease severity as rankings, but there were some notable differences, particularly in the second component.

First component terms for both severity categorisation methods were similar, and revealed strong associations with humoral responses, myeloid cells, and neutrophil responses. In conjunction with the associations with *Streptococcus pneumoniae* related immunological signatures, this suggests that the first component may be primarily characterised by strong innate myeloid responses, that likely contribute towards more severe disease due to a lack of capacity to properly control viral infection. The genes on the first component also consistently linked CD4 T cells, CD8 T cells and B cells with less severe disease.

Fewer terms were found to be significantly enriched when using loadings from the second component of both disease severity metrics. Where there were few significantly enriched terms it was difficult to infer distinct biological functions for genes in this component, however there were consistently more significantly enriched terms in the immunological database, which made it possible to draw conclusions about functions of genes in the second component.

The second component of PLS-DA shows both IFN-γ, IL-4 and activated and memory CD4 T cell type responses associate with more severe disease, but also that B cell and naïve CD4 T cell type responses, correlate with less severe disease. This contradicts much of the traditional associations that have been observed between RSV disease and lymphocytes, and indeed the associations seen in this study on other components. Increases in infiltration of CD4 T cells have been observed in the bronchioalveolar lavage of patients experiencing an LRTI(61), as well as populations of activated CD4 T cells(379), although their role during the peak of disease remains poorly characterised. RSV is capable of infecting CD4 T cells, and

populations of activated RSV positive CD4 T cells have been shown to correlate with disease severity(380). It is possible that these results (and in particular the contradiction of what is seen on the first component) are due to a subset of RSV infected lymphocytes (or a subset of lymphocytes with different function) that associate with more severe disease, but which account for a much lower proportion of variation in the data set, and thus are difficult to observe due the more dominant protective associations between severe disease and lymphocytes seen in the first component. Populations of virally infected cells with a different profile of expression have been observed previously. IL-10 expressing B cells infected with RSV have been observed(82).T cells infected with herpesvirus have a distinct pro-inflammatory gene expression pattern(381).

While many terms in this study associate very consistently with severe disease (in particular terms relating to neutrophil and monocyte gene expression), this does not mean that the cell types associated with these patterns of expression do not have some protective roles in RSV. In RSV infection neutrophils may have some role in controlling infection while contributing towards respiratory immunopathology. Neutrophils have been shown to aid in recovery from severe influenza infection in the mouse model(382) and defensins released from human neutrophils have been shown to encourage phagocytosis of Influenza(383), suggesting an early neutrophil response may help to control and limit the scope of infections that do not develop into severe disease. While being an important component in the pathogenesis of RSV, neutrophils do have some protective roles, though these have yet been demonstrated in human RSV infection.

4.6.4 Limitations

The sampling and time points in this study were not designed to capture either gene expression or disease severity at the peak of disease. Instead they represent the time of presentation in a hospital. As this time point is variable, gene expression data for each participant is at a slightly different time and is not reflective of the peak of severity. As the populations of cells change over the course of an RSV infection, this may prove in some ways to be a benefit, as the mechanistic effects of cell types both before and during the peak of disease may be assessed.

As gene expression data were only taken early in acute infection, the role of factors that emerge later in infection on severity during acute infection (such as antibodies, which

typically emerge two to four weeks after infection or vaccination(384)) are underrepresented. Within the RESCEU study this will be compensated for in future analyses by integration of neutralising antibody titre data for all participants. A study where sampling and disease measurements are conducted each day over the course of infection to analyse how the changes in cell populations and gene expression effect the ongoing development of disease would be an ideal method by which to answer these questions. Human challenge studies may also provide valuable information about the development of the immune response and disease severity over time. Due to ethical issues RSV human challenge studies have been conducted in adults(385), however due to allowing for baseline sampling directly prior to disease, and allowing for sampling and disease severity measurements on each day of infection, challenge models provide an invaluable model for development of disease over time.

The correlations between gene expression and disease shown here do not necessarily reflect a role for those genes in RSV disease. In the case of each of these relationships some functional analysis will be required to determine if these changes in gene expression are directly related to disease severity or are merely side products of an excessive harmful inflammatory response. Traditionally animal models have been used to provide answers to these questions, as gene knockouts and transgenic animals allow for the examination of the roles of specific genes in disease. The relevance of T lymphocytes and antibodies in animal disease is reviewed in chapter 6 of this thesis.

4.6.5 Conclusions

Neutrophil and monocyte responses have a strong relationship with more severe disease, while CD4 T cell, CD8 T cell and B cell responses are consistently associated with milder disease. PLS-DA analyses also revealed that Th1 type signalling profiles associated with milder disease, that was masked in correlative analyses by a strong myeloid response that associated more strongly with disease severity.

For the purposes of developing disease biomarkers, the robust statistical associations observable at time of presentation with disease are ideal, as they allow for prediction of progression of disease throughout the course of an acute infection. While PLS-DA is an algorithm that is very well suited to exploratory and functional analyses(377), as maximally different sets of genes are useful for characterising disease states, it is desirable to use a

broader set of algorithms due to the tendency of PLS-DA towards overfitting to the data set on which it is trained. The usage of a panel of statistically robust algorithms and biomarker discover methods (such as cross validation of samples, and employing training, validation, and test sets) are further investigated in chapter 5 of this thesis.

Chapter 5: Biomarkers for severe RSV disease

5.1 Introduction

In chapters 3 and 4 I showed the distinct signatures of gene expression that associate with acute RSV disease, and that associate with RSV disease severity. In the current chapter I will show the development of a minimal, predictive biomarker from whole blood transcriptomic data for RSV disease severity.

Machine learning is the study of and application of algorithms that can be automatically improved through exposure to data. These algorithms allow for visualisation and quantification of patterns and relationships in large bodies of data that may be extremely difficult to detect and quantify by other means, while also providing a set of tools that allow scientists to easily utilise these patterns in predictions (for example predicting relationship between gene expression responses and immunological outcomes) and in understanding the patterns that drive clinical outcomes. To do this, machine learning algorithms are often trained on a set consisting of the majority of participants available (often 75-85% of all participants in a given study, with the remaining percentage being reserved as a "test" set). The algorithms are provided with data as well as labels for those data to develop the model and uses repeated cross validation (by internally specifying multiple conformations of training and validation sets within the "training set"). The trained model can then be used to predict the labels of unlabelled samples. The success of the model in prediction in unlabelled samples can be used to assess its efficacy(386). This model is then tested in the remaining "test set" for its ability to correctly predict the outcome measure. Ideally, this model will then be evaluated in an independent study cohort to assess its generalizability.

Disease severity was assessed contemporaneously with blood samples, and so gene expression could be used to develop a clinical biomarker that could be of use in discriminating between participants with a higher risk form of disease and those with a lower one. The metric used for severe disease was requirement for invasive ventilation, due to its advantages as a severity metric discussed in chapter 4.

A transcriptomic biomarker for disease severity ideally consists of a small number of genes, the expression of which can be tested quickly and in an easily accessible tissue, and which provides information that can be used to correctly identify severity of disease such that appropriate treatment can be applied quickly. As the most severe RSV cases are often treated with respiratory intervention, a score aiming to predict the likelihood of an infection requiring that would be ideal for this purpose. This biomarker would need to provide clinicians with information that would aid in stratifying patients in a quick and effective manner that would improve confidence in the decision to rule out PICU admission and invasive respiratory support.

In this chapter I aim to isolate a biomarker for more severe RSV, and to validate this biomarker using publicly available data. To accomplish these aims I will first utilize an unbiased PCA based feature selection process to select for genes with the highest variability in the data set by selecting genes from the components that account for the highest amount of variation in the data set, and then demonstrate the efficacy of a panel of machine learning algorithms in predicting more severe RSV infection using selected genes. I will then systematically remove genes with low variation, to isolate an effective model with a minimal number of genes and validate the predictive power of that model using publicly available data.

5.2 Methods

5.2.1 SIMON

SIMON (Sequential Iterative Modeling “OverNight”) is a machine learning platform that simultaneously runs up to 128 machine learning algorithms. Machine learning models are run through SIMON and performance metrics for each model (ran at a single pre-selected set of hyperparameters) are shown(387). A proportion of the data set was reserved prior to running for performance testing. Models were subjected to tenfold cross validation repeated three times and the best model selected. SIMON produces ranking metrics of variable importance in each model.

5.2.2 Area under the receiver operating characteristic curve (AUC) analysis

Area under the curve analysis is one of the most often used methods of assessing performance of binary categorisation models where numbers of samples in each category are roughly equal(388). The receiver operating characteristic curve is a plot of 1 – specificity (true positive rate) on the x axis against the sensitivity (true negative rate) on the y axis. Categorical machine learning models often produce a probability value between 0 and 1. A threshold is usually assigned at some point between these two numbers such that if the probability assigned for a sample falls above the threshold it is considered as one category, and if it falls below it is considered another. How well a model assigns true positives and false positives is dependent on where this threshold is set. In a receiver operating curve these rates are calculated for a range of probability thresholds. The area under the curve is calculated by calculating the sum of the area between the x-axis and a line connecting each two adjacent data points, and serves as an assessment of the performance of the algorithm as a whole, independent of individual probability thresholds.

The AUC statistic ranges from 0 to 1, where a score of one reflects correct identification of every participant, 0 reflects incorrect identification of every participant and a score of .5 reflects correct identification of participant 50% of the time (the same as if identification were random).

5.2.3 PCA feature selection

Principal component analysis was employed to select features in an unsupervised manner, as described in Song et al(389). The “elbow” of the scree plot (the point where the curve of the scree plot turns from steep to shallow) was located, and all components falling above it were considered for further analysis and the top 5% of all genes in each of these top components were selected. The loading of a gene represents the proportion of the variance occurring within a component accounted for by that gene. The loadings (according to the percentage of the total variation accounted for by a given component) for of the top 5% of genes in each of these components were combined into a matrix of loadings and scaled to each other to ensure that the largest and smallest values of loadings from each component were identical while retaining the distribution of loadings from the original PCA. The

loadings from each component were then scaled according to the proportion of the total variance in the data set accounted for by their component of origin. The lists of genes with the scaled loadings were combined into a single list of all genes in the top 5% of each component, ranked by the absolute weighted loadings of each gene. The new list functions as a ranked list of all genes that contribute the most to the variance observed in the data set. This list was used to select genes for use in downstream machine learning analyses. Filtration of genes to reduce size of gene set was performed by filtering genes based on their rank in this list. Method is described in figure 5.2.3.1 A. Usage of this method for biomarker discovery is described in figure 5.2.3.1 B. Lower ranked genes were iteratively removed and the remaining gene list used in SIMON analysis until no models had an AUROC of over .75.

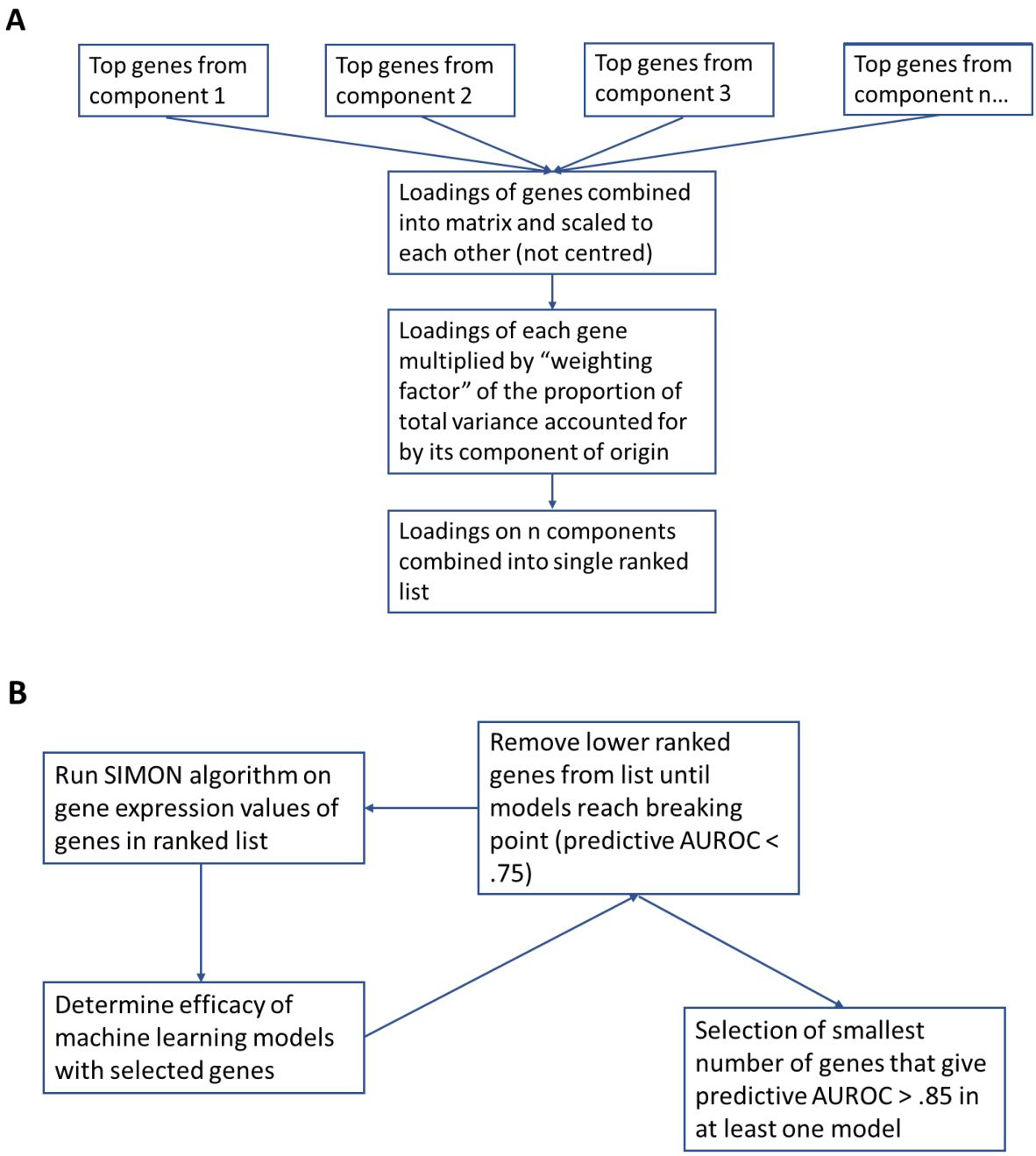


Figure 5.2.3.1 Flow chart of feature selection methods demonstrating A: methods by which genes are ranked for selection of relevant features in an unbiased manner and B: the use of this method in biomarker discovery.

5.2.4 Validation in published data

The corresponding data set for the study by Heinonen et al(285) was used for validation of this data set in an independent cohort. This data set was chosen due to similarities in method (microarray), outcome metrics (one of the components of the GRSS score used in this study was requirement for CPAP/BiPAP or a more invasive form of respiratory support)

and a comparable population (children between the age of 0 and 24 months, children older than 12 months were removed from this study before validation). Severe disease was defined as having a GRSS score for respiratory intervention of 3, which corresponds to requirement of CPAP/BiPAP or a more invasive method of respiratory intervention. Gene expression data for the validation study were normalized together to ensure comparability. ComBat batch correction was used to correct for differences between the two arrays. After model was trained on study data, model was applied to the independent validation cohort, and performance metrics were plotted.

5.3 Preparation of gene set

As a biomarker must be able to predict disease based on expression of a single sample at a single time point, only samples corresponding to cases of acute RSV infection were included, while healthy controls and convalescent samples were excluded. Quality control was performed on this subset of samples in the same manner as described in chapters 3.2 to 3.4, including sex chromosome filtration, normalization, and batch correction. This analysis was performed on the subset of 92 samples taken at the time of acute infection with available severity data described in chapter 4.3.1. Data of participants with acute infection prior to quality control are shown in appendix 3.1.

Principal component analysis (as described in chapter 3.2.7) was performed on the data set at this point and showed that batch effects due to microarray batch (fig 5.3.2 A) and due to cDNA yields (fig 5.4.2 B) were removed from the data. Principal component analysis also showed that after batch correction clustering according to disease severity occurred on the first and second principal components (fig 5.3.3 A, fig 5.3.3 B). The loadings of genes in the first three principal components were used to prepare a ranked list for feature selection as described in 5.2.3. Rankings of genes according to these loading for each component are shown (fig 5.3.4). The genes that ranked most highly were *IL1R2* and *TFDP1* on the first component (fig 5.3.4 A), *OLFM4* and *JMY* on the second component (fig 5.3.4 B) and *TMEM259* and *IRF1* on the third component (fig 5.3.4 C). The absolute loadings of genes on the first three principal components were scaled to each other, and used as a ranked list, where genes with the lowest loadings would be considered less important, and could be filtered out when reducing size of the gene set for biomarker discovery. The top 821 genes

from each component (the top 5% of genes from each component) were included in the ranked list, resulting in that list containing a total of 2463 genes.

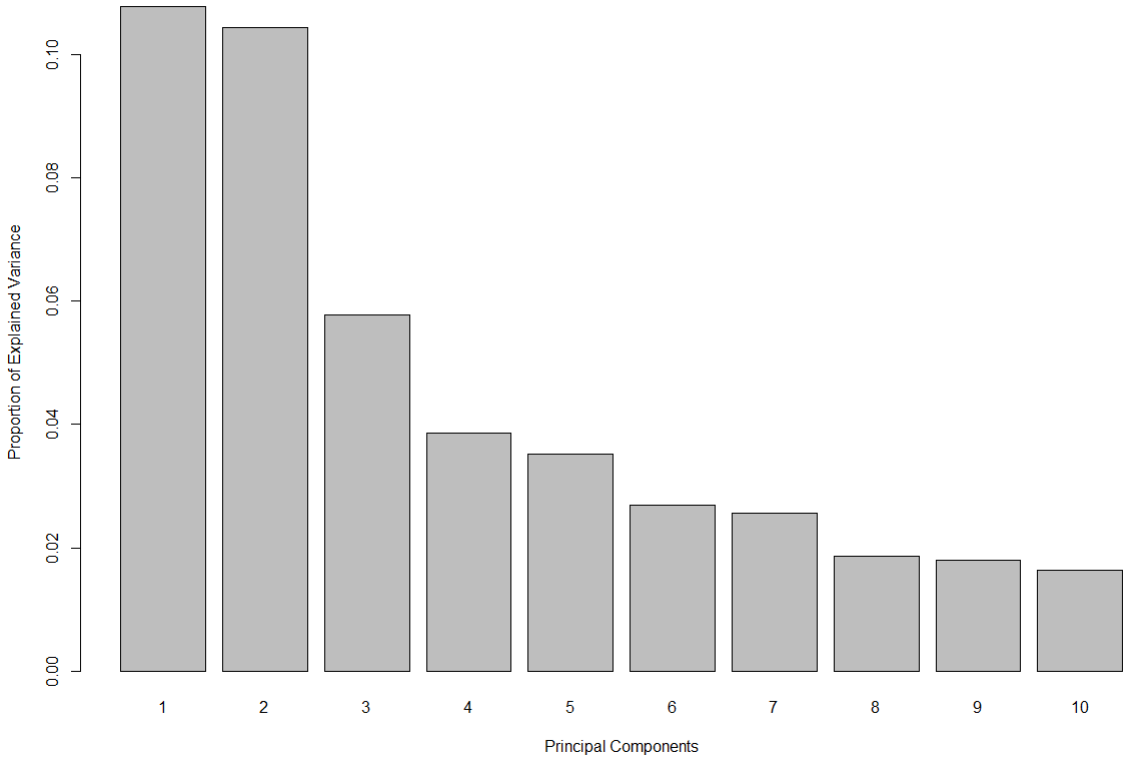


Figure 5.3.1 Scree plot of the first ten principal components of the samples from RSV infected patients at the time of acute disease after normalization and batch correction. Components account for more than 5% of the variation in the data set were used for feature selection

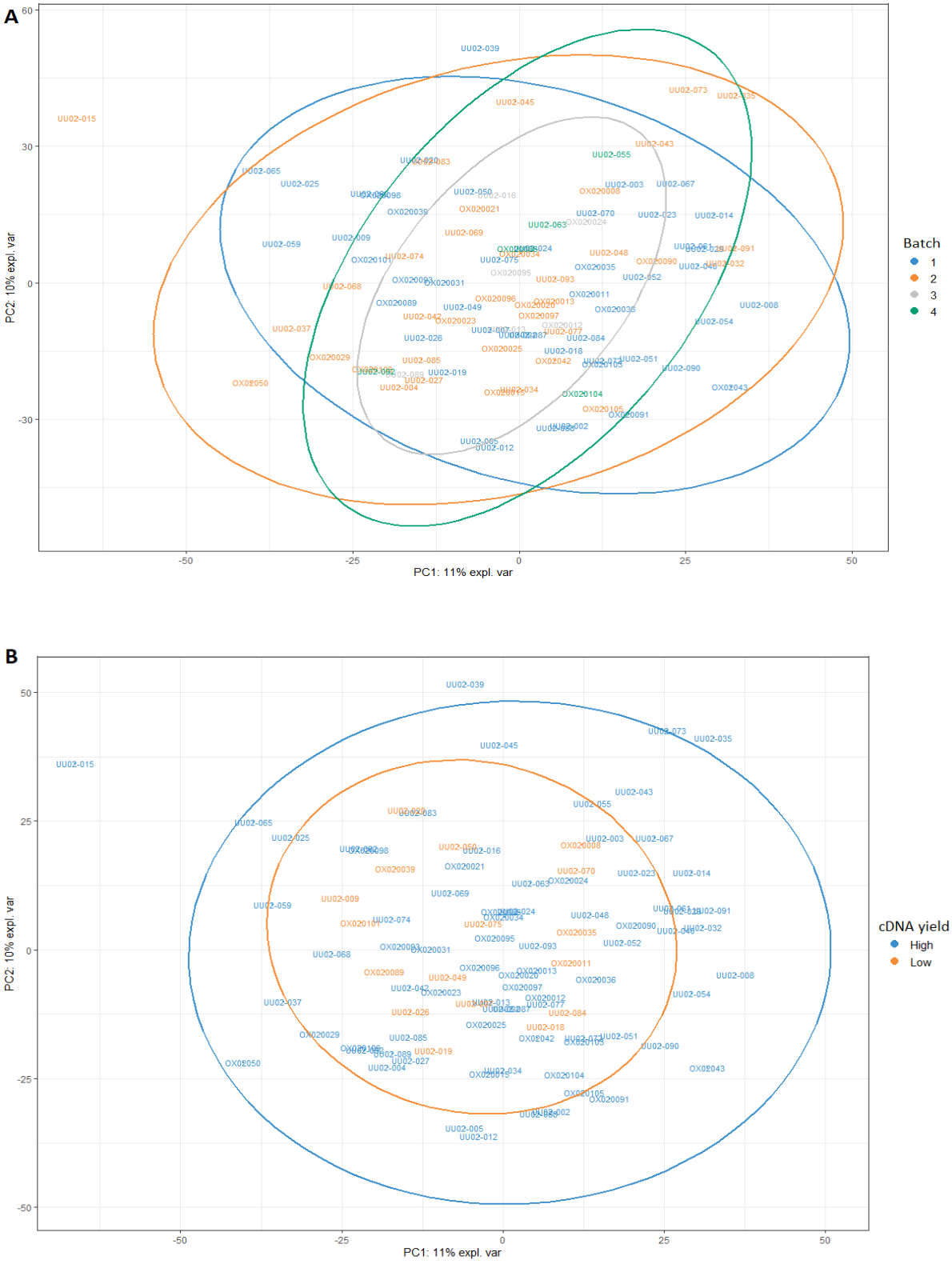


Figure 5.3.2 Mappings of samples at the time of acute disease on the first and second principal components, coloured for A: batch effects according to microarray batch and B: cDNA yield, coloured according to high and low yield, where low yield samples had less than 5ng/µl

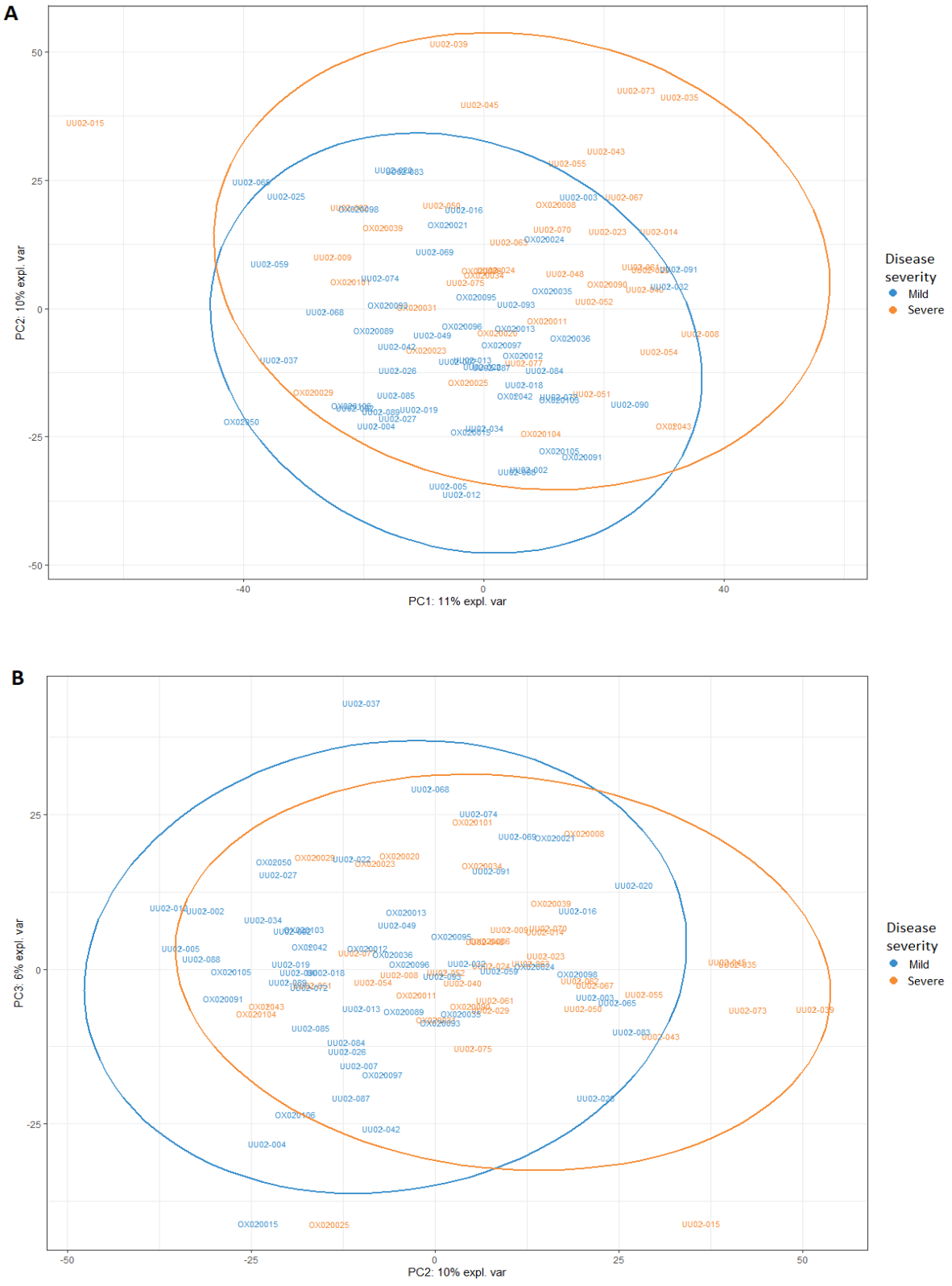


Figure 5.3.3 Mappings of samples at the time of acute disease on A the first and second and B the second and third principal components, coloured for disease severity as defined by requirement for invasive respiratory support. Separation of samples according to disease severity in the first and second principal components is visible

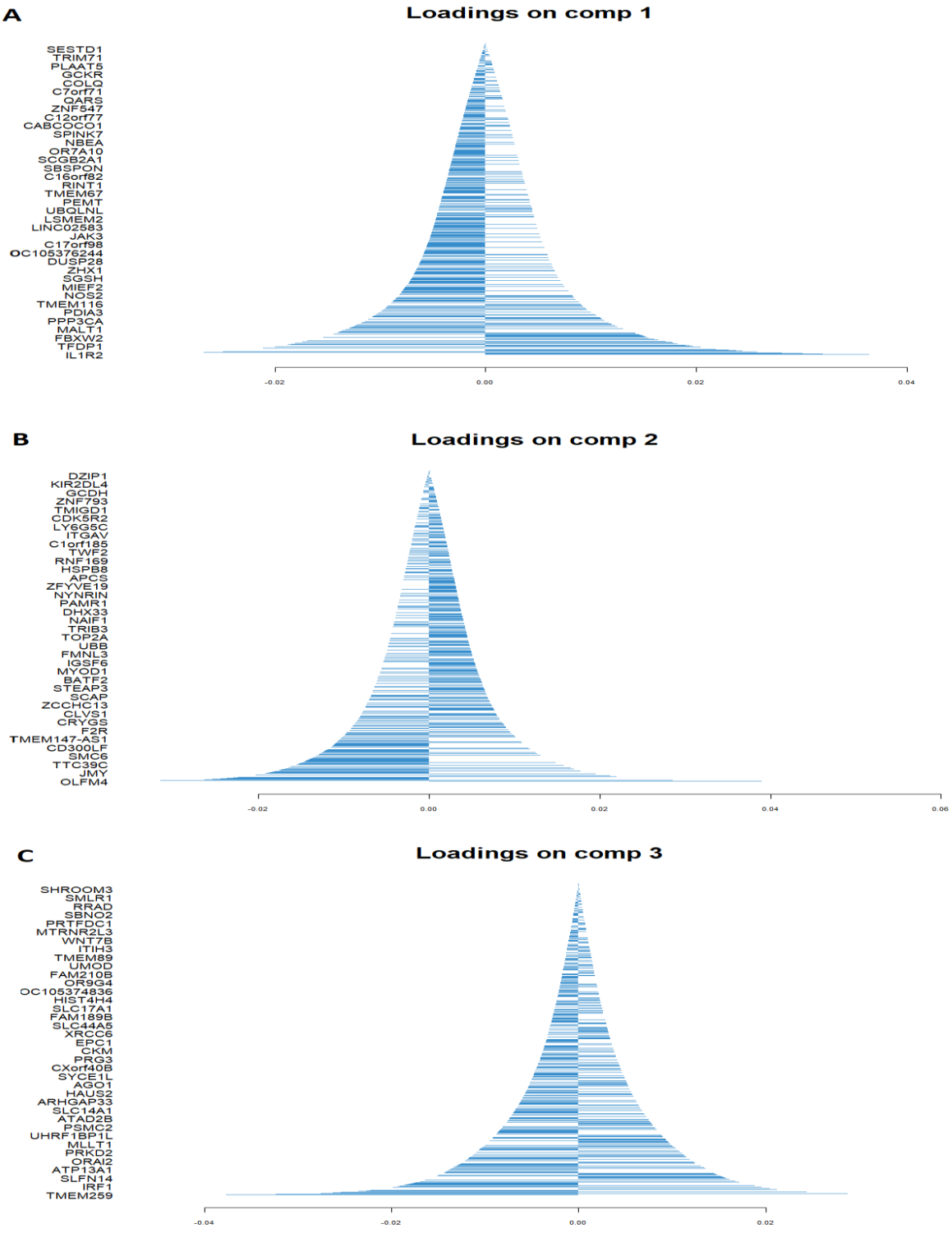


Figure 5.3.4 Loadings of genes on A: the first principal component, B: the second principal component and C: the third principal component are shown, ranked according to their contribution to the total variation within that component. The top 5% of genes contributing to each component were ranked according to their scaled loadings for feature selection

5.4 Multiple algorithms are capable of predicting RSV disease severity using all selected genes

The expression values of all 2463 genes selected by the PCA feature selection method were provided to the SIMON algorithm (described in 5.2.1) along with the age of each participant in days, and the 30 models (full list in appendix 3.2) were selected. 75% of the samples were used in training the algorithms, while 25% were reserved for testing the results of the algorithm. Models were trained on and cross-validated within the training set to make predictions on the completely reserved test set. Severity was defined as requirement for invasive respiratory support, as described in chapters 2 and 4.

The models that performed best in the test set according to AUROC (described in 5.2.2) (fig 5.4.1) were nearest shrunken centroids (pam) (AUC = .8295), Radial SVM (AUC = .8182), partial least squares (AUC = .8295), Naïve Bayes and Random forest (AUC = .8295). ROC curves for the top five highest performing models are shown.

The ranked list of the importance of genes in the top five most highly performing models is shown in table 5.4.2. The genes *PGLYRP1, ITGA2B, CD96, TPM1, MGAM2* and *TMEM19* were the most important genes in each of the top five models. Due to the large size of the gene set used in model training, these models and the genes that contribute most highly to their efficacy are suitable for exploratory analyses of the relationship between expression of individual genes and disease severity, however further feature selection was required for biomarker discovery. The use of feature selection was to produce a minimal set of genes for biomarker discovery analysis, as discussed in chapters 5.5 and 5.6.

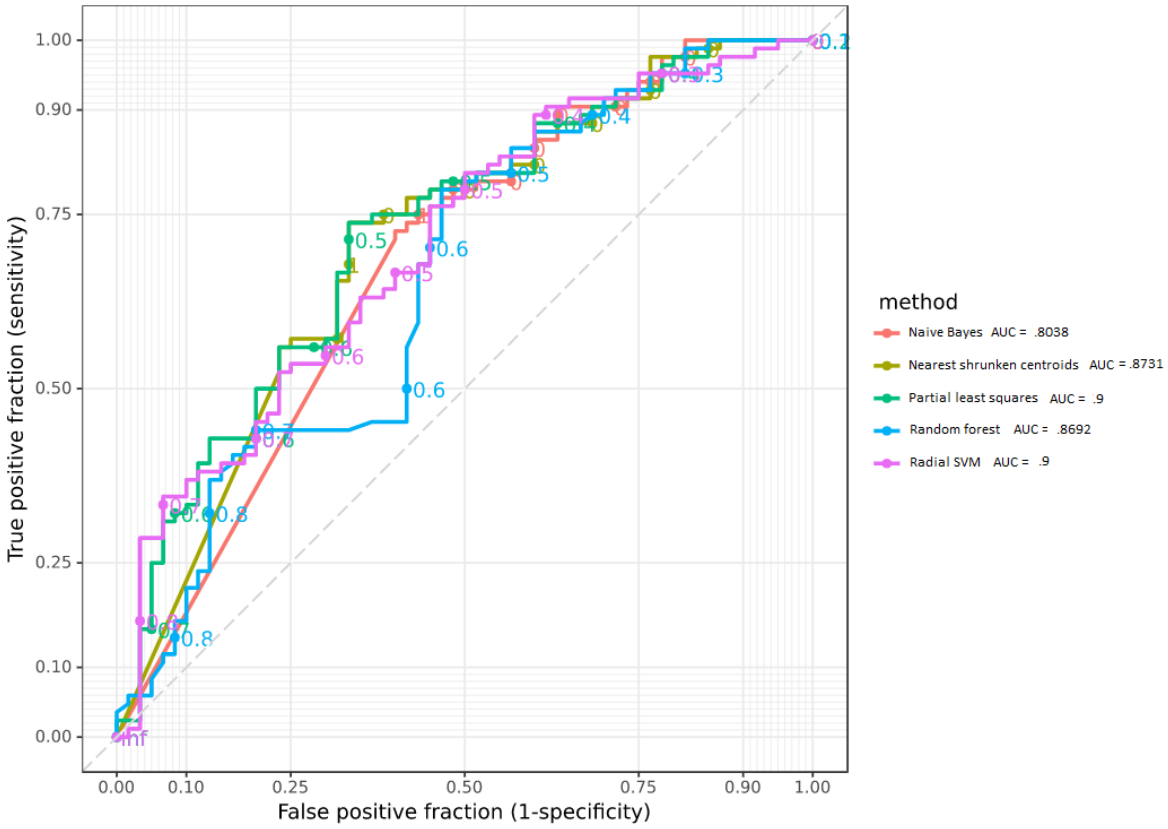


Figure 5.4.1 AUROC statistics for predictions of disease severity in the reserved test fraction of the population for highest performing machine learning models where models were trained on top 2500 selected genes with age used as a cofactor. Numbers presented on each curve represents the probability threshold at which the given true positive and false positive fractions are recorded.

Model Name	Feature	Weight in model
Radial SVM	*PGLYRP1*	100
Random Forest	*ITGA2B*	100
Partial least squares	*PGLYRP1*	100
Nearest shrunken centroids	*PGLYRP1*	100
Naive Bayes	*PGLYRP1*	100
Radial SVM	*CD96*	98
Partial least squares	*TPM1*	98
Nearest shrunken centroids	*TPM1*	98
Naive Bayes	*CD96*	98
Radial SVM	*ITGA2B*	95
Naive Bayes	*ITGA2B*	95
Radial SVM	*TMEM19*	94
Radial SVM	*MGAM2*	94
Partial least squares	*ITGA2B*	94
Naive Bayes	*TMEM19*	94
Naive Bayes	*MGAM2*	94
Radial SVM	*NSUN7*	93
Partial least squares	*SLC2A14*	93
Naive Bayes	*NSUN7*	93
Radial SVM	*CEACAM4*	92
Radial SVM	*CEACAM1*	92
Partial least squares	*CEACAM1*	92
Nearest shrunken centroids	*ITGA2B*	92
Naive Bayes	*CEACAM4*	92
Naive Bayes	*CEACAM1*	92
Radial SVM	*TPM1*	91
Radial SVM	*MTERF2*	91
Radial SVM	*SLC2A14*	91
Partial least squares	*ITGB3*	91
Naive Bayes	*TPM1*	91

Table 5.4.2 Genes ranked by model contribution rank for the top 5 best performing models in the test set where models were trained on top 2463 selected genes

5.5 100 gene signature is predictive of more severe RSV disease

The expression values of the top 100 genes selected by the PCA feature selection method were provided to the SIMON algorithm along with the age of each participant in days, and the 30 models (full list in appendix 3.2) were selected. 75% of the samples were used in training the algorithms, while 25% were reserved for testing the results of the algorithm. Models were trained on and cross-validated within the training set to make predictions on the completely reserved test set. Severity was defined as requirement for invasive respiratory support, as described in chapters 2 and 4.

The models that performed best in the test set according to AUROC (described in 5.2.2) (fig 5.5.1) were nearest shrunken centroids (AUC = .8846), Radial SVM (AUC = .8615), partial least squares (AUC = ..8846), Naïve Bayes (AUC = .8692) and Random forest (AUC = .8846). The top five models with the highest performance in the test set are shown.

The ranked list of the importance of genes in the top five most highly performing models is shown in table 5.5.2. The features *PGLYRP1,* age*, CEACAM1, CD96, SLC2A14, LCN2* and *PSTPIP2* were the most important features in the top five models. It was decided that a smaller model would be more suitable for use as a biomarker, and that these models and the genes that contribute most highly to their efficacy are more suitable for exploratory analyses of the relationship between expression of individual genes and disease severity. Feature selection was used to produce a minimal set of genes for biomarker discovery analysis by removing genes until model breaking point is discussed in chapter 5.6.

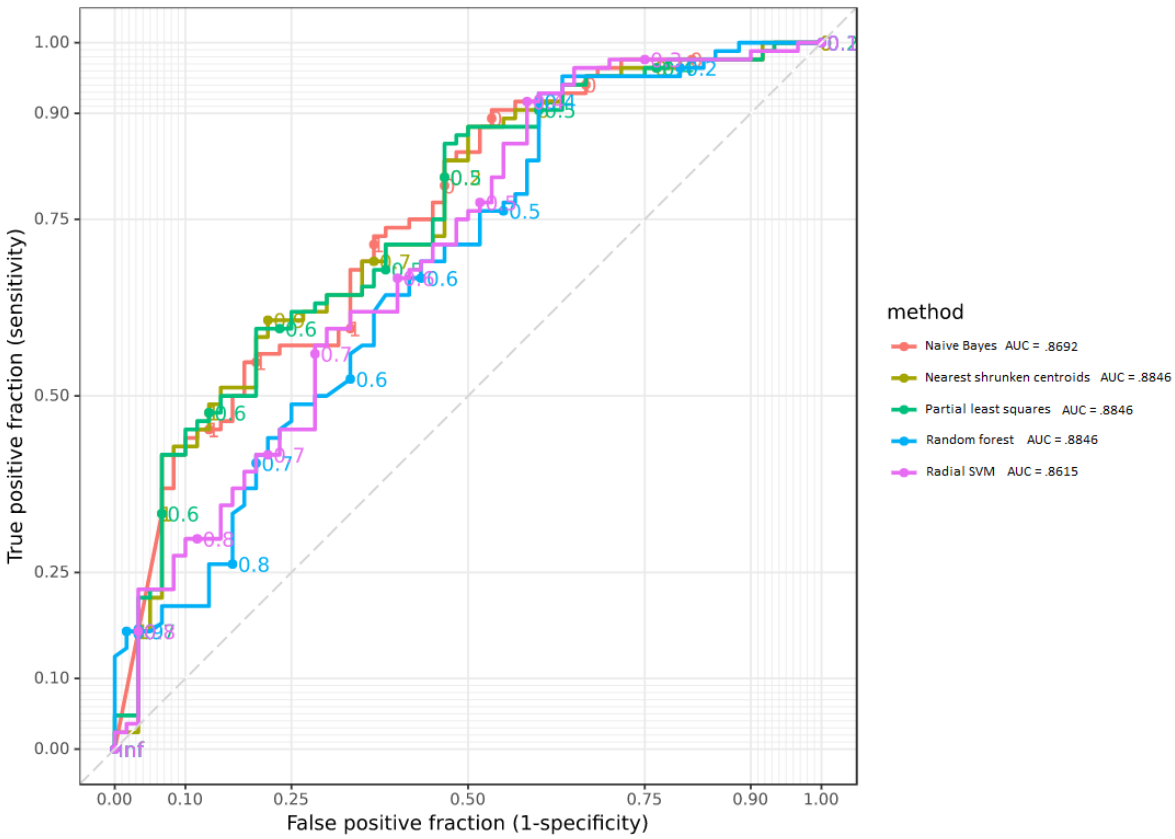


Figure 5.5.1 AUROC statistics for predictions of disease severity in the reserved test fraction of the population for highest performing machine learning models where models were trained on top 100 selected genes with age used as a cofactor. Numbers presented on each curve represents the probability threshold at which the given true positive and false positive fractions are recorded.

Model name	Feature	Weight in model
Radial SVM	*PGLYRP1*	100
Random Forest	age	100
Partial least squares	*PGLYRP1*	100
Nearest shrunken centroids	*PGLYRP1*	100
Naive Bayes	*PGLYRP1*	100
Partial least squares	*SLC2A14*	93
Radial SVM	*CEACAM1*	92
Naive Bayes	*CEACAM1*	92
Naive Bayes	*SLC2A14*	91
Partial least squares	*CEACAM1*	91
Radial SVM	*SLC2A14*	91
Nearest shrunken centroids	*SLC2A14*	90
Partial least squares	*LCN2*	89
Naive Bayes	*PSTPIP2*	88
Radial SVM	*ARG1*	88
Nearest shrunken centroids	*CEACAM1*	88
Radial SVM	*PSTPIP2*	88
Naive Bayes	*ARG1*	88
Naive Bayes	*LIN7A*	86
Nearest shrunken centroids	*LCN2*	86
Radial SVM	*LIN7A*	86
Partial least squares	*IL1R2*	85
Naive Bayes	*IL1R2*	85
Radial SVM	*ORM1*	85
Radial SVM	*IL1R2*	85
Naive Bayes	*ORM1*	85
Nearest shrunken centroids	*IL1R2*	84
Partial least squares	*FCAR*	84
Partial least squares	*LIN7A*	84
Partial least squares	*NQO2*	83

Table 5.5.2 Genes ranked by model contribution rank for the top 5 best performing models in the test set where models were trained on top 100 selected genes

5.6 A five gene signature is predictive of requirement for invasive ventilation

As a gene signature made up of a hundred genes may have limited usefulness as a biomarker due to the difficulty of simultaneously assessing expression of large amounts of genes, it was necessary to remove genes from the analysis in a robust manner to isolate the smallest signature that power to predict more severe disease.

Genes were iteratively removed from analysis based on their ranking according to PCA feature selection and the algorithms re-run until the smallest number of genes with satisfactory predictive power were found. All parameters of machine learning algorithms remained the same as those described in chapter 5.4. The smallest gene signature with predictive power found consisted of five genes, *IL1R2, OLFM4, MMP8, BCL6,* and *ANXA3*. Age of participant was also included as a covariate in model training.

The models that performed best in the test set according to AUOC values (fig 5.6.1) were random forest (AUC = .8769), Radial SVM (AUC = .8615), nearest shrunken centroids (AUC = .8154), naïve Bayes (AUC = .8231) and partial least squares (AUC = .8154). Performance according to AUROC of the top five highest performing models is shown.

The ranked list of the importance of genes in the top five most highly performing models is shown in table 5.6.2. The genes *IL1R2, OLFM4* and *BCL6* were the most important genes in each of the top five models, with age being the second most important feature in random forest.

The expression of each of these genes in severe and mild groups is shown in figure 5.6.3. All genes except are more highly expressed in severe disease than in mild.

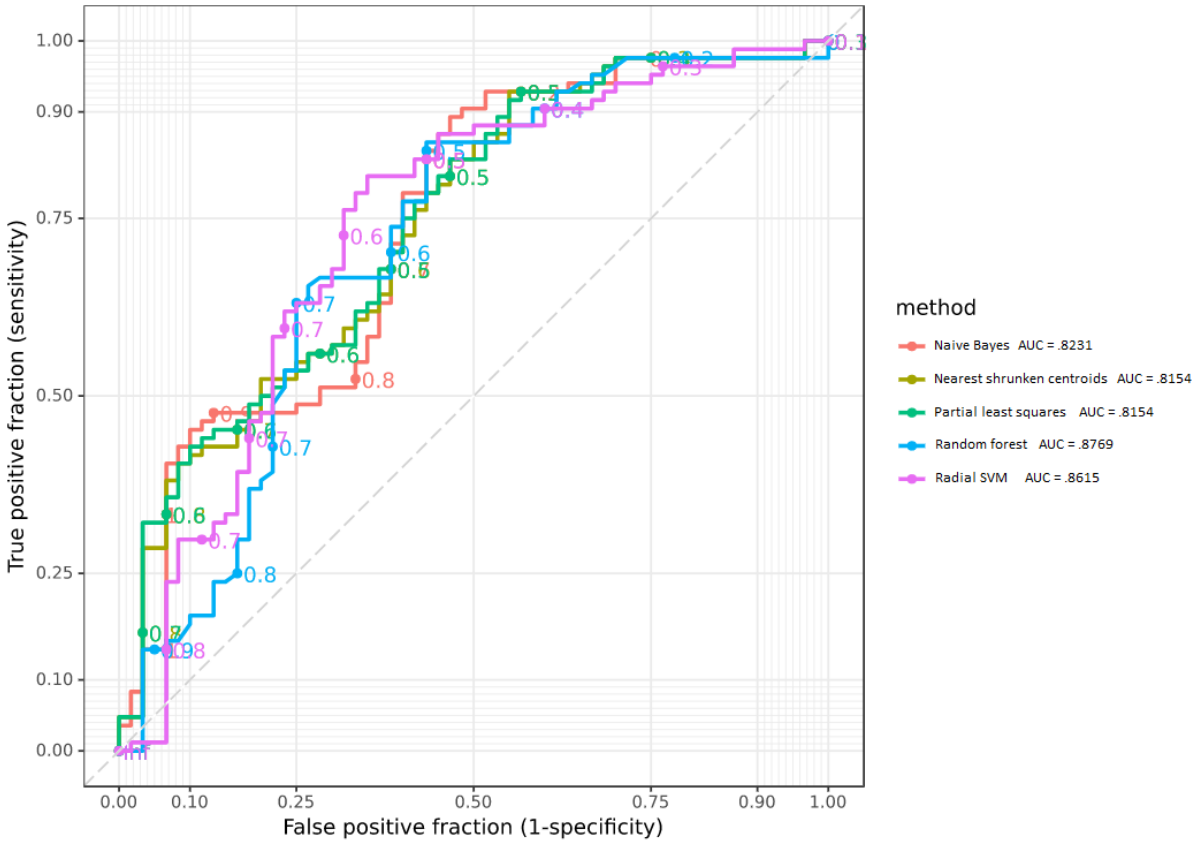


Figure 5.6.1 AUROC statistics for predictions of disease severity in the reserved test fraction of the population for highest performing machine learning models where models were trained on top 5 selected genes with age used as a cofactor. Numbers presented on each curve represents the probability threshold at which the given true positive and false positive fractions are recorded.

Model name	Feature	Weight in model
Partial least squares	*IL1R2*	100
Random Forest	*BCL6*	100
Naive Bayes	*IL1R2*	100
Radial SVM	*IL1R2*	100
Nearest shrunken centroids	*IL1R2*	100
Nearest shrunken centroids	*OLFM4*	97
Partial least squares	*OLFM4*	96
Random Forest	age	94
Naive Bayes	*BCL6*	93
Radial SVM	*BCL6*	93
Naive Bayes	*OLFM4*	90
Radial SVM	*OLFM4*	90
Partial least squares	*MMP8*	84
Nearest shrunken centroids	*MMP8*	82
Partial least squares	*BCL6*	81
Partial least squares	*ANXA3*	79
Radial SVM	*MMP8*	77
Nearest shrunken centroids	*BCL6*	77
Naive Bayes	*MMP8*	77
Nearest shrunken centroids	*ANXA3*	76
Naive Bayes	*ANXA3*	76
Radial SVM	*ANXA3*	76
Random Forest	*OLFM4*	50
Random Forest	*IL1R2*	38
Random Forest	*MMP8*	18
Nearest shrunken centroids	age	4
Radial SVM	age	0
Partial least squares	age	0
Random Forest	*ANXA3*	0
Naive Bayes	age	0

Table 5.6.2 Genes ranked by model contribution rank for the top 5 best performing models in the test set where models were trained on top 5 selected genes with age as a covariate.

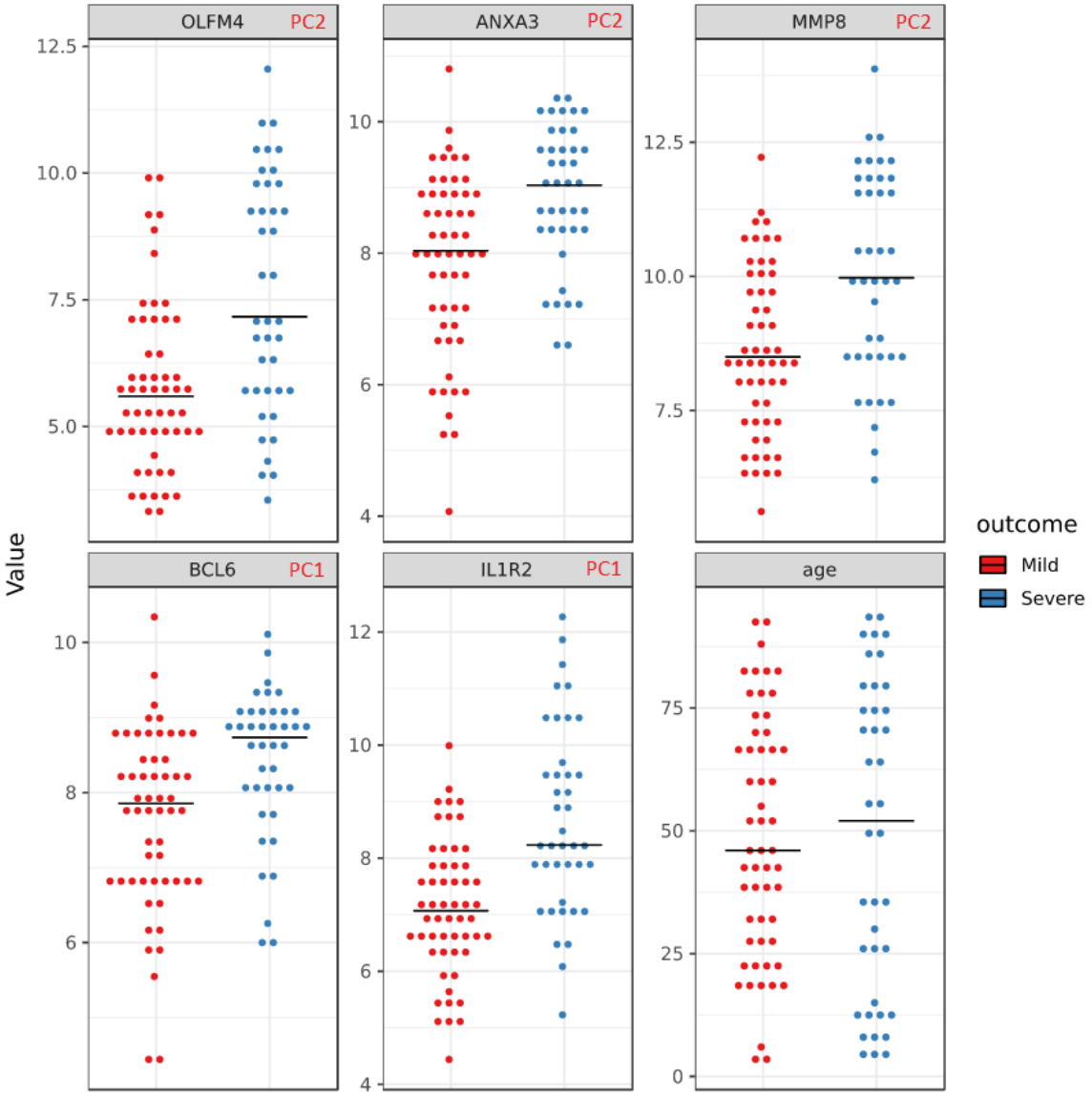


Figure 5.6.3 Expression of the six features in final signature for both mild and severe disease. Black line represents mean of the values of the feature. Principal component of origin of the feature is indicated for each feature except age.

5.7 Validation in published data

Gene signatures predictive of more severe disease were further tested in a publicly available independent validation cohort published by Heinonen et al(285). Series matrix of normalized gene expression values was downloaded from GEO, and clinical metadata (specifically relating to disease severity according to GRSS) were requested from the authors. Data from both studies were scaled to each other and centred around a mean of 0. Batch correction was employed using ComBat (described in 3.2.9) to ensure comparability

between the studies. Participants above the age of 12 months were removed from the study, along with healthy controls so that only participants with acute RSV infection remained. Participants were categorised as having severe or mild disease according to their respiratory intervention score within GRSS, where receipt of CPAP/BiPAP or a more invasive method of respiratory support was given the maximum score of 3. Details of participant selection and statistical transformation of data were described in 5.2.4. Predictive power of the top five models using the small signature outlined above were assessed in these participants, and the outcomes assessed according to AUROC. Distribution of disease severity in the population of participants used in the independent validation cohort is shown in table 5.7.1.

The gene and age signature using the random forest model that showed and was found to predict disease severity in the validation cohort with an AUROC of .6 (fig 5.7.2), indicating some predictive power.

Total participants	Mild Disease (hospitalised, Cannula or no O2)	Severe disease (CPAP/BiPAP or intubation)
53	39	14

Table 5.7.1 Distribution of disease severities in population used for validation study

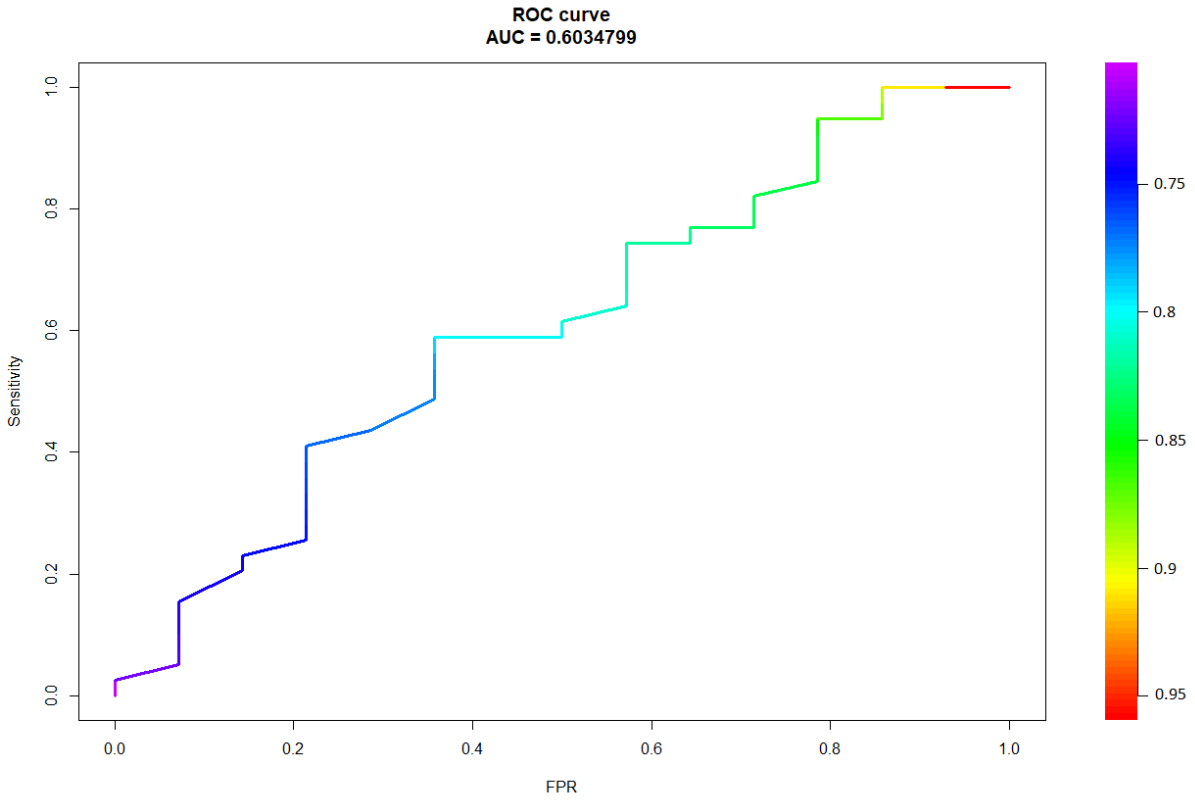


Figure 5.7.2 Receiver operating characteristic curve of predicted disease severity based on random forest model of 5 genes and age in the independent validation cohort. Colour of bar represents probability thresholds at which true positive and false positive rates were calculated

5.7 Discussion

5.7.1 Quality control

Principal component analyses of the data set after removal of batch effects revealed that signals associating with RSV disease severity account for the variation seen in the first two principal components of this data set, suggesting a possible role for the patterns of gene expression associated with these components in RSV severity.

An unsupervised feature selection method using PCA was employed to ensure robustness of machine learning analyses. This was employed rather than a supervised method, where genes were removed from analyses according to their discriminatory capacity within models in order to prevent a “peek forward” effect that can contribute to overfitting, where features of the data set that are identified as being highly associated with one outcome during model training are retained, thus potentially retaining overfitted variables in the model at the expense of potentially more general and informative ones.

5.7.2 Models and validation

The models with the highest predictive power in test sets in this analysis were random forest, naïve Bayes and linear discriminant analyses. While random forest performs well in both the larger training set of 1000 and the smaller one with five genes and participant age, the other models do not tend to perform well in the larger data sets. Bayesian methods in particular due to their assumption of no inter-relatedness of variables can fail to account for sources of structure in larger data sets(390). Linear discriminant analysis attempts to find a linear combination of variables that are different between classes in a similar manner to how principal component analysis finds linear combinations of variables that are similar between classes. Linear discriminant analysis thus tends to overfit due to maximising distance(391).

Random forest analysis performed the best out of all models in the smaller data set. Random forest generates a multitude of decision trees, each of which makes a classification prediction for the outcome variable. Within each decision tree some aspect of the variables is used to separate and assign categories to each sample. The mode of the assigned classes for each sample is chosen as the final output for that sample(392).

5.7.3 Genes

Aside from the capacity to predict disease severity, the usage of these machine learning methods allows for the distillation of huge amounts of information down to a handful of the most informative genes that can act as stand ins for biological processes that influence disease outcomes while excluding genes that do not have a strong relationship with severity.

The gene with the highest predictive power in the minimal signature shown here was *IL1R2* (interleukin 1 receptor type 2). *IL1R2* is a decoy receptor for the IL-1 family of cytokines(393), and can be considered a negative regulator of IL-1 activity as higher *IL1R2* expression results in decreased IL-1 family signalling. Expression of *IL1R2* occurs rapidly during an infection, and is induced by signals from common inflammatory molecules such as NF-kB and IL-1β that are commonly expressed in most infections, including RSV infection(394)(395). *IL1R2* has been found to be differentially expressed in acute RSV infection in this study, and is also more highly expressed in more severe RSV disease(275).

IL1R2 is also more highly expressed in acute influenza infection(238) and is upregulated during severe influenza, but not mild influenza(396). It is possible that either robust IL-1 production in response to RSV infection results in robust clearance and decreased symptoms, or that higher *IL1R2* expression in more severe disease is as a response to excessive inflammation occurring through other pathways such as NF-kB, and that the robust *IL1R2* expression seen in more severe disease is indicative of a feedback loop aiming to control harmful inflammation in these participants. In the mouse model lower IL-1 signalling results in less lung immunopathology due to decreased recruitment of neutrophils alongside decreased survival rates, suggesting both a role in disease severity through facilitating tissue damage and a requirement to resolve disease(397). *IL1R2* is commonly expressed in neutrophils(398), and overexpression has been shown to have anti-inflammatory properties in mice and in cell lines(399), supporting the hypothesis that the *IL1R2* expression in more severe disease in this study originates in an attempt to control excessive neutrophil activity.

The gene with the second highest predictive power in the predictive signature using five genes and participant age shown here was *OLFM4*, which has previously been published as a potential biomarker for more severe RSV disease(275). *OLFM4* has not been described in the context of a respiratory viral infection in any other studies. *OLFM4*, and is associated with the specific granules of neutrophils(357), further implicating neutrophils in severe RSV disease. However, much like as had been observed in *IL1R2,* it is possible that increased expression of *OLFM4* in more severe disease may be part of a signalling feedback loop induced by excessive inflammation via NF- κB signalling, as *OLFM4* has previously been observed to negatively regulate this pathway(292). *OLFM4* knockout mice have a decreased capacity to develop sepsis(400). *OLFM4* expression is also associated with a subfamily of neutrophils, however the role of this subpopulation and the role of *OLFM4* expression in neutrophils remains poorly characterised. It is possible that a pro-inflammatory role exists for *OLFM4* in neutrophils that has not yet been described(357). *OLFM4* is also highly expressed in Th1 cells, and so may be indicative of an active Th1 population during more severe disease(275).

Similarly, *MMP8*, another gene in the six gene signature is highly expressed in neutrophils. *MMP8* has previously been associated with more severe lower respiratory tract

infections(401)(402) and has been observed in other studies to be significantly upregulated in RSV infection(276). MMP8 is a cytokine linked to neutrophil migration(356), and lower expression of *MMP8* results in decreased neutrophil recruitment and activity(403). Expression of *MMP8* in mice has been linked to antiviral signalling via RIG-1-like receptors and through type-I interferons(404). Due to the demonstrated role of *MMP8* in increased inflammatory activity of neutrophils, it may be that *MMP8* plays a role in exacerbating disease severity, rather than being simply expressed in association with it, as we have hypothesised is the case for some other genes in this predictive signature. MMPs are one of the key agents of tissue remodelling, and *MMP8* has previously been shown to digest collagen tissues(405), suggesting a potential role in lung immunopathology.

BCL6 is a gene involved in T-follicular helper (Tfh) differentiation(406), as well as acting as a checkpoint inhibitor for type-I interferon signalling in mice(407). Both of these functions have potential roles in severe RSV disease. Tfh cells have a well characterised role in aiding B cell production of antibodies(408), and are downregulated in RSV infection, similarly to other lymphocytes(409)(410). Increased expression of BCL6 has previously been associated with both increased RSV disease severity and viral load(284). A similar trend was observed in murine influenza(411). The relationship between Tfh cells and RSV disease severity remains poorly characterised. As a type-I interferon inhibitor, the expression of *BCL6* is controlled through IRF7, and thus occurs in response to increased signalling through type-I interferons(412). Increased *BCL6* signalling correlated with increased inflammation and a greater capacity to clear viral infections, while decreased signalling increased the susceptibility of mice to viral pneumonia. While this relationship has not yet been characterised in humans, it is possible that similarly to OLFM4, increased expression of BCL6 in more severe disease is a response to rapid production of early inflammatory mediators such as type-I interferons, and occurs to help control harmful inflammation during infection. It is also possible that decreased interferon signalling due to *BCL6* results in increased disease severity due to a decreased capacity to adequately respond to and control viral infections due to inhibited interferon signalling. Reduced capacity to induce interferon signalling due to *BCL6* has been shown to increase susceptibility of Tfh cells to viral infection in vitro(413).*BCL6* has also been implicated in neutrophil inflammation in the lungs in Influenza infections(414).

ANXA3 has been implicated in RSV disease severity in previous transcriptomics studies(275)(278), and has been implicated in the immune responses associated with worsened outcomes in sepsis(415). *ANXA3* expression has been found to be highly expressed in neutrophils(415), and is associated with neutrophil and monocyte granules(416). It is possible that *ANXA3* associates with severe RSV due to the increased numbers of neutrophils and neutrophil activity and is expressed more highly in severe disease due to increased activity of neutrophils.

Other genes were implicated in more severe disease from the larger gene signatures of 2463 genes and 100 genes. The gene *PGLYRP1* is the gene with the strongest relationship with disease severity in both gene sets, and is a peptidoglycan receptor related to neutrophil activity in response to detection of gram positive bacteria, the cell walls of which contain peptidoglycan(417), while also contributing towards increased Th2 and Th17 bias alongside decreased regulatory T cell and plasma cell differentiation bias(418). *PGLYRP1* been proposed as part of biomarker panel for influenza infection due to its induction by influenza infection and high correlation with neutrophil responses, suggesting some general function in viral infections(419). Altogether this makes *PGLYRP1* a large potential contributor to severe RSV disease, although the means by which it is signalled to in viral infection remain unclear. It is possible that *Streptococcus pneumoniae* proliferation in the respiratory tract during RSV infection is inductive of *PGLYRP1* signalling, although it has also been shown that *PGLYRP1* can occur in the absence of a bacterial infection by dimerization with TREM-1(417). *ITGA2B* is an integrin shown to be involved in NF-kB mediated activation of neutrophils(420). *CD96* is a checkpoint inhibitor involved in the limitation of $CD8^+$ T cell responses(421). TPM1 controls the responsiveness to IL-12 signalling, and therefore regulates Th1 activity(422). *CAECAM1* is also a regulator of $CD8^+$ T cell activity(423). Involvement of the genes with the strongest relationships with disease severity in neutrophil responses, CD8 responses and Th1 type responses is telling of the strong relationship between these processes and RSV disease.

Overall, the genes that make up the biomarker identified here broadly relate to innate signalling induced on infection. The association between genes that are induced by and regulate responses to early immune response signallers such as IL-1 and NF-kB suggests that severe RSV disease may be associated with strong early innate immune signalling via IL-1

and NF-kB, and that this induces a strong feedback response in more severe disease to control pathological inflammation. This early inflammation likely occurs due to increased activity of neutrophils, causing local tissue remodelling and inflammation. It is also possible that the immunoregulatory roles of these proteins are themselves contributing towards increased disease severity, however immunoregulatory responses have traditionally been associated with decreased disease severity in RSV, as discussed in chapter 1. In parallel to this, antiviral signalling via type-I interferons may induce *MMP8* and *BCL6*. Increased expression of *BCL6* causes a downregulation of interferon related signalling, preventing the development of an appropriate immune response capable of robustly clearing infection. It is difficult to say if these responses merely coincide with more severe disease, or actively contribute towards increased disease severity. Functional analyses will be required to further develop these causative relationships.

5.7.4 Model performance

The models shown here can predict disease severity status of an individual based on whole blood gene expression during an active infection with an AUROC of .87 in the test set and .6 in the independent validation set. This indicates that the model has very strong predictive power in the participants from the current study and can categorize participants much more effectively than random. However, the model can only successfully categorise participants as having severe disease in the independent validation set slightly better than random. There are numerous potential explanations for this. It is possible that the criteria for the decision to treat a participant with more invasive ventilation is different between populations, as different clinicians may have different standards for the application of more invasive treatment. It is also possible that a different method of selecting features may show better results in model training as genes may be selected based on their predictive capacity. While selecting genes by their contribution to the overall variance in the data set reduces chances of overfitting, it is not necessarily ideal for the selection of genes with the highest influence on severity. A hybrid approach may show better performance, where PCA based feature selection would be used to select for a set of genes that account for a large proportion of the variance in the data set, and more traditional forward and backward selection methods are employed to reduce that gene set to a minimal, predictive set. The integration of age into the signature provides a key advantage in that it allows for the

determination of predictive power of genes beyond what is accounted for by age related effects.

Validation of a gene signature in the independent data set was limited by differing microarray technologies. The present study used an Affymetrix Clariom S human HT array, while the validation study used two different Illimuna arrays, the Illumina Human WG-6 v3 and Human HT-12 v4(276). Comparability of microarray results across different platforms has been controversial, with many studies showing conflicting results indicating either good comparability(424)(425)(426)(427) or a lack of comparability(428)(429)(430). Affymetrix arrays and Illumina arrays use different probe lengths in arrays, meaning that they are not measuring exactly the same thing(424). Different array types may also use different isoforms of genes in probes, further limiting potential comparisons between arrays. It has been demonstrated that usage of probes for a single gene that do not map to the exact same sections of the gene results in very low comparability(424). Agreement between platforms is also highly dependent on expression levels of probes, with more highly expressed probes showing higher levels of agreement than less highly expressed probes(424).

Clinical implementation of a biomarker to improve patient stratification would require evidence of supplementary value of the signature to the decisions of a clinician. An idea of the efficacy of a transcriptomic biomarker as a supplement to clinician decisions may be gained by a study in which a biomarker is used to predict requirement of treatment in parallel to a clinician's initial decisions on presentation. A prospective head to head comparison between the efficacy of the biomarker in predicting progression to more severe disease versus the physician's initial decisions would show the potential value of the biomarker as supplemental to clinician decisions. Additionally, the inclusion of the biomarker alongside the clinician's assessment versus the clinician's assessment alone would also provide evidence of whether the biomarker brings added clinical value. If the value of the biomarker was shown in this way it could become a useful routine tool to assist clinical decision making in ambiguous cases, such as in cases where decisions must be made to admit, refer to PICU or intubate.

Due to the inherent constraints for biomarker discovery of sampling and disease severity measurement occurring at the same time point, a potential direction for future work could

be in the validation of a biomarker using samples at the time of presentation to predict the peak of disease severity, to determine if this signature can be used to predict the severity of the peak.

5.7.5 Limitations

One of the largest limitations of the current work is the lack of a full body of meta data representing longer term respiratory sequelae of RSV infection. One unmet clinical need within RSV research is the difficulty in predicting which patients will develop long term respiratory sequelae after infection with RSV. Data are being gathered within the RESCEU trial that aim to address this issue through clinical biomarker research, however due to the nature of long term sequelae, this information is only available months or years after all samples are gathered and all participants recruited to the study, and so the development of biomarkers of respiratory sequelae is beyond the scope of this thesis.

Transcriptomics samples from participants recruited in the final year of the trial were not available at the time of writing this thesis, and so analyses using these samples was not possible. A further analysis in the future using the full complement of samples gathered (and thus additional power) within the RESCEU study will be possible in the future.

5.7.6 Conclusions

A robust and predictive RSV biomarker is shown with limited capacity to predict disease severity of a participant at the time of presenting using only six genes. The genes with the most discriminatory power for severe RSV infection are *IL1R2*, *OLFM4*, *MMP8*, *ANXA1* and *BLC6* which are involved primarily in neutrophil and type-1-interferon responses. Age is an important cofactor of this signature. Many of the genes included in the final severity associated gene signature are both associated with neutrophils and induced by NF-kB expression, suggesting that much of the activity of pro-inflammatory cells that is linked to more severe RSV disease occurs as a product of NF-kB signalling, suggesting that it may be a useful therapeutic target in future drug design.

Chapter 6: Systematic review of the role of Antibodies and T lymphocytes in RSV disease severity in mice and cattle

6.1 Introduction

In chapters 4 and 5 I presented a transcriptomic biomarker to predict development of more severe disease in humans and functional analyses of gene signatures associating with more severe RSV disease. One method of further elucidating the relationship between the immune system and disease severity (briefly discussed in chapter 4) is the use of animal models. Initially immunological data (such as lymphocyte data and antibody data) were going to be integrated with transcriptomics data for the final chapter of this thesis. Due to the ongoing pandemic, these data were not generated in time to be included in this thesis. Systematic review of the animal literature was conducted to better understand the immunological parameters to include in analyses once it becomes possible to generate further data.

RSV disease has been extensively studied in animal models, which recapitulate many (but not all) aspects of human infection. The inflammatory cell infiltration of the lung in mice can resemble that seen in infantile bronchiolitis (62), and the cotton rat is well established as semi-permissive for infection with human RSV strains (431). Calves infected with bovine RSV (bRSV) recapitulate some aspects of human disease but also have distinctive features (such as acute emphysema) but there are logistical limitations of working with large animal species. Other animals that have been used to study RSV include lambs(432)(433) and chimpanzees(434), but non-human primates have major cost and ethical limitations.

In addition to neonatal disease, RSV also causes severe but sometimes insidious disease in older adults (435)(436)(437)(438). Infection of mice enables the effects of age and other factors on the likelihood of developing severe RSV disease to be studied. This model has been used to study the effects of age on the immune response to RSV(439)(440).

Serum antibody (Ab) and T lymphocyte frequency and phenotype are two aspects of the immune response that have a critical role in RSV disease. Human infant studies demonstrate that neutralising Ab titre correlates inversely with susceptibility to reinfection(441) and there is an inverse correlation between maternally-derived Ab titre (MDA) and RSV severity (442)(107)(443)(444)(445)(446). In addition, monoclonal Ab prophylaxis can prevent RSV disease in the lower respiratory tract (447). However, low levels of neutralising Abs correlate with enhanced disease severity after RSV infection(445)(448)(444). Isotypes of serum Ab and both self-derived and maternally-derived serum Ab titres are reviewed separately.

Similarly, T lymphocytes can be both pathogenic and protective in RSV infection (449). Children with a defective T-cell response have higher viral loads, delayed viral clearance and prolonged disease(450)(451). However, T-cell reconstitution in immunodeficient patients can enhance disease(452). However, measures of T lymphocyte frequency or phenotype in the periphery may not predict disease severity because the relevant cells are concentrated in the site of disease.

In this chapter I aim to systematically review the animal model literature on the roles of T lymphocytes in mice and RSV-specific antibody in mice and cattle in protection and pathogenesis of RSV disease and to assess the potential value of these readouts in animal studies as predictive or concurrent markers of RSV disease severity in humans.

6.2 Methods

This review was structured according to PRISMA guidelines(453).

6.2.1 Search strategy

6.2.1.1 Selection criteria

Studies were included that either investigated Ab titres (any isotype) or T lymphocyte activity or numbers as a variable. For Ab titres both mice and cattle were considered, while for T lymphocyte activity only mice were considered. Human, in vitro and epidemiological studies were excluded. Studies must have involved infection with either human or animal RSV. Studies without a quantitative scoring of RSV disease severity were excluded. Studies must have been structured as controlled trials, cohort studies or case-control studies.

6.2.1.2 Search strategy

A search was conducted on 28th May 2018 using Embase, Pubmed and Zoological Record databases from January 1980 to May 2018. The search terms were 'RSV' or 'Respiratory Syncytial Virus' using filters: 'Journals only', 'Animal experiments' and 'English language'. The inclusion criteria were applied via title and abstract screening, followed by full-text screening. Full texts of identified articles were gathered.

6.2.1.3 Variable measures included

T lymphocyte measures: cell number, proportion, activity and molecular markers, were measured from samples of bronchoalveolar lavage fluid (BALF), lung tissue and serum. Data from the following cell types were analysed separately: CD4 (Th1, Th2, Th17), Treg, CD8, T-cells (both CD4 and CD8) and lymphocytes (T and B cells, and natural killer cells). Antibody measure was isotype titre via optical density measurements from serum samples.

6.2.1.4 Severity measures included

Primary measures were weight loss, airway function, lung histopathology and illness severity scores, and were assumed to be proportional to disease severity. Weight loss is a common, reproducible and objective measure of disease severity in animal models(454). Airway function measures reflect the reduced lung function of infants with severe RSV infection. Lung resistance (RL) gives consistent results for airway function in both mild and severe inflammation models(455), and lung dynamic compliance (Cdyn) is a reflection of peripheral lung function(456). Some studies dichotomised their severity measure as high or low severity, using a cut-off of '2 or more symptoms present' for bovine respiratory disease (BRD), which is severe, or 'fewer than 2 symptoms' for 'non-BRD', which is mild or sub-clinical.

6.2.2 Data extraction

Tables were created with summaries of information concerning study design and study results. Ab titres and severity scores were extracted where available. The studies were divided and analysed in subgroups based on animal type, Ab isotype or T-cell measurement. When results were portrayed graphically, data were extracted using GetData Graph Digitiser. Forest plots were created using Cochrane's 'Review Manager'(457). The tabled observations are independent and are self-reported correlations from the authors of the studies.

6.2.3 Standardised mean difference

Where studies reported continuous outcomes (temperature, weight loss or symptom scores), forest plots were drawn using standardised mean difference (SMD) of severity measures to make different outcomes comparable. Standardized mean difference is a method of standardising measurement types that measure the same outcome but with different measurements. It is a measure of the effect sizes of a given intervention in each study relative to the total variability in the study. In this study standardised mean difference is used to compare relationships between antibodies or lymphocytes and disease severity as measured by a range of different methods.

6.2.4 Odds ratio calculation

Where studies reported a dichotomous outcome measure (i.e. high or low disease severity), effect sizes were calculated as odds ratios and these values are shown in a separate forest plot, as the axis uses a different scale to the effect sizes of continuous outcome. An odds ratio calculation measures the strength of an association between two events. In this study this statistic is used to quantify the association between high antibody and severe disease.

6.2.5 Predictive analyses

Due to the commonality of measures of T lymphocyte number and disease severity across some studies, scatter graphs were also plotted for the analysis of T lymphocytes. Spearman's rank correlation coefficient (rs) was calculated to indicate the association between each cell and severity measure. Cell number and activity data were taken at least one day earlier than severity measures to make the plots predictive. One day before severity measurement was considered to be close enough to consider a marker predictive.

6.3 Studies included

The literature search yielded 5023 studies, which were screened and assessed for eligibility by two reviewers. Twenty-six (fig 6.3.1 A) studies met the eligibility criteria regarding Ab titres and 37 (fig 6.3.1 B) for T lymphocyte frequency or activity. Results for both biomarkers are presented separately to independently establish their potential as concurrent or predictive biomarker for RSV severity.

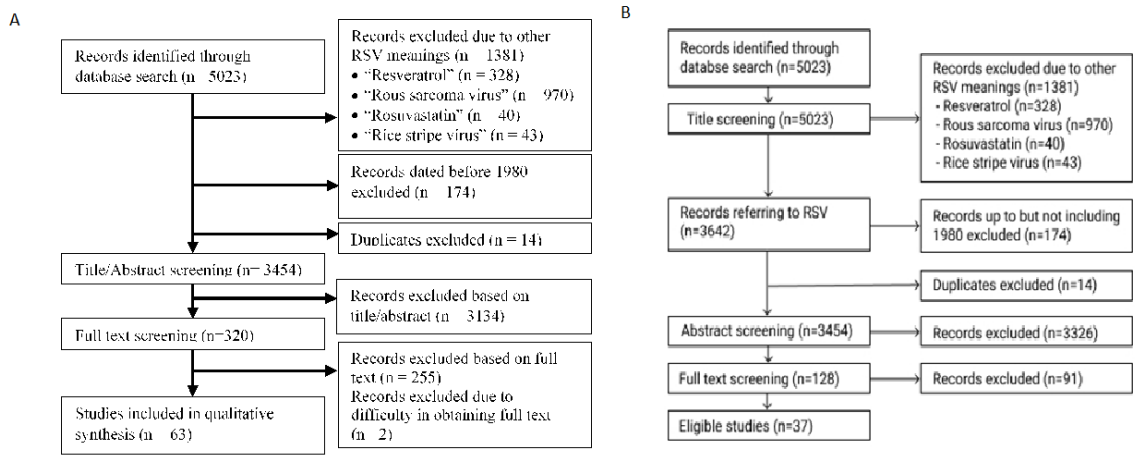


Figure 6.3.1 Literature Search Flowchart, based on PRISMA methods for conducting systematic reviews for A: systematic literature search for contribution of lymphocyte and cellular responses and their relevance to animal disease severity and B: systematic literature search for contribution of immunoglobulin responses and their influence on animal disease severity

6.4 Antibody titres are not indicative of protection in cattle

Most preclinical studies on RSV Ab titres (17 of 26, 65%) (table 6.4.1) were conducted in bovine RSV (bRSV) infected calves, while 9 (35%) (table 6.5.1) were conducted in mice infected with hRSV. Plots of mean difference were generated to demonstrate the effects of antibody subtypes on disease severity. Meta-analyses of calf studies (fig 6.4.2) did not demonstrate any significant association between Ab titre measured at the time of severity measurements and disease severity over the course of disease, with the exception of maternally derived antibodies, which was associated with decreased disease severity.

Fourteen of the 17 calf studies measured infant Ab and the remaining 3 examined the effect of maternal Ab (MDA) on disease severity, using colostrum feeding or deprivation. Seven studies observed a significant correlation between a given Ab titre and a severity measure. Each isotype was examined independently. While 'bRSV-specific antibody titre' (all isotypes) was infrequently associated with severity measures (2 of 7, 29%), 'bRSV-specific IgE titre' was associated with disease severity in three out of four studies. MDA titre was associated with severity in only one of three studies, but all studies comprised only small numbers of animals. There were few studies that analysed RSV specific antibody isotypes (with the exception if IgE) in isolation (Table 6.4.1), and these were not found to be associated with severity by any studies (Figure 6.4.2)

In the studies with a continuous severity outcome, no overall association can be drawn with Ab titre, nor did any isotype variable indicate a substantial association with severity. Combining data from studies that shared outcome measures (e.g. weight loss) was not possible due to incomparable measures of Ab titre. Although most studies included a small number of animals (averaging five animals per group), the three experiments with dichotomous outcome measures (i.e. high or low severity) had higher sample size (>100). Those outcomes (Figure 1) showed no overall significant effect of a different Ab titre on the incidence of BRD (a severity score of 2+).

Three articles reported titres of maternally-derived Ab (MDA) in young calves in relation to RSV infection severity. These studies were conducted in colostrum-fed and colostrum-deprived calves to simulate high and low MDA titres in calves. Overall MDA did not significantly affect severity of disease in these studies (fig 6.4.2). However, 2 out of 3 studies analysing MDA reported an association between more severe disease and lower antibody titres (table 6.4.1) with the third study showing a high variance in measured disease severities, and no statistical significance.

Study	Variable	Primary outcome (continuous / dichotomic data)	Association (P<0.05)	Study limitations	Number of animals (cattle)	Day of Antibody measurement	Age of cattle
Gaeta 2018(458)	Specific Ig Titre	BRD	**No**	Dichotomous outcome	115	Time of recruitment	1-12 months
Antonis 2010(459)	Specific Ig Titre	Symptom score	**Yes**	Confounding (high/low Ab groups were different ages)	10	Days 0, 3,6,9, 13	0-6wks
Kalina 2006(460)	Specific Ig Titre	Symptom score	**Yes**	Confounding (coinfection with Alternaria)	14	Days 0 - 120	8-10wks
Bingham 2000(461)	Specific Ig Titre Specific Ig Titre	BRD Weight gain	**Yes** **No**	Dichotomous outcome	254	Days 0, 28 post infection	N/A
Elvander 1998(462)	Specific Ig Titre	Rectal temp	**No**	Confounding (coinfection with Bovine viral diarrhoea virus (BVDV))	12	Days 0 – 11 post infection	14-17wks
Gershwin 1994(463)	Specific Ig Titre	Symptom score	**No**	Confounding (coinfection with Micropolyspora faeni)	NA	NA	NA
Allen 1992(464)	Specific Ig Titre	BRD	**No**	Dichotomous outcome	136	Day 28 post infection	6-8wks
Blodorn 2015(465)	MDA titre	Symptom scores	**No**		6	Day 7 post infection	< 1 month

Uttenthal 2000(466)	MDA titre	Rectal temp	**No**		5	Days 0 – 30 post infection	3-12wks
Belknap 1991(467)	MDA titre	Respiratory Rate Rectal Temp	**Yes**	Temperature data were not shown	17	Days 0 – 12 post infection	0-2wks
Orro 2011(468)	Specific IgG_1 Titre	Symptom score	**No**		10	Days 0 – 42	9-32 wks
Stewart 1990(469)	Specific IgM Titre	Symptom score	**No**		15	Day 14 post infection	7-8wk
Kalina 2004(470)	Specific IgE Titre	Symptom score	**Yes**	Correlation testing used ranked data	35	Days 7 - 52	6-8wks
Gershwin 1990(471)	Specific IgE Titre	Symptom score	**Yes**		NA	NA	NA
Stewart 1989a(472)	Specific IgE Titre	Symptom score	**No**	Poor reporting of data	15	Day 14	N/A
Stewart 1989(473)	Specific IgE Titre	Symptom score	**Yes**		6	Days 0 -14	NA

Table 6.4.1 A summary of calf studies, with details of observed correlations and study limitations. BRD denotes 'bovine respiratory disease', defined as a symptom score of 2 or more, and is the more severe outcome. MDA denotes maternally derived anitbodies, Days indicates number of days post (or prior) to infection that outcome variable measurements were taken. Outcome measurements include both measurements of antibody and cell abundances as well as measurements of disease severity.

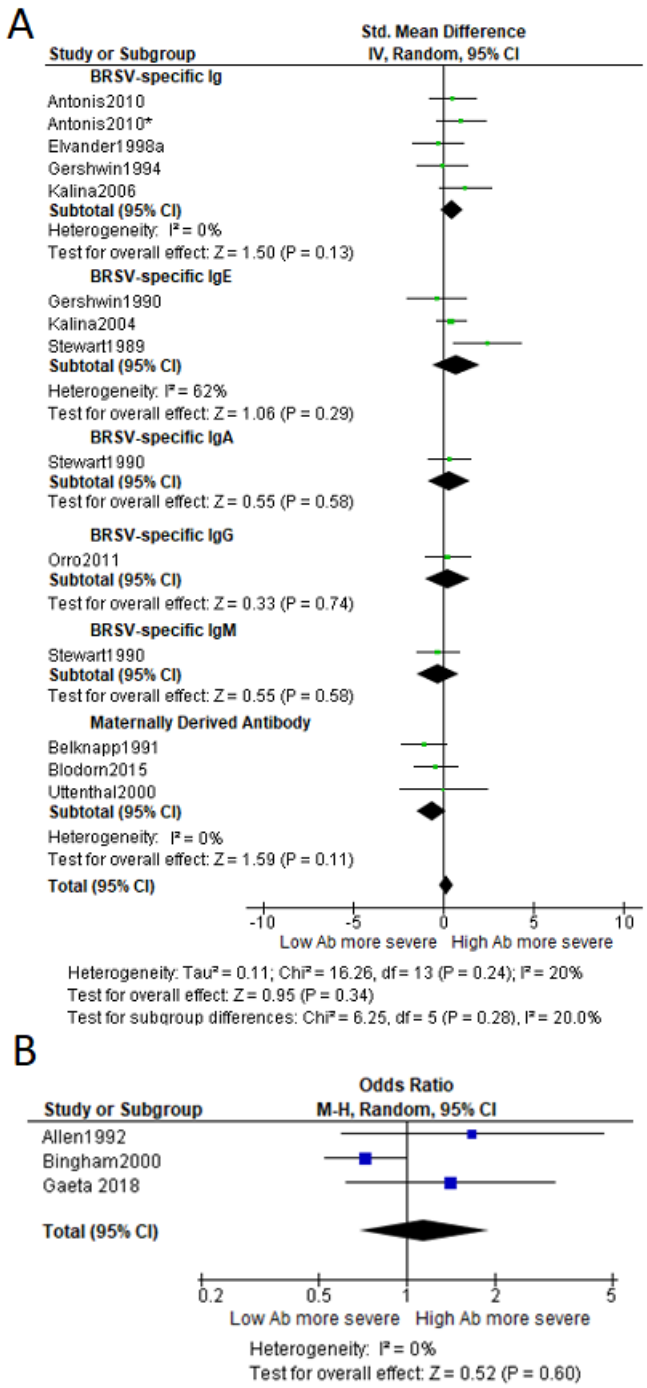


Figure 6.4.2 Forest plots of disease severity outcome data from calf studies that used; A: a continuous outcome measure of disease severity, such as a multifactorial scoring system, or a measure with some other non-binary outcome such as temperature; B: a dichotomous outcome measure of disease severity, where cattle were defined as severe or mild based on the development of any observable form of respiratory disease

6.5 Mouse antibody titres associate with less severe disease

Eligible studies and designs are shown in table 6.5.1. Plots of standard mean difference were plotted for each eligible study that showed a relationship between RSV specific IgG titre, RSV specific IgE titre and total RSV specific Ig.

A significant association with severity was observed for 'total RSV-specific antibody' (all serum isotypes) demonstrating high Ab titres protect from severe symptoms (fig 6.5.2). No significant association with severity was observed for 'RSV-specific IgG titre' or 'IgE titre' (total and RSV-specific IgE). For the purpose of the Ab titre analyses no distinction was made between "predictive" and "correlative" as relatively few studies with Ab as an outcome measurement had depletion of cells or measurements taken prior to infection, so all studies were considered together for a correlative analysis.

Study	Variable	Outcome	Association? (P<0.05)	Limitation	Number of mice	Days of outcome measurement	Age of mice
Chang 2004(474)	Specific Ig Titre	Weight loss Symptom scores	**Yes**	Confounding (IL-2 overexpression in one group)	32	Days 0, 2, 8, 45	6wks
Graham 1991(475)	Specific Ig titre	Weight loss	**Yes**	Confounding (B cells and therefore all antibody depleted)	44	Days 0, 4,5,7,8	4-8wks
Graham 1991a(476)	Specific Ig titre	Weight loss	**No**	Poor data reporting	24	Days 7, 10, 28, 42	6-8wks
Shao 2015(477)	Specific IgG titre	Lung pathology	**Yes**	Confounding (Anti-CD25 antibody given to one group)	10	Day 0, 4, 8, 28	6-8wks
Harker 2010(478)	Specific IgG titre	Weight loss	**Yes**	Confounding (IL-18 overexpression in one group)	15	0, 2,4,6,8,12	7-8wks
Dakhama 2009(479)	Specific IgE titre	Airway resistance	**Yes**	Confounding (IL-4/IL-13 knockout)	62	Day -1, 0, 2,3,6	8-10wks
Dakhama 2004(480)	Specific IgE titre	Airway resistance	**Yes**	Confounding (anti-IgE antibody given to one group)	114	Days 7, 14, 21, 28	8-10wks
Barends 2002(481)	Total IgE titre	Weight loss	**No**	Confounding (ovalbumin allergy)	24	Day 0, 4, 6	7-9wks
Matsuse 2000(482)	Total IgE titre	Lung pathology Airway resistance	**Yes** **Yes**	Confounding (co-infection with Dermatophagoides farinae (Df))	24	Day 4,7,8,10,32,91	6-8wks

Table 6.5.1 A summary of mouse studies, with details of observed correlations and study limitations. Days indicates number of days post (or prior) to infection that outcome variable measurements were taken. Outcome measurements include both measurements of antibody and cell abundances as well as measurements of disease severity.

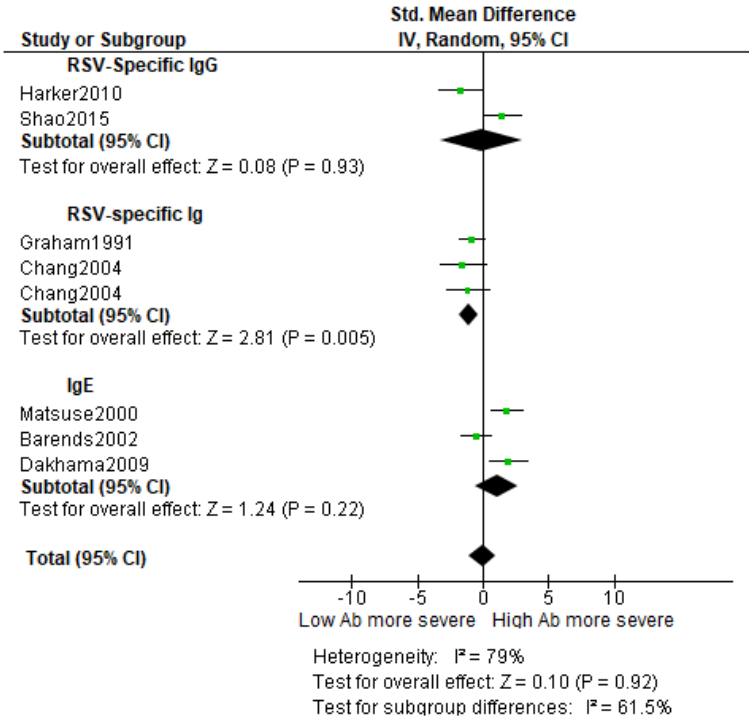


Figure 6.5.2 Forest plot of disease severity outcome versus neutralising antibody levels data from mouse studies that used a continuous severity outcome such as a multifactorial scoring system, or a measure with some other non-binary outcome such as temperature.

6.6 Mouse lymphocytes associate with protection

6.6.1 Concurrent biomarker analyses

The number of studies reporting positive and negative associations of lymphocytes with RSV disease severity were assessed (table 6.6.1.1) and the association between these cell types and RSV disease severity is shown (table 6.6.1.2). Four studies considered $CD4^+$ T cells (and not any specific subtype) and all of them demonstrated increased $CD4^+$ T cell abundance to be associated with more severe disease. Twenty-one studies contained some measurement of $CD8^+$ T cells and their association with more severe RSV disease, of which 17 showed a positive association, two showed a negative association, and two showed an inconclusive association. Thirteen studies measured total T lymphocytes association with RSV disease, of which nine showed a positive association, two showed a negative association, and two showed an inconclusive association. Where appropriate studies were analysed using standardized mean difference analyses to demonstrate effect size of a given cell type on severity (fig 6.6.1.3).

A variety of cell types have been studied, including total lymphocytes (n=13 articles), as well as subsets of lymphocytes, such as CD4 (n=18) and CD8 (n=22) T cells and Tregs (n=8) (table 6.6.1.2). In these studies, many different immunological variables were measured, including cell number and/or proportion (n=30), adoptive transfer of RSV infected lymphocytes to uninfected mice (n=7), immunodeficiencies (n=10), activity of lymphocytes in response to infection (n=9) and various molecular markers (n=19). Samples were from BALF (n=12), lung tissue (n=11), both (n=6) or serum (n=1) and were collected at variable time-points post-infection. A diversity of severity measures was used in these studies, including weight loss (n=16), lung resistance (RL) (n=11), lung dynamic compliance (Cdyn) (n=4) and a composite severity score (n=2). Ages of animals ranged from 6 to 10 weeks old. For our analysis, studies were included that focused on both primary and recurrent infection.

5 of 7 studies investigating T lymphocyte numbers indicated a positive correlation with severity (Table 1). Of these, 4 showed positive correlation with the duration of airway hyperreactivity (AHR) across mouse and rat models and in both primary and recurrent infection(483)(484)(485)(486). One article demonstrated an association between T lymphocyte numbers, illness severity and viral load(158). Two of 7 studies of T lymphocyte numbers reported a negative correlation with RSV titre, illness severity(476), and weight loss(487)(Table 1). In 6 of 7 studies that investigated CD8 cell number a positive correlation was found between CD8 cell number and various measures of disease severity, included RL, Cdyn(488), weight loss(489)(490) (Table 1), airway restriction, pulmonary inflammation, and renal injury(491). These studies included animals with both primary and secondary infection and various interventions, including recombinant RSV (rRSV)/ IFNγ(490) and anti-TNF-treatment(492). The results of one study conflicted with the results of the other studies that used CD8 T cells as a variable, reporting that reduced RSV-specific CD8 cell numbers corresponded with lower weight loss, yet greater viral load(492). All four studies of Treg proportions found a negative correlation with disease severity, despite no significant difference in RSV titre(181)(183)(493)(494)(Table 1).

Studies considered for this category involved transfer of lymphocytes into mice that were infected with RSV. CD8 T cell transfer in 2 of 3 studies was associated with increased disease severity, as measured by weight loss(152) and AHR(232); the third CD8 T cell transfer study, however, found increased lung viral clearance(495). Both CD4 T cell transfer studies found

greater weight loss(152), eosinophilia and AHR(232), and more severe immunopathology than CD8 T Cell transfer. One of the two studies of Tregs found increased cell numbers suppressed airway inflammation and accelerated recovery (based on weight loss)(181). However, another study showed mixed results, as Treg transfer from uninfected tolerised mice suppressed inflammation, whilst transfer from infected mice exacerbated airway inflammation(496). This suggests normal Treg cells have a protective, suppressive activity that is dysfunctional in infected mice. γδ T-cell transfer was associated with increased Th2 cytokines, airway inflammation, and allergic asthma in both studies(497)(498).

Many studies analysed the effect of cell types on RSV disease through depletion of those cell types. These studies often involved the depletion of the capacity of the mouse to produce immune cells, often by irradiation of lymph nodes, or using immunodeficient mouse strains without the capacity to produce certain types of cells. One study of immunodeficient nude mice with depleted lymphocytes found increased viral titres and more severe histopathology compared with wild type (WT) mice(151). Five of six Treg depletion studies found reduced Treg numbers correlated with increased weight loss, airway restriction and inflammatory responses(184)(180)(181)(493)(499). However, reports had mixed effects on viral clearance, with evidence for unchanged(181), delayed(180)(477) and enhanced(181) clearance. In three of five studies, CD8 T cell depletion was associated with increased weight loss, AHR, illness, viral load, and eosinophilia(158)(232)(500). Two of five studies had mixed results, with increased AHR, but reduced airway inflammation(501) and eosinophilia(157). CD4 T cell depletion also reduced AHR(232) and weight loss(158), and depletion of both CD4 and CD8 cells completely abrogated weight loss(158).

Two studies indicated positive correlation between CD8 cell responses as measured by some metric of activity (e.g. proliferation, cell-mediated killing and IFNγ production) and illness, but negative correlation with viral titre(500)(439). Similarly, four of six studies measuring T-cell activation exhibited an association with illness severity, weight loss, and RL(158)(502), as measured by Granzyme B, IFNγ(181) and CD11a(184) markers. However, one of those studies found a correlation with increased viral load(158), whilst another found no such association(502). Moreover, two studies showed no association between T-cell activation and weight loss based on T-cell CD69 expression(181)(499).

All 9 studies investigating CD8 T cells found a positive correlation between IFNγ+ and/or TNFα+ CD8 cells and weight loss(503), airway restriction(490)(158)(184)(181)(492)(499), viral titre and illness(487)(500).

Downregulated Th2/Th17 (GATA3, STAT6, RORγT) and upregulated Th1 (T-bet) responses were associated with decreased AHR, lung inflammation and RSV replication, although this was in cell culture, not *in vivo*(504).

Th1 responses are characterised by expression of IFNγ and TNFα. IFNγ+ and/or TNFα+ CD4 cells were measured in six studies, five of which exhibited positive correlation with viral replication, weight loss(499)(181)(503), AHR and lung pathology(183) across a range of interventions. However, one study found that the reduced IFNγ+CD4+ cell numbers also corresponded to an increase in weight loss(489).

Th2 markers include IL-4 and IL-13. All 5 studies that measured Th2 markers found a positive correlation with disease severity. Increased IL4+CD4 cells corresponded to greater weight loss(489), worse lung function, and greater inflammation(158), eosinophilia and mucus secretion(153). CD4-specific deletion of IL-4Ra attenuated these pathologies, suggesting Th2 cells have a role in disease(153). In addition, increased IL-13+ CD4 and CD8 levels in BALF corresponded to greater lung resistance abolished by anti-IL13 Ab(486).

Increased Th17 cells (defined by RORγT expression or IL-17+CD4+ numbers) were associated with more severe lung pathology and increased weight loss in all studies(183)(493)(494).

Standardized mean difference analysis showed a statistically significant associations between more severe RSV disease and CD4+ T cells (fig. 6.6.1.3 A), CD8+ T cells (fig. 6.6.1.3 B) and overall T cell populations (fig. 6.6.1.3 C) and a statistically significant association between Tregs and less severe RSV disease (fig 6.6.1.3 D). This indicates overall that most lymphocyte activity in the mouse model of RSV contributes towards more severe disease in this model.

Study	Animal	Sex	Age	Animals per group	Infection protocol and RSV strain	Main intervention/ observation	Sample	Control	Outcomes reported
Alwan 1992(152)	BM	F	3-4M	3-8	Primary, A2	T: CD4, CD8, TC	N	T: irradiated cells	WEIGHT LOSS, H, C, VL
Bukreyev 2001(505)	BM	F	11Wks	6	Primary, rRSV/mGM-CSF	N: CD4	L	rRSV, mock-infected	VL, C
Culley 2002(489)	BM	U	1D-12Wks	U	Recurrent, A2	P: CD4, CD8	L	UV-inactivated RSV	WEIGHT LOSS, VL, C, H
Dakhama 2005(486)	BM	U	0-3Wks	4-8	Recurrent, A2	N: lymphocytes	B	UV-inactivated RSV	RL, H, C, VL
Durant 2013(184)	CM	U	6-10Wks	11-13	Primary, A2	D: Treg. N: CD4, CD8	M	Uninfected	WEIGHT LOSS, H, C, VL
Fulton 2010(180)	BM	U	6-8Wks	4	Primary, A2	D: Tregs. N+P: Treg, CD8	M	Non-depleted	VL, IS, H, WEIGHT LOSS, Penh
Graham 1991(476)	BM	F	8-10Wks	24	Recurrent, A2	Lymphocyte aggregates	N	None	WEIGHT LOSS, VL, H
Harker 2007(490)	BM	F	4-8Wks	≥4	Recurrent, rRSV/IFNγ, rRSV/IL-2	P: lymphocytes. N: CD4, CD8	M	WT A2 RSV	VL, WEIGHT LOSS, H, C

Jiang 2012(488)	BM	F	8Wks	U	Primary, A2	N: lymphocytes	B	OVA/PBS only	RL, Cdyn, C, H
Jiang 2009(501)	BM	F	6-10Wks	10	Primary, A2	D: CD8. N: lymphocytes. P: TC, CD4, CD8	B	Control Ab, OVA/PBS only	RL, H, C
Krishnamoorthy 2012(496)	BM	B	3-8Wks	4-5	Recurrent, A2	N: lymphocytes, Treg. T: Treg	B	UV-inactivated RSV, uninfected	RL, C, H
Lee 2010(499)	BM	F	8Wks	5	Primary, A2	D: Treg. P: Treg. N: lymphocytes, CD4, CD8	M	PBS, control Ab	WEIGHT LOSS, VL, H, C
Loebbermann 2012(181)	BM	U	6-10Wks	4-5	Primary, A2	D: Treg. IL-2cx stimulation. N: CD4, CD8	B	Uninfected. Non-depleted	WEIGHT LOSS, H, C, VL
Munoz 1991(495)	CM	M	U	13	Primary, A2	T: CD8	N	No transfer	VL
Nagata 2015(493)	CM	F	6-10Wks	4	Primary, A	D: Treg (via SMYD3 KO). M: Th17	N	Non-depleted	C, H
Ostler 2003(487)	BM	U	6-12Wks	3-6	Primary, A2	N: CD8	B	UV-inactivated RSV	WEIGHT LOSS, C, VL
Peebles 1999(484)	BM	F	8Wks	6-10	Primary RSV+OVA, A2	N: lymphocytes	B	OVA only. RSV only	RL, H

Rutigliano 2004(500)	BM	F	8-10Wks	30	Primary, A2	D+N: CD8	L	Control Ab	WEIGHT LOSS, IS, VL, C, H
Schwarz 1999(157)	BM	F	8-12Wks	12	Primary, A	D: CD8	N	Mock-infected, no depletion	Penh, H, C
Schwarz 1999(232)	BM	F	8-12Wks	12	Primary, A	T: CD4, CD8, TC. D: CD4, CD8	N	Mock-infected, no transfer/ depletion	Penh, H, C
Shao 2015(477)	BM	U	6-8Wks	5	Primary, B1	D: Treg. P: Treg, CD8	L	Control Ab	H
Sun 2011(502)	BM	U	9-12Wks	5-10	Primary, A2	IL-10 KO, anti-IL-10R. N: CD8. P: CD4, CD8	L	WT animal, control Ab	WEIGHT LOSS, C, Cdyn, RL, VL, H
Tregoning 2010(492)	BM	F	6-10Wks	≥3	Primary, A2	Anti-CCL3 Ab and CCL3 KO. N: CD4, CD8	L	Control Ab. WT mouse	WEIGHT LOSS, VL, H, C
Tregoning 2008(158)	BM	F	4D-6Wks	≥4	Recurrent, A2	D/N/P: CD4, CD8.	L	Mock-infected	WEIGHT LOSS, C, Penh, H
Wang 2014(494)	BM	F	4Wks	20-40	Primary, A2	N: lymphocyte. P: Th17, Tregs	B	Mock-infected. OVA only	WEIGHT LOSS, VL, RL, Cdyn, H

Wang 2014(504)	BM	F	4Wks	20	Primary OVA/RSV, A2	N: lymphocytes. M: Th1, Th2, Th17	B	OVA only. RSV only	VL, RL, Cdyn, H, C,
Weiss 2011(183)	BM	F	6-8Wks	8	Primary, A2	IL-10 KO, anti-IL-10R. N: CD4, CD8	M	Control Ab. WT mouse	C, WEIGHT LOSS, VL, H, Penh
Yao 2015(503)	BM	U	8-10Wks	≥4	Primary, A2	Anti-PD-L1 Ab. N: CD4, CD8	L	PBS only, control Ab	WEIGHT LOSS, VL, C
You 2013(153)	BM	F	6-8Wks	5	Recurrent, A2	IL4Ra deletion on TCs. N: lymphocytes, CD4	B	WT animal. Mock infection	RL, C, H
Zeng 2014(497)	BM	F	8Wks	5	Primary, A2	T: γδ TC	L	PBS only. OVA only	C
Zhang 2013(498)	BM	F	6-10Wks	5	Primary, A2	T: γδ TC. N: lymphocytes, γδ TCs	M	Mock-infected, OVA	Penh, H, C
Zhang 2002[11]	BM	F	2-22M	10-12	Primary, A2	P: CD4, CD8	M	Uninfected	VL
Zhou 2008[95]	BM	F	8-12Wks	8	Primary, A2	D: nude mouse. N: lymphocytes	B	Mock-infected. WT mouse	VL, H, C

Table 6.6.1.1 Table of all studies used in analysis of relationship of lymphocytes with disease severity in this review. Abbreviations are as follows: ANIMAL - BM (BALB/c mouse), CM (C57BL/6 mouse; SEX - F (female animal), U (unknown sex), M (male), B (both male and female animals); AGE - W (weeks), M (months), D (days), U (unknown age); INFECTION AND RSV STRAIN – primary (single infection), recurrent (multiple infections), A/A2/ B1 (refers to the strain of virus), rRSV (recombinant RSV); MAIN INTERVENTION/

OBSERVATION – N (number of cells), P (proportion of cells), D (depletion of cells), T (transfer of cells), KO (knockout), TC (both CD4 and CD8 cells); SAMPLE – N (no sample), L (lung sample), B (bronchoalveolar fluid sample), M (both lung and bronchoalveolar fluid samples); CONTROLS – WT (wild-type), OVA (ovalbumin, used to immunologically challenge animals), PBS (phosphate-buffered saline control medium); OUTCOMES REPORTED – WL (weight loss), VL (viral load), H (lung histopathology), C (cytokine responses), RL (lung resistance), Cdyn (lung dynamic compliance), Penh (enhanced pause)

	Severity	Murine models	Citations
Lymphocyte responses	↑	$CD4^+$ [4/4] $CD8^+$ [17/21] Total T cells [9/13] Th17[3/3] γδ T-cells[2/2] Th1[5/6] Th2[5/5]	(184)(180)(490)(499)(181)(158) (500)(501)(183)(493)(494)(489) (153)(486) (503)(152)(232)(484) (485)(488)(491)(439) (502)
	No effect	Tregs [1/8] Total T cells[2/13] $CD8^+$[2/21]	(496)(501)(157)(181)(499)
	↓	Tregs [7/8] Total T cells [2/13] $CD8^+$[2/21]	(184)(180)(499)(181)(489) (476)(487)(492)(495)

Table 6.6.1.2 - Summary of reported associations with disease severity and lymphocytes in papers included in the systematic review. Each reviewed cell type or antibody isotype is listed as contributing positively to severity in the discussed animal model, with number of studies that support that relationship out of total studies with information on that cell type or antibody isotype. Where different subtypes of CD4 T cells were considered independently they were not counted towards the total number of CD4 studies. References to the studies data were acquired from are also listed

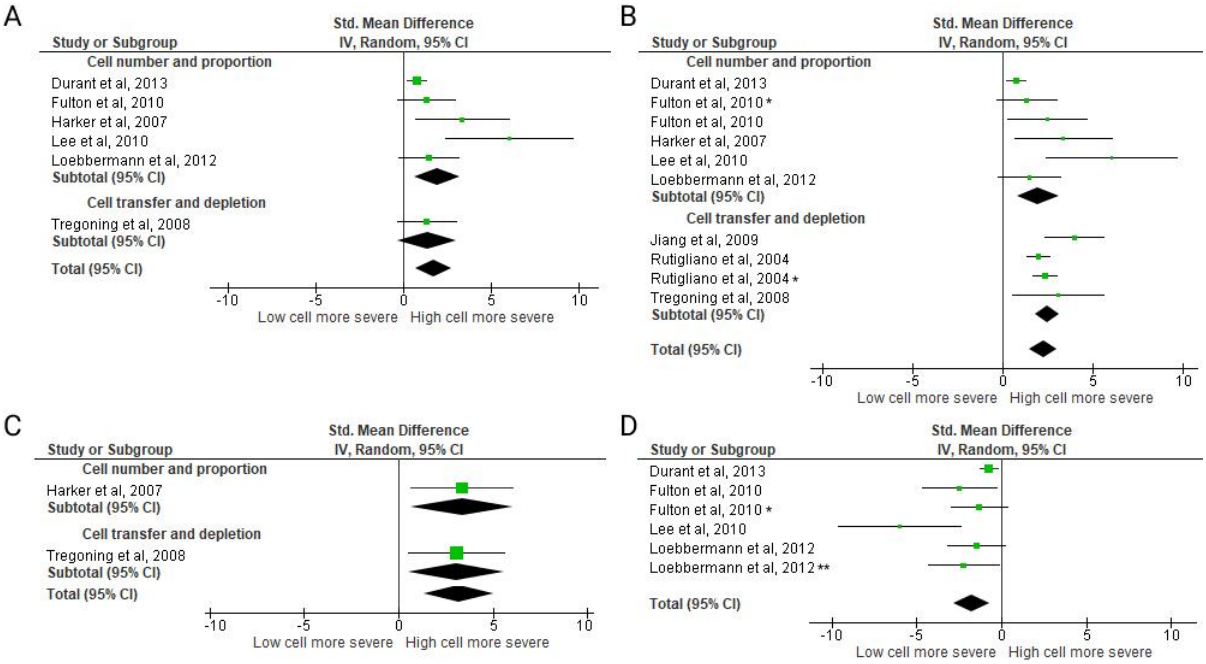


Figure 6.6.1.3 – Forest plots presenting random-effects meta-analysis with standard mean difference of disease severity measures between high cell and low cell number groups in mice, in cases where data on abundance of cell types is known prior to infection, such as in knockout mouse models generated with Review Manager 5.3 software. The 4 separate plots are for CD4 cells (A, p=0.001, I2=58%), CD8 cells (B, p<0.00001, I2=73%), T cells (C, p=0.0007, I2=0%) and Treg cells (D, p=0.0009, I2=57%). For panels B and D, * refers to studies using illness score as their severity measure, as opposed to weight loss in unindicated studies, and ** refers to a second experiment by Loebbermann et al, 2012[34], which uses IL-2 complex stimulation to alter cell numbers and disease severity, rather than cell deletion.

6.6.3 Predictive biomarker analyses

Some studies utilised time points prior to the onset of RSV disease, so analysis for these studies were considered separately to understand which elements of the immune system existing before infection have an influence on disease severity upon infection. A time point before the onset of disease or measurement of symptoms is required to assess earlier states that may lead to more severe disease later in disease. These biomarkers identified from data from earlier time points can be considered “predictive” of likelihood of developing severe disease upon subsequent infection. For predictive biomarkers, scatter plots showed there were significant correlations between CD4 and CD8 lung cell number and weight loss (Table 6.6.2.1 E, table 6.6.2.1 F), with r_s values of -0.83 and -0.68, respectively ($P < 0.05$).

With respect to correlations between measured cell abundances in mice and disease severity metrics, there was no statistically significant correlation between CD4 populations in the lung and weight loss (Table 6.6.2.1 A)(Rho = -.168, P = 0.32), but there was a statistically significant correlation with CD4+ proportion in lungs and weight loss (Table 6.6.2.1 B) (Rho = 0.94, $p < 0.05$), as determined by sensitivity analysis where neonatal data were excluded because these mice exhibited weight gain (342% and 502%) rather than weight loss (75-105%) (combined $r_s = -0.17$). Moreover, the statistically significant r_s value of -0.31 for CD8 number in lungs versus weight loss (Table 6.6.2.1 C) ($p < 0.05$, 2-tailed) supports that of the predictive plot. Finally, lymphocyte number was found to positively correlate with Cdyn (Table 6.6.2.1 D).

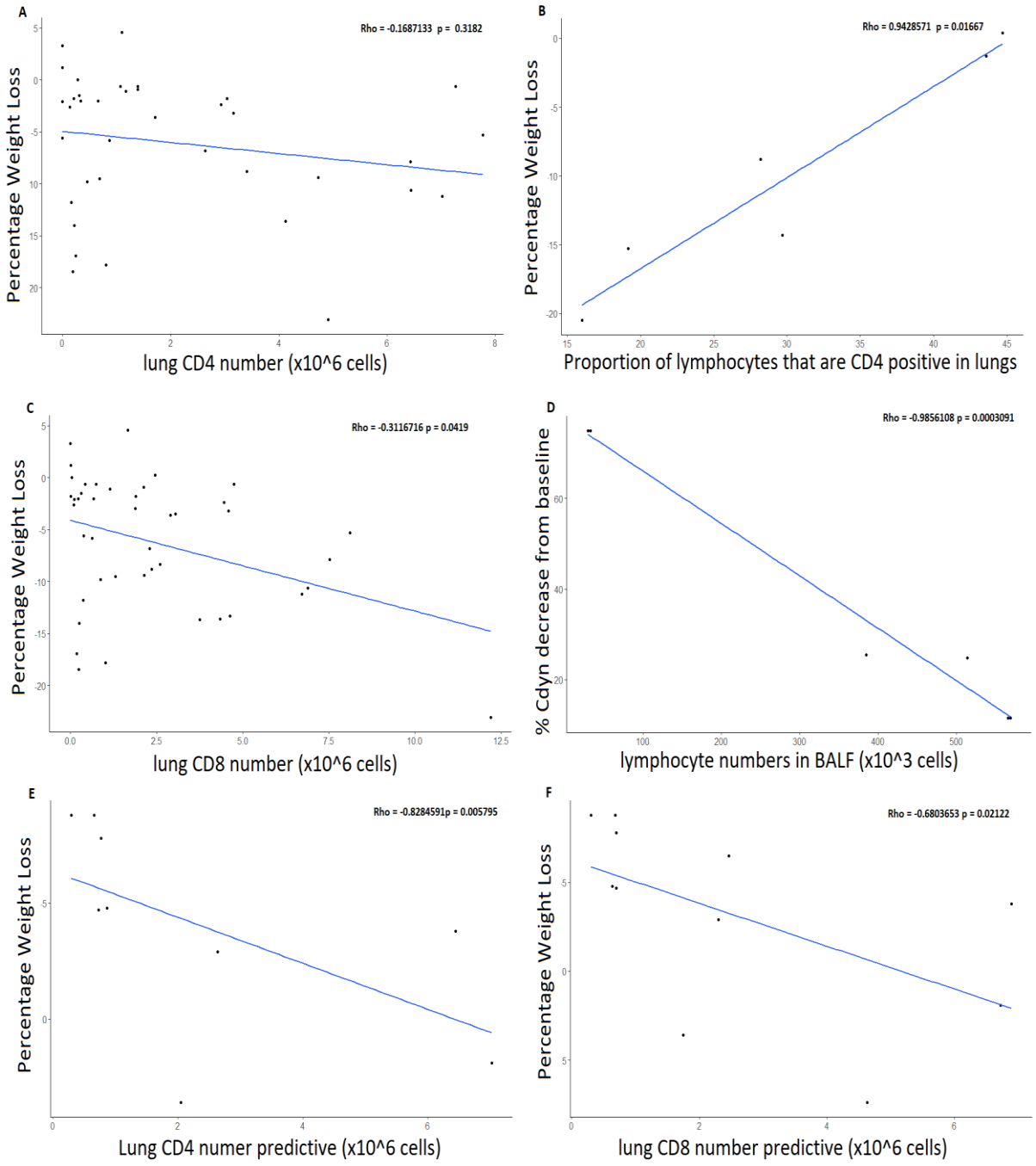


Figure 6.6.2.1 - Scatter graphs showing the relationship between various T lymphocytes (x-axis) and disease severity measures (y-axis, commonly weight loss but also Cdyn). Rho value was calculated for each plot. Increasing weight loss, with a smaller % of initial weight, indicates more severe disease, as does a smaller Cdyn. Plots A-D use all study groups where cell measurements are taken at the 4-8 days post infection, whilst E and F involve studies where lymphocyte measurements were taken prior to infection. Plot D refers to 12.5-15mg/ml of acetylcholine or methacholine used to challenge mice in whole body plethysmography, to calculate lung dynamic compliance (Cdyn). All studies measuring weight loss or Cdyn as an outcome measure studied mice between the age of 6-10 weeks of age. Each data point indicates one study group.

6.7 Discussion

6.7.1 Summary of main findings

Based on published data, markers of CD4 lymphocytic responses (especially, markers of highly active or Th2-skewed lymphocytic response and a relative lack of immunoregulatory Treg responses) are associated with severe disease in RSV-infected sensitized mice. CD8+ T cells may enhance pathogenic effects early during infection, while assisting viral clearance in the later stages. RSV neutralising antibody in mice is a potential biomarker of protection against severe disease, especially in the lower respiratory tract. However, evidence from bovine studies is less clear, only maternally derived antibody showing predictive effects in our meta-analysis.

6.7.2 Lymphocytic responses

There is clear evidence that strong inflammatory responses in the airway are associated with more severe disease in mice and that lymphocytes are central to the pathogenesis of lung disease. Higher lymphocyte numbers are significantly correlated with lower lung dynamic compliance (Cdyn), which is associated with increased airway occlusion. In humans and in cattle this can be attributed to increased populations of apoptotic or necrotic cells in the airway during a heightened immune response, however, this not yet been demonstrated in mice(506). It has been shown that pathogenesis in neonatal mice can be reduced by knockout of IL-4α, and that expression of IL-4α in response to RSV infection contributes towards a further Th2 bias(507) and increased disease severity.

Increased activation and numbers of CD8 T cells in the lungs of RSV infected animals is associated with more severe disease, but also with viral clearance. Moreover, a statistically significant positive correlation between lung CD8 number and weight loss supports the further investigation of low CD8 T cell numbers as a potential biomarker to predict severe RSV disease. Based on this evidence, CD8 T cells have a role in enhanced disease severity in acute disease but also in eventual viral clearance.

The proportions of lymphocyte subtypes in the mouse model are among the most promising biomarker candidates for disease severity. Increased Th1, Th2 and Th17 responses may all be associated with greater disease severity. Th2-skewed responses in general were more severe than Th1-skewed responses, but this may need additional study. Th1 responses having a stronger capacity to control and limit infection than other T lymphocyte subtypes and accordingly have a lower correlation with disease severity than other T lymphocyte subtypes that are not as capable of controlling or limiting viral infection. Treg numbers were generally associated with accelerated recovery and reduced disease severity in mice. Regulatory T cells inhibit the development of other T cell subtypes that are associated with more severe disease in mice.

It has been observed in humans that frequency of activated Tregs in the blood is reduced in RSV infection(176). Dampened Treg responses correlate with increased Th2 responses(508), which are associated with more severe disease in humans(171). Decreased Treg numbers may result in more severe infection primarily due to a decreased capacity to control inflammation in infection.

6.7.3 Antibodies and RSV disease severity

The most prominently identified association between antibodies and disease was between total RSV-specific Ab titre and more severe RSV infection in mice. However, Abs are not only protective against reinfection but are also involved in preventing severe disease through reducing viral replication prior to disease onset and cellular immune responses(509). Indeed, some associations between Ab titres and severe disease were observed,(459) but age, particularly, was a confounder in many studies. Neutralising antibodies do not appear to be an appropriate biomarker for disease severity in the bovine model.

The collective bovine data presented here shows little or no association between RSV specific Ig and disease severity. However, the studies that revealed a positive correlation between neutralising antibody and severe disease had notable confounders, such as age or allergic sensitisation, making them less reliable to assess biomarkers for disease severity(459)(463)(460). Conversely to the ambiguous role of RSV neutralising antibody, maternally derived antibodies in the bovine model demonstrate a consistently protective role and exist primarily in the group at the highest risk of disease.

Interestingly, the results of this study show that the different isotypes of Igs are poor biomarkers of disease severity on their own in both cattle and in mice, due to inconsistent relationships with disease severity. This is unexpected in particular for IgE, which has frequently been associated with airway hypersensitivity and Th2 responses in RSV infection(510)(479). As IgE is more often implicated in airway hypersensitivity, it is possible that the metrics of disease severity used in these studies do not capture the associated longer-term allergic symptoms. Similarly, it is not unexpected that evidence for IgM or IgA as a biomarker in either model would be conflicting or insubstantial. These isotypes have relatively poorly characterised roles in RSV disease, although they are consistently detectable after primary RSV infection in humans(511).

From our analysis, the mouse offers a better model for understanding the role of antibodies in human disease severity than the bovine model, since consistently protective antibody response observed in this study more closely resembles that in humans(442)(107)(443)(444)(445)(446)(509) than the inconsistent bovine neutralising antibody response.

6.7.4 Limitations

It is conceivable that some associations investigated in this review were masked by poor methodology or reporting in some studies. Mice are only semi permissive to hRSV replication, and do not naturally experience severe respiratory disease on hRSV exposure. bRSV, however, can result in fatal respiratory disease in calves and thus potentially offers a representation of the pathology closer to that of humans. Another complicating factor is the time at which biomarkers were assessed in animal models in comparison to humans. For example, many of the studies of Ab in infants measured Ab soon after birth to predict RSV disease over the next few months. Ab was often tested in animal models just prior to challenge with RSV (usually no more than 24 hours prior to challenge). These differences in timing of measurement of biomarkers may significantly influence the findings from each study.

Measures of disease severity used in animal models are not necessarily reflective of the human experience of disease, and as such biomarkers for disease in these models do not

necessarily reflect disease in other species with differing manifestations of disease. The most commonly used elements of severity scores in animal models were weight loss and histopathological features of the respiratory tract, which do not have commonly used counterpart disease severity readouts in humans. Aspects of the pathogenesis of disease also differ between models, as despite being one of the primary cell types driving severe RSV disease in humans, neutrophils may have a more protective role in mice(512). Nevertheless, these models remain important for the improvement of our understanding of RSV disease as they allow studies to be undertaken that are unethical or impossible in humans.

Many mouse experiments used genetically altered mice infected with human RSV to assess various cell functions. Due to the regulatory role of certain cells (in particular CD4 lymphocytes) in regulating other aspects of the immune response (for example antibody development) it may be questionable to attribute the effects of these interventions to these cells alone. While the importance of these cells as regulators of RSV disease based on the evidence shown here cannot be denied, the fact that they have regulatory roles in other processes creates difficulties in understanding which conditions created by changes to these relatively central regulators are in fact responsible for changes in disease severity.

6.7.5 Conclusions

Bovine models of RSV infection do not show a clear association between RSV specific serum Ab titres and ensuing severity of RSV disease. Studies in mice suggest that low Ab concentrations are predictive of greater disease severity, but these studies may be confounded and show risk of bias. The most promising biomarkers of disease severity in mice are measures of lymphocyte number and differentiation, especially if measured directly at the site of infection. These findings suggest that T lymphocyte numbers, activation and differentiation are possible biomarkers that might be explored also in human studies. Due to differences in both pathogenesis of disease, and measurements of disease between these models and humans, applicability and relevance of these biomarkers in a clinical setting is difficult to predict and requires further targeted investigations, especially those focused on mucosal responses. The animal data can suggest hypotheses to be tested

in clinical studies, but for some weaker biomarkers more evidence from animal model studies is required.

Chapter 7: Discussion

7.1 Introduction

RSV is one of the leading causes of childhood mortality worldwide, however the relationship between the inflammatory response to the virus and clinical disease has remained poorly characterised. As the inflammatory response is central to the pathogenesis of RSV, this results in a lack of capacity to adequately stratify patients according to risk of more severe disease. Advances in "Omics" technologies in the last two decades have resulted in a never before seen capacity to comprehensively view and understand the labyrinthine complexity that is characteristic of biological systems. This thesis uses this capacity in analysis of gene expression to characterise the immune response to RSV, its relationship with severe disease, and to create a robust discriminatory biomarker for severe disease.

The RESCEU consortium has facilitated the gathering of the data in the present study by enabling the gathering of a huge variety of samples from children with RSV from an equally varied set of countries, enabling the largest and most comprehensive study of RSV immunology to date. The scope of this study encompasses many fields, including immunology, genomics, virology and epidemiology, and by combining and comparing results across these disparate fields, it may be possible to reach a new level of understanding of RSV disease that could catapult us towards new treatments and perhaps a functional vaccine.

7.2 Summary of the principal findings of this study

7.2.1 The immune response to RSV

The immune response to RSV in infants in this study is characterised primarily by upregulation of neutrophil, monocyte and type I interferon associated gene expression and downregulation of T and B lymphocyte related gene expression. Neutrophils and monocytes show increases in both numbers and activity in the whole blood of children with RSV, and gene expression associated with signalling by early inflammatory mediators is observed, particularly of IL-1, IL-2 and IL-6. This signalling is in line with what has been observed previously in both RSV and influenza, where signalling by early responding myeloid cells,

such as macrophages, neutrophils and monocytes produces early inflammatory mediators such as interferon alpha and beta(513) and IL-6(514) that in turn encourage the proliferation of B cells(98) and eventually the differentiation of T cells into Th1 CD4 T cells and CD8 T cells to encourage clearance of infection.

Gene expression associated with markers of humoral and anti-microbial immunity is also seen early in infection in this study and is reflective of the increase in gene expression of neutrophil secreted products that have previously been observed in at this time point, particularly of traditionally anti-bacterial secreted products such as azurophilic granules(239). Production of these has been previously characterised in RSV infection *in vitro*(515). It has been hypothesised that production of these molecules helps to control the expansion of commensal bacteria such as *Streptococcus pneumoniae*(239) that have been associated with worsened outcomes in RSV and Influenza infections(516)(373), however this hypothesis remains in its infancy and the role of these secreted neutrophil products in viral clearance remains unclear.

Downregulation of gene expression associated with lymphocytes, such as antigen receptor mediated signalling is also seen in acute infection and is reflective of the pronounced lymphopenia consistently observed in RSV and Influenza infection. The downregulation of B cell related gene expression in early infection is unexpected, as early proliferation of B cells in response to type I interferon secretion is typically expected in respiratory viral infections. This may be a product of the elevated populations of B cells observed in convalescence in this study, as the convalescent time point was used as a control. Memory B cell populations have been observed to increase significantly across the first six months of life(368). Decreased numbers of both CD4 and CD8 T cells are also commonly seen in RSV infection. This lymphopenia has been hypothesised to be due to increased expression of checkpoint inhibitor receptors such as Fas and TRAIL in response to viral infection(517), potentially as a viral immune evasion mechanism.

7.2.2 RSV and severity

More severe RSV disease, as defined by requirement for invasive respiratory support is characterised by increased neutrophil and monocyte associated gene expression alongside

decreased CD4 T cell, CD8 T cell and B cell associated gene expression, similar to what has been previously characterised in the literature.

The strongest relationships between disease severity and gene expression were secreted products of neutrophils such as azurophilic granules. Production of these granules by neutrophils in response to RSV has been observed *in vitro* where it was found to associate with some cytotoxic activity(515). RSV has also been shown to induce neutrophil elastase secretion in children, however no link was found with disease severity(518). In Influenza, release of NETs by neutrophils in response to infection has been found to correlate with more severe disease(519). Further studies analysing the relationship between neutrophil granules and disease are required to fully understand the influence of these granules on disease severity.

Less pronounced cases of lymphopenia in both RSV and influenza have correlated with less severe disease, so it is not surprising that a similar relationship was observed here. Lymphopenia has been identified as a specific risk factor for RSV LRTI(520), and other microarray studies have shown association between reduced T cell numbers and disease severity in RSV(276). Similarly, lymphopenia associates with worsened outcome in influenza infection(521) and SARS-COV-2 infection(522).

Similarly, lower B cell numbers have been associated with more severe disease in other respiratory viruses, a relationship that has also been observed in this study. Neutralizing antibodies have been demonstrated to reduce immunopathology in RSV infection(445). Presumably, any decreased capacity to generate these responses in a timely manner due to reduced B cell numbers would result in greater immunopathology. In mice never exposed to influenza and kept in germ free conditions, some IgM activity attributable to “innate-like” B cells has been shown to reduce viral loads and severity of infection(523)(524)(525), suggesting potential protective roles for even non-specific B cells early in respiratory viral infections in infants. Transfer of antigen specific antibody has been shown to reduce immunopathology in Lymphocytic choriomeningitis virus (LCMV) vaccine induced cytokine storm(526), suggesting that generation of antibodies over the course of an infection could reduce immunopathology.

An association between gene expression related to CD4 lymphocytes and disease severity was observed on the second component of disease severity PLS-DA. This could be reflective of a population of differentially activated proinflammatory RSV infected lymphocytes(380), or some other small population of lymphocytes with a different activation state. Viral infection of immune cells that induces a different activation state has been previously observed in RSV(82) and resulted in an immunoregulatory phenotype. It is possible that this gene expression signature represents such a population and would be an interesting target for future confirmation work using cytometry.

7.2.3 Biomarkers for severe RSV

The biomarker for severe RSV identified in this study comprises five genes and participant age. These genes were *IL1R2*, *OLFM4*, *MMP8*, *BCL6* and *ANXA3*. These genes can be grouped according to function, with the different functions reflecting aspects of the inflammatory response with some importance in RSV disease severity. Interestingly, many of these genes (in particular *IL1R2* and *OLFM4*) are traditionally associated with immunoregulatory effects(527)(292), despite showing increased expression in more severe disease. While this increased expression in more severe disease may seem counterintuitive in light of the associations with severe disease observed in chapter 4, it is possible that these genes show such high expression in severe disease because the participant in question is already experiencing immunopathogenesis, resulting in upregulation of these genes to compensate. Both genes are regulated by NF-kB, and *IL1R2* shows a rapid response profile to increased NF-kB expression upon infection(292)(528), suggesting that it may be a key factor in an anti-inflammatory feedback loop that controls excessive inflammation during infections. A correlation between the dampening of excessive inflammation and improved viral clearance has been characterised in Influenza infection(529).

These results are broadly in line with other studies that have aimed to identify biomarkers. Brand et al identified *OLFM4* as a discriminatory biomarker for severe RSV previously(275). Other studies that have used gene expression to develop predictive signatures for severe disease have identified similar families of genes, with the signature identified by Jong et al containing *ANXA3*, *MMP8* and *OLFM4*(278).

Biomarkers in animal models identified by systematic review highlight the issues of using animal models to explore disease phenotypes. The relationship between $CD4^+$ T cells in mice and disease severity show some differences between that observed in humans, with higher $CD4^+$ and $CD8^+$ T cell activity associating with more severe disease and faster viral clearance. A skew towards Th2 type immunity in mice also associated with greater disease severity. Markers of specific T cell subsets, such as Th2 cells or regulatory T cells were the best potential biomarkers for RSV disease identified in animal models in this study.

7.3 Limitations of this study

As the majority of participants recruited to this study were recruited in hospital, there is a bias towards those who developed significant symptomatic disease. This may be desirable for the generation of a biomarker within the hospital setting, however to better understand the mechanisms that underly disease severity, a larger population of infants with mild infection (infants who were not admitted to hospital, discharged early or recruited from the community) is desirable. More samples that meet these criteria were collected in the final year of the study after the generation of this transcriptomic data and will be included in the generation of subsequent transcriptomic data.

Usage of convalescent samples as control samples introduces potential sources of bias (e.g. age) in functional analyses of gene expression, while also providing the benefits of pairing, and thus removing interindividual variation. The effect of age can be very pronounced in participants below the age of six months and can result in differences in gene expression due to different cell populations. B cells in particular show a large amount of change over the first six months of life(368).

More modern technologies also exist that may have increased the power of this study. In particular RNA sequencing allows for a more granular view of gene variant expression than is typically possible with a microarray, is not limited to a pre-specified probe set and captures very low and very high expression values better than microarray. However, RNA-sequencing is more expensive; therefore, financial constraint may have limited the number of samples that could be run. As the number of participants has an impact on study power, any reduction in samples size will likely have detrimental effects on study power, as evidenced in power calculations shown in chapter 2.

This study used microarrays on whole blood, which potentially shows a different picture of gene expression in response to disease than sampling from the tissues of the respiratory tract, where accumulations of cells have been shown to associate strongly with both disease and recovery(149). While in biomarker analyses the tissue sampled is irrelevant if predictive power is sufficient, functional analyses of results may show significant differences dependent on tissue that reflect differing immune responses in different tissues. Indeed, blood may be the most suitable tissue for biomarker analyses due to its relative ease of access compared to lung tissue.

7.4 Future directions

One potential utilisation of this data set is the exploration of respiratory sequelae of RSV infection. Analyses of these was not possible within the scope of the current analysis as data on sequelae is gathered in update questionnaires in years following the study, and as such a full complement of data were not available for the majority of participants. In future it will be possible to use these data to determine the role of gene expression at the time of disease or in convalescence in contributing to the development of asthma and persistent wheeze in the months and years post RSV infection.

The relationship between age and the patterns of gene expression associated with disease in this study is an area for potential future development. Age has been considered one of the primary markers of susceptibility to severe RSV disease. It is possible that one of the key mechanisms by which age contributes towards more severe disease is through lower B cell populations that exist at those early time points.

Challenge models, while restricted to adults may provide a context in which we can further understand the mechanisms of severe disease, particularly the relationships between gene expression associated with secreted neutrophil granules and severe disease observed here. Using a challenge model, it may be possible to assay for expression of specific secreted products in the lung and assess the statistical relationships between those products and both viral load and severe disease. Once these relationships are confirmed in this way, further work may be done in mice and cell lines to determine the direct interactions between both the virus and these molecules, and the virus and host tissues. In this way we might develop a more comprehensive understanding of both the means by which

neutrophils contribute to more severe disease, and the factors that trigger their immunopathogenic activity.

Additionally, further data are being collected within the RESCEU consortium that will complement this data set in the future. CyTOF, GWAS, respiratory microbiome sequencing, and neutralizing antibody analysis are just a few of the analyses that will be performed in the future. The combination of all of these diverse types of data produces unique challenges in integration and comparison, but equally creates an opportunity to fully understand what drives RSV disease.

A validation study for the identified biomarker using samples from the respiratory tract would be an interesting direction in which to take this work that would also aid in understanding the between the transcriptomic response to RSV in whole blood and in respiratory tract, which remains poorly understood.

Any usage of a biomarker in a clinical setting is likely to be performed using a point-of-care test, such as a qPCR based test. A validation study using the samples collected within this trial to determine whether the gene signature is readily identifiable with qPCR, and whether it retains its discriminatory power would be desirable to further emphasise potential clinical applicability of the biomarker.

7.5 Conclusions

This thesis provided a comprehensive overview of the relationship between gene expression and RSV disease severity, which is defined by lymphopenia and overexuberant neutrophil and monocyte responses, as well as showcasing an effective minimal biomarker capable of discriminating between mild and severe disease. Further assays conducted within this trial will help to understand the mechanisms underlying these relationships and will generate hypotheses that will drive the next stage of research in host responses. Additionally, further validation of the biomarker shown in this study and further honing of the methods used to create it may allow for the realisation of a robust in clinic biomarker for RSV severity in the coming years.

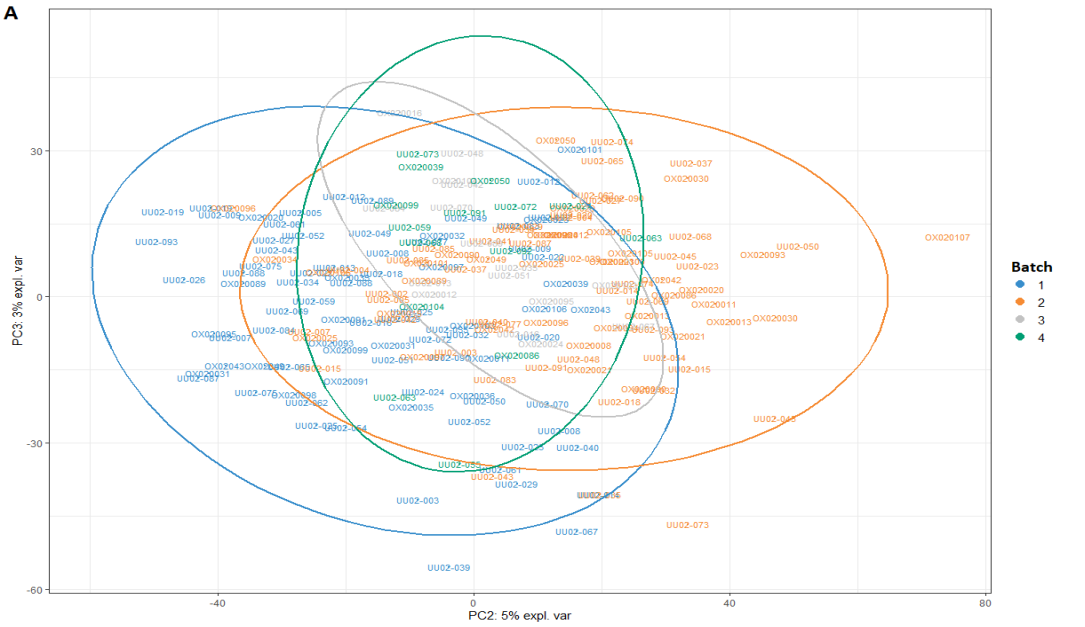
Appendix chapter 1: Supplemental materials for chapter 3

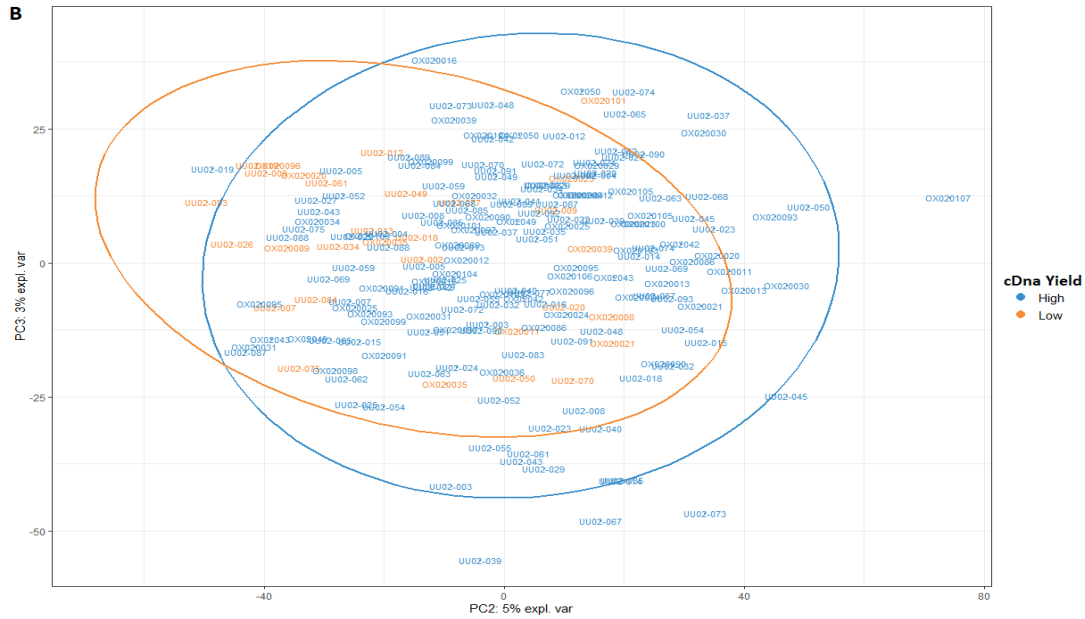
S1.1 Quality control of microarray data reveals signal associated with RSV infection

S1.1.1 The majority of variation between samples prior to quality control is attributable to technical variation

Principal component analysis was also performed on the raw data prior to any further quality control steps to determine the potential systematic effects introduced due to experimental design prior to correcting for these sources of bias. Principal component and probe intensity analyses after implementation of quality control steps is a useful tool in determining the effects of quality control measures on data.

Principal component analysis of raw data prior to normalization and filtration found the first principal component to account for 60% of the variance in the data set (fig 3.3.2.2), with the second principal component account for 5% of the variance in the data set, while the third principal component accounted for 3%. No distinct clustering was visible on the second and third principal components (Sfig 1.1.1.1). Some separation due to batch effects (sfig 1.1.1.1 A) and cDNA yield (sfig 1.1.1.1 B) is seen on the second and third components, however the majority of technical variation attributable to both appears on the first principal component (fig 3.3.2.3). No separation or clustering due to disease status is visible on the first and second principal components (sfig 1.1.1.1 C).


A
PC3: 3% expl. var
PC2: 5% expl. var
Batch
1
2
3
4


B
PC3: 3% expl. var
PC2: 5% expl. var
cDna Yield
High
Low

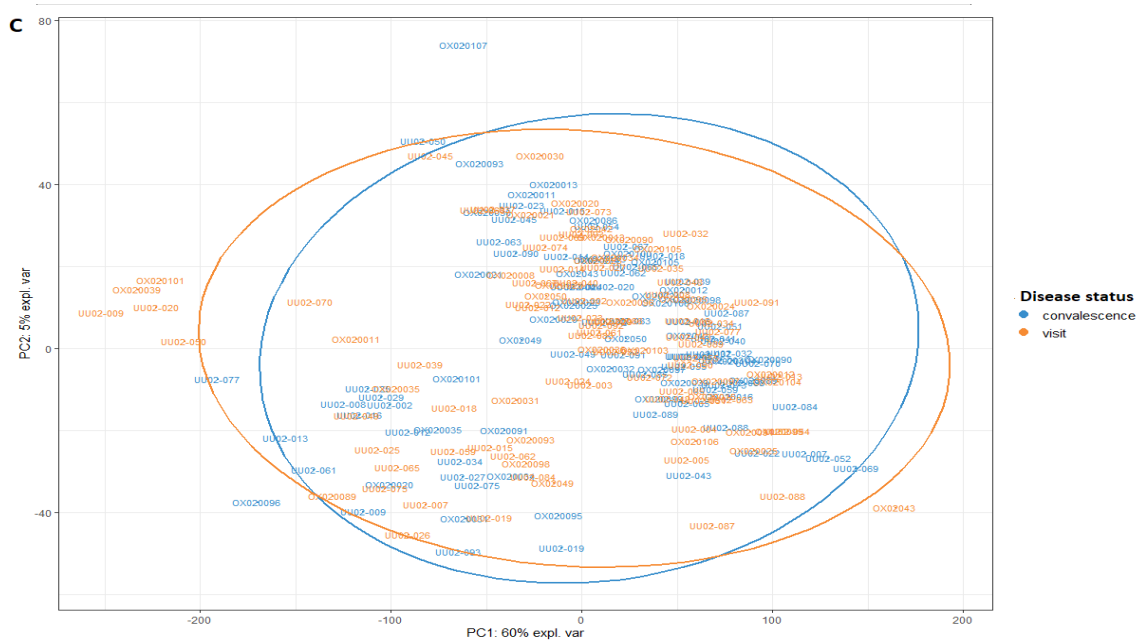


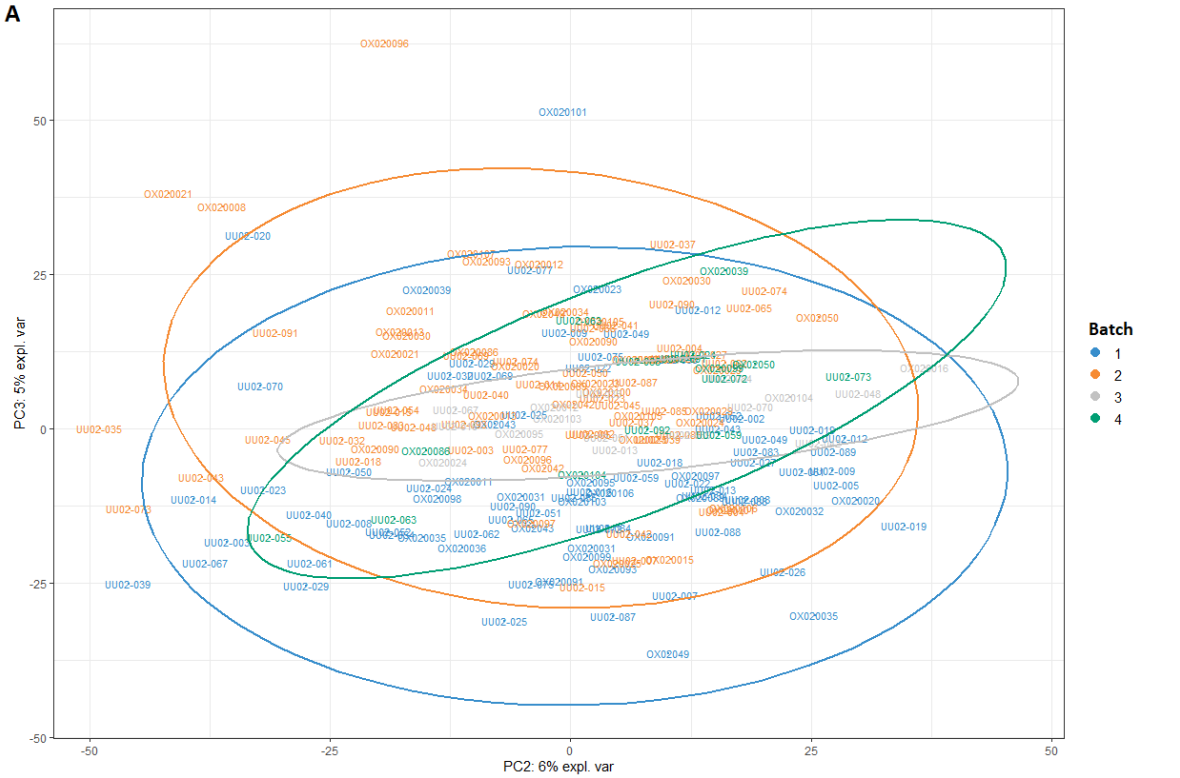
Figure S1.1.1.1 Principal component analysis of samples before quality control Grouping of samples according to expression of probes in the A: second and third principal components coloured by microarray batch, B: second and third principal components coloured by cDNA yield and C: first and second principal components coloured by disease status is shown. Expression of the first principal component is mapped on the x axis and expression of the second principal component is mapped on the y axis. Samples originating from patients recruited in Oxford are labelled with the study title "OX02", while the samples from patients recruited in Utrecht are labelled as "UU02". Samples are coloured according to infection status of the patient they were taken from.

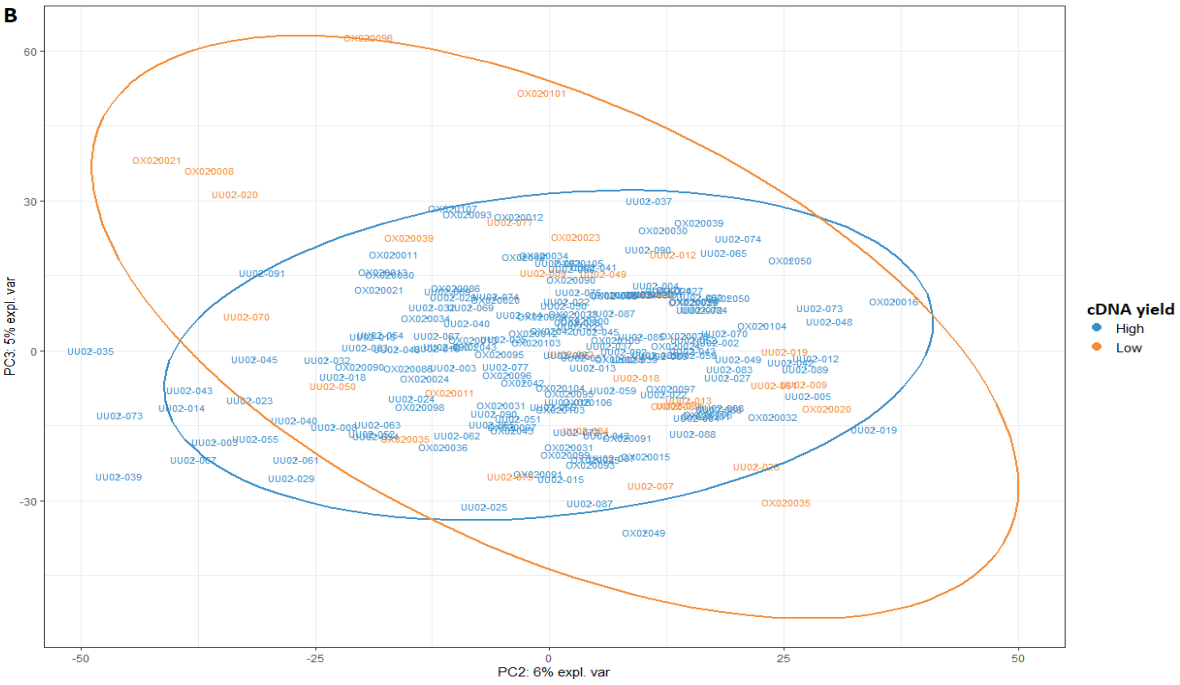
S1.1.2 Normalization of array data reveals distinct signal of RSV infection

To ensure comparability between samples quantile normalization(340) was used on the probe intensities, as well as filtration of probes with no variance in at least half of all samples and filtration of sex chromosome associated probes.

PCA was also performed on all samples after normalization to show effects that normalization may have had on sources of variation that may exist within the data.

Principal component analysis of data after normalization and filtration and filtration of low variance probes found the first principal component to account for 28% of the variance in the data set (fig 3.3.5.2), with the second principal component account for 6% of the variance in the data set, while the third principal component accounted for 5%. No distinct clustering was visible on the second and third principal components (Sfig 1.1.3.1). Some separation due to batch effects (sfig 1.1.3.1 A) and cDNA yield (sfig 1.1.3.1 B) is seen on the second and third components, however the majority of technical variation attributable to both appears on the first principal component (fig 3.3.3.3). Some clustering due to disease status is visible on the first and second principal components (sfig 1.1.3.1 C).


A
PC3: 5% expl. var
PC2: 6% expl. var
Batch
1
2
3
4


B
PC3: 5% expl. var
PC2: 6% expl. var
cDNA yield
High
Low

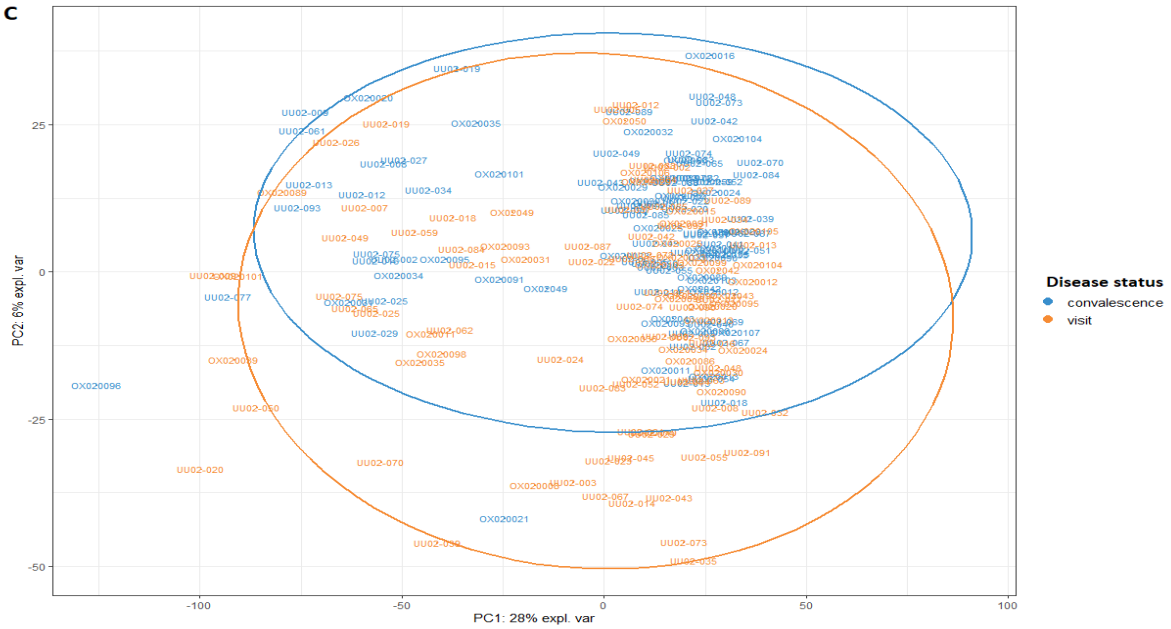


Figure S1.1.3.1 PCA mappings of samples after RMA normalization, sex chromosome removal and probe filtration Grouping of samples according to expression of probes in the A: second and third principal components coloured by microarray batch, B: second and third principal components coloured by cDNA yield and C: first and second principal components coloured by disease status is shown. Expression of the first principal component is mapped on the x axis and expression of the second principal component is mapped on the y axis. Samples originating from patients recruited in Oxford are labelled with the study title "OX02", while the samples from patients recruited in Utrecht are labelled as "UU02". Samples are coloured according to infection status of the patient they were taken from.

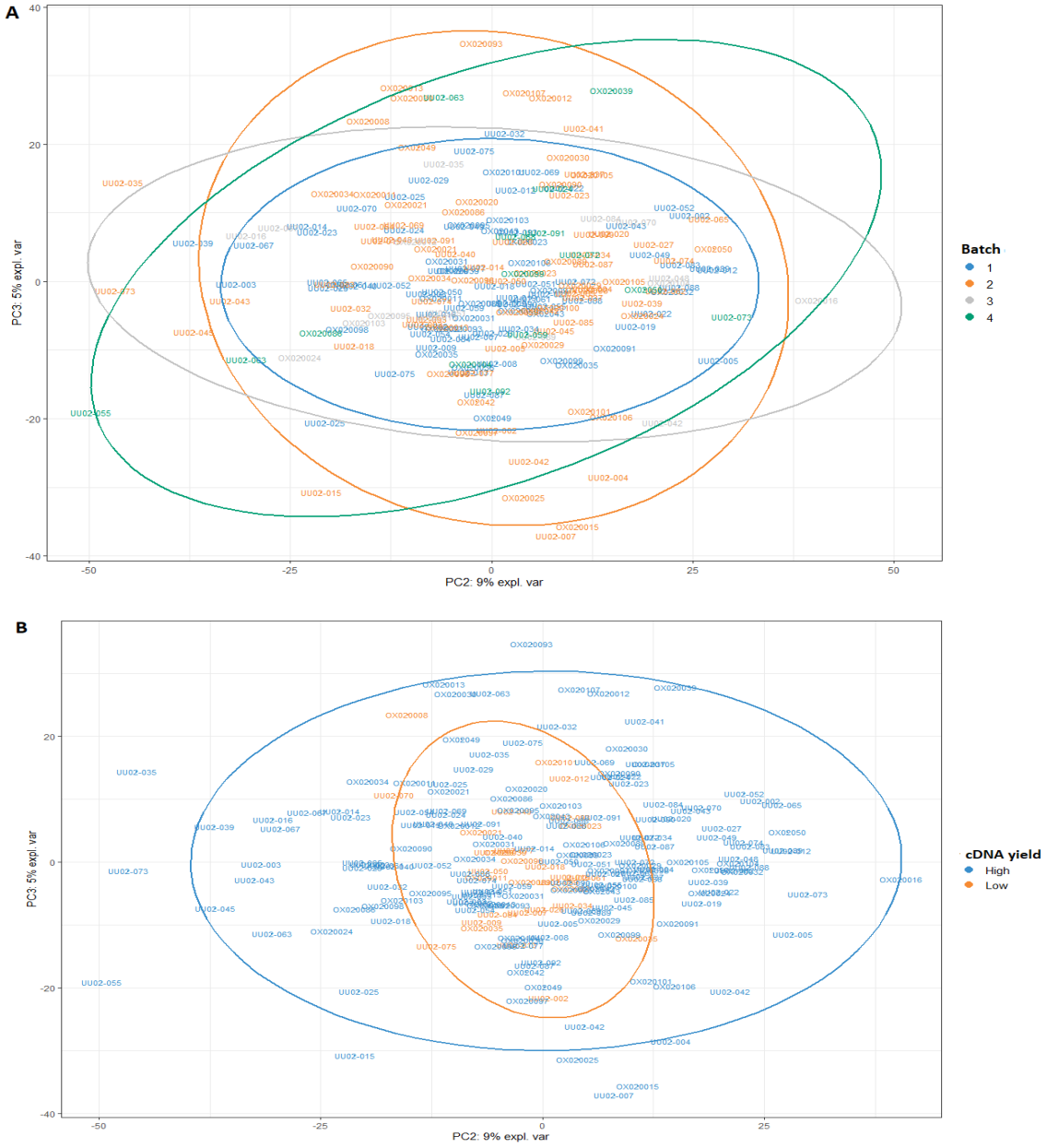
S1.1.3 Batch correction with ComBat removes effects of microarray batches and cDNA yields

After normalization, batch correction was used to remove the systematic effects resulting from the batches in which the samples were run on the microarray, and the effects of cDNA yields.

PCA was also performed on all samples after normalization to show effects that normalization may have had on sources of variation that may exist within the data.

Principal component analysis of data after normalization and filtration and filtration of low variance probes found the first principal component to account for 12% of the variance in the data set (fig 3.3.3.4), with the second principal component account for 9% of the variance in the data set, while the third principal component accounted for 5%. No distinct

clustering was visible on the second and third principal components (Sfig 1.1.2.1). Some separation due to batch effects (sfig 1.1.2.1 A) and cDNA yield (sfig 1.1.2.1 B) is seen on the second and third components, however the majority of technical variation attributable to both appears on the first principal component (fig 3.3.2.3). Distinct due to disease status is now visible on the first and second principal components (sfig 1.1.2.1 C).



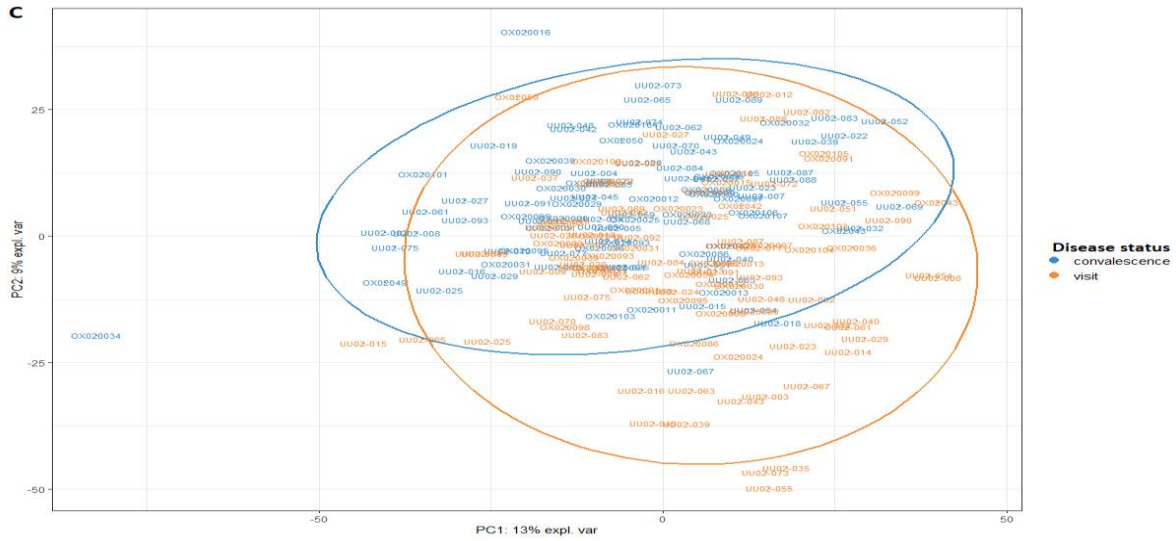
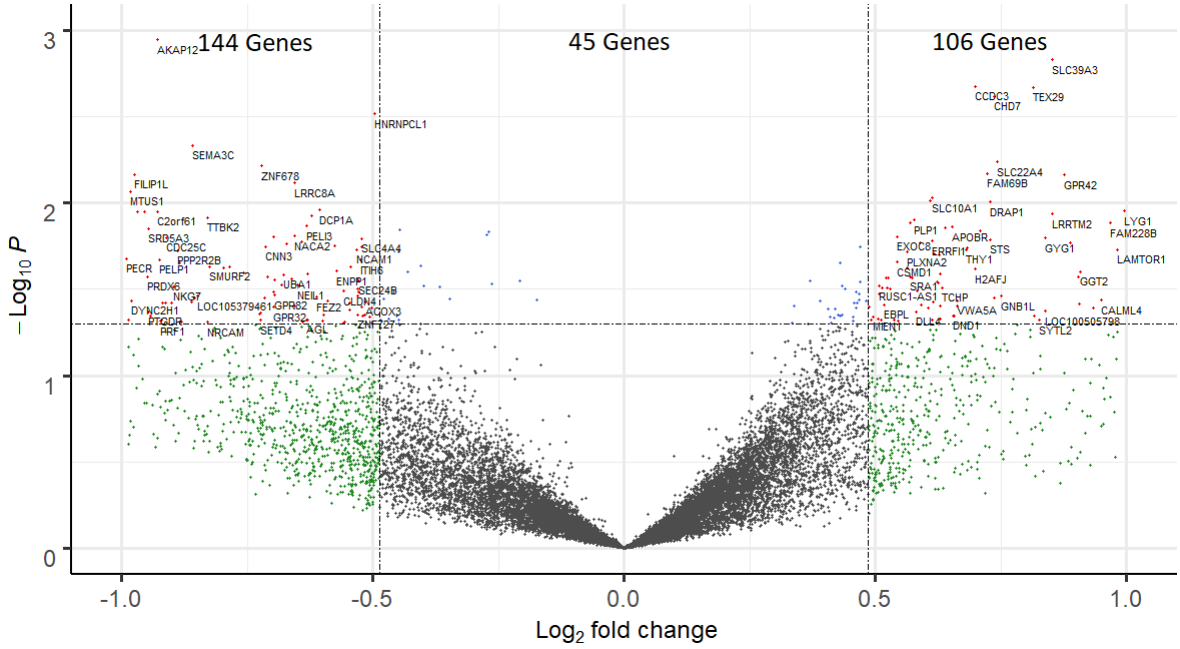


Figure S1.3.1 PCA mappings of samples after normalization and background correction Grouping of samples according to expression of probes in the A: second and third principal components coloured by microarray batch, B: second and third principal components coloured by cDNA yield, C: first and second principal components coloured by disease status. Expression of the first principal component is mapped on the x axis and expression of the second principal component is mapped on the y axis. Samples originating from patients recruited in Oxford are labelled with the study title "OX02", while the samples from patients recruited in Utrecht are labelled as "UU02". Samples are coloured according to infection status of the patient they were taken from.

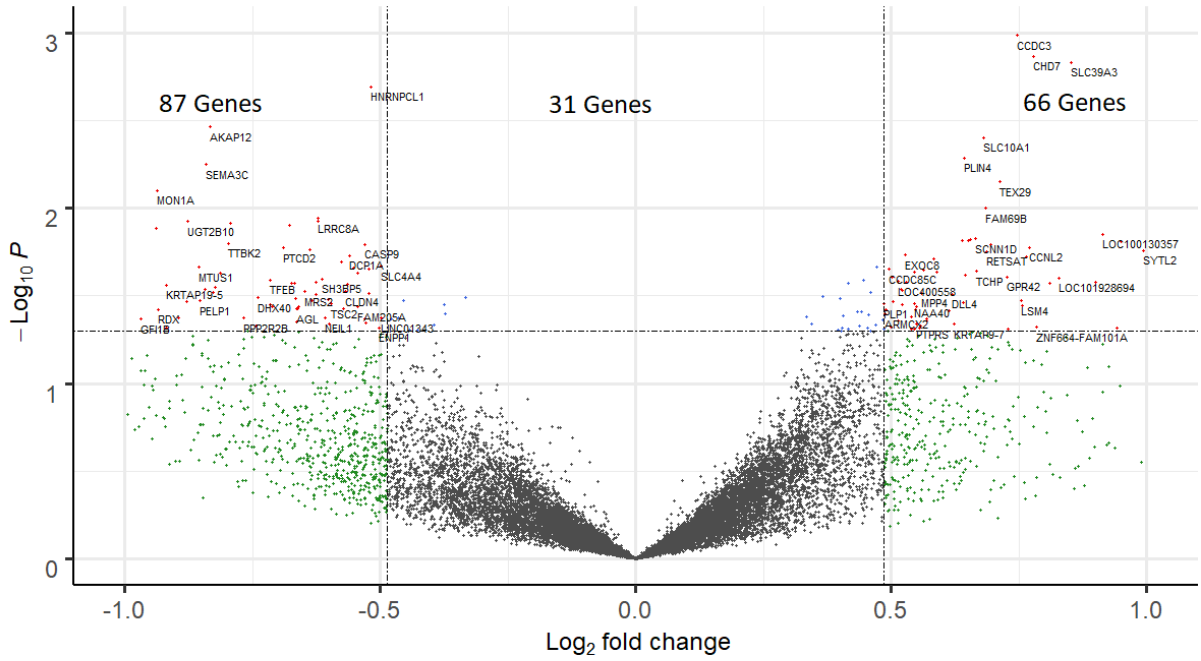
S1.2 Differential expression of genes between participants with RSV infection and healthy controls

In addition to the comparisons between participants with acute RSV and participants in convalescence, comparisons with the seven healthy control participant samples available in this study were also made.

Between the seven healthy controls and ninety-eight RSV infected patients 250 genes were found to be differentially expressed (sfig 1.2.1), according to a nominal p value threshold of .05 and a fold change threshold of 1.4. 144 of these genes were downregulated and 106 were upregulated. Between the healthy controls and the patients in convalescence (sfig 1.2.2), 153 differentially expressed genes were seen, with 87 being downregulated and 66 being upregulated. Nominal p value was used as significance threshold due to a lack of significance of any genes when false discovery rate adjustment was used.



sFigure 1.2.1 Volcano plot illustration of transcriptional response to RSV infection when comparing RSV infected children to healthy controls. N=98 patients at the time of RSV infection compared with N=7 healthy control patients with no RSV infection. Red dots denote significantly differentially expressed probes (p <.05, fold change > 1.4), blue dots significantly differentially expressed probes with fold change < 1.4, green dots indicate genes with a fold change > 1.4 but which are not significantly differentially expressed, grey dots indicate probes that are not significantly differentially expressed. Horizontal dotted lines represent FDR cutoffs, vertical dotted lines represent log2 fold change cutoffs.



Sfigure 1.2.2 Volcano plot illustration of transcriptional response to RSV infection when comparing children in convalescence to healthy controls. N=98 patients during convalescence from an RSV infection compared with N=7 healthy control patients with no RSV infection. Red dots denote significantly differentially expressed probes (p <.05, fold change > 1.4), blue dots significantly differentially expressed probes with fold change < 1.4, green dots indicate genes with a fold change > 1.4 but which are not significantly differentially expressed, grey dots indicate probes that are not significantly differentially expressed. Horizontal dotted lines represent FDR cutoffs, vertical dotted lines represent log2 fold change cutoffs.

Appendix Chapter 2: Supplemental material for chapter 4

S2.1 Introduction

Studies aimed at understanding the relationship between the immune system with RSV disease severity have used different methods to classify patients as having more or less severe disease, such as severity scores (like ReSVinet(291) or GRSS(286)), hospitalisation(276), or requirement for respiratory support(275). RSV disease severity has also been stratified according to the respiratory support intervention required(278), by grading these interventions according to their invasiveness and categorising participants according to that grade. It was decided to perform exploratory analyses with metrics other than the primary metric used in this thesis to determine comparability between methods.

S2.2 Distributions of disease severity

The methods used to quantify disease severity in this study were ReSVinet score (a composite score consisting of many clinical variables), receipt of supplemental oxygen (and requirement of more invasive supplemental oxygen when compared to less) and requirement for hospitalisation. Different thresholds were used for severe disease within each of these metrics for the purpose of supervised statistical analyses of the relationship between disease severity and gene expression.

Using the ReSVinet score, a score of fourteen (as recommended by the developers of the score) was used as thresholds for severe disease(291). The median of the distribution of ReSVinet scores of participants in this study was nine. Forty-five participants were classed as having severe disease when using a score of nine as the threshold for severe disease, while only fourteen participants had severe disease when a score of fourteen was used as the threshold for severity (fig S2.1.1).

Requirement of hospitalisation for participants was also used as a metric by which disease severity could be determined. Using this metric 20 participants were classed as having mild disease, while 72 participants were classed as having severe disease (fig S2.1.2).

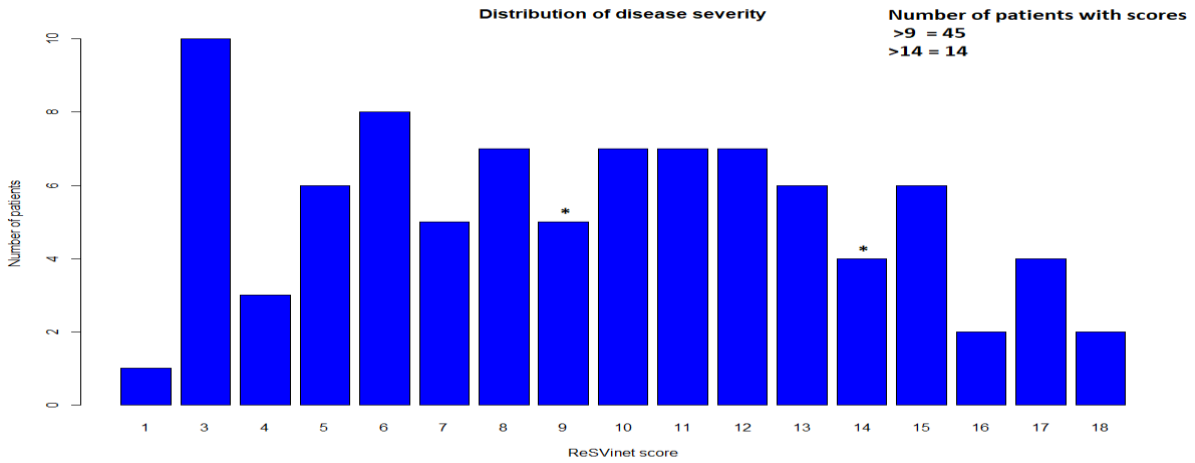


Figure S2.2.1 Distribution of disease severity scores in patients recruited to the RESCEU case control study. ReSVinet scores were plotted against number of patients with that score. * indicates thresholds for severe disease used in analyses, 9 and 14

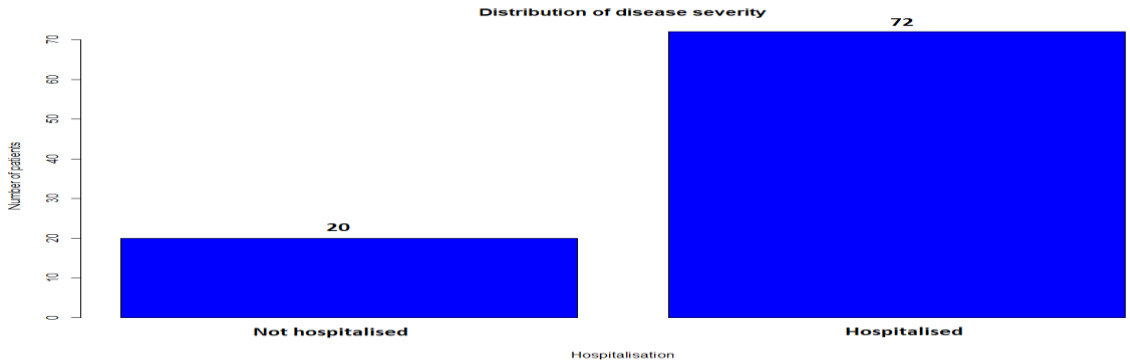


Figure S2.2.2 Number of patients recruited to the RESCEU case control study who were hospitalised and who were not hospitalised.

S2.3 Correlations of gene expression with clinical disease

S2.3.1 Gene expression correlates significantly with clinical disease

Expression of all 1707 differentially expressed during acute disease was used in correlation analyses with different outcomes representing disease severity from clinical data. These outcomes were the scoring of the participant according to the ReSVinet score, the requirement for more invasive respiratory support (graded from one to five) and the patients hospitalisation status. Severity scoring according to invasiveness of respiratory

support received was described in chapter 2.5. P values for each correlation were adjusted for multiple testing using the "FDR" method(369), and a significance threshold of .05 FDR adjusted p value was applied. The Pearson correlation values were used to rank each gene according to its correlation with a disease severity metric for gene set enrichment analyses. The top 30 genes that were most highly correlated with disease severity as measured by each metric are shown (table S2.3.1.1).

Severity metric	Genes correlated with more severe disease	Genes correlated with less severe disease	Total
ReSVinet score	180	47	227
Respiratory support	50	32	82
Invasive respiratory support	228	132	360
Hospitalisation	40	3	43

Table S2.3.1.1 Summary table of genes that have a statistically significant correlation (FDR <0.05) with disease severity as measured by ReSVinet score, requirement for respiratory support, a scoring of invasiveness of respiratory support, and requirement for hospitalisation. All significantly differentially expressed genes according to FDR p.value cutoff of .05 were used in correlation analyses.

S2.3.2 Majority of genes correlated with each severity metric are also correlated with others

To determine how similar the sets of genes that correlate with different metrics of disease severity were, a Venn diagram showing the intersection of all genes that significantly correlate with any of the metrics of disease severity used in this study was generated (fig S2.3.2.1). The largest overlap was between genes correlated with ReSVinet score and genes correlating with invasiveness of respiratory support, at 120 genes, which was 53% of the genes that significantly correlate with ReSVinet score and 33% of genes correlating with invasiveness of respiratory support. 27/227 (12%) genes correlating with ReSVinet score were uniquely correlated with ReSVinet score, with 143/360 (40%) uniquely correlated with invasiveness of respiratory support, 6/82 (7%) uniquely correlated with requirement for respiratory support and 1/43 (2%) correlating with requirement for hospitalisation.

Agreement plots for the shared group of significantly correlated genes for every possible pair of two severity metrics were generated (fig. S2.3.2.2). It was found that no gene positively correlated with disease severity by one metric and correlated with less severe disease by another.

With the exception of the six uniquely correlated genes with requirement of supplemental oxygen and the one significantly correlated gene with hospitalisation, all other genes that were significantly correlated with either of these terms were significantly correlated with either ReSVinet score or invasiveness of supplemental Oxygen. Based on these observed redundancies between genes correlated with requirement for supplemental oxygen and genes correlated with hospitalisation with other severity metrics, and the relatively low numbers of genes that correlate with these clinical outcomes, it was decided to remove these disease severity metrics from further exploratory analyses.

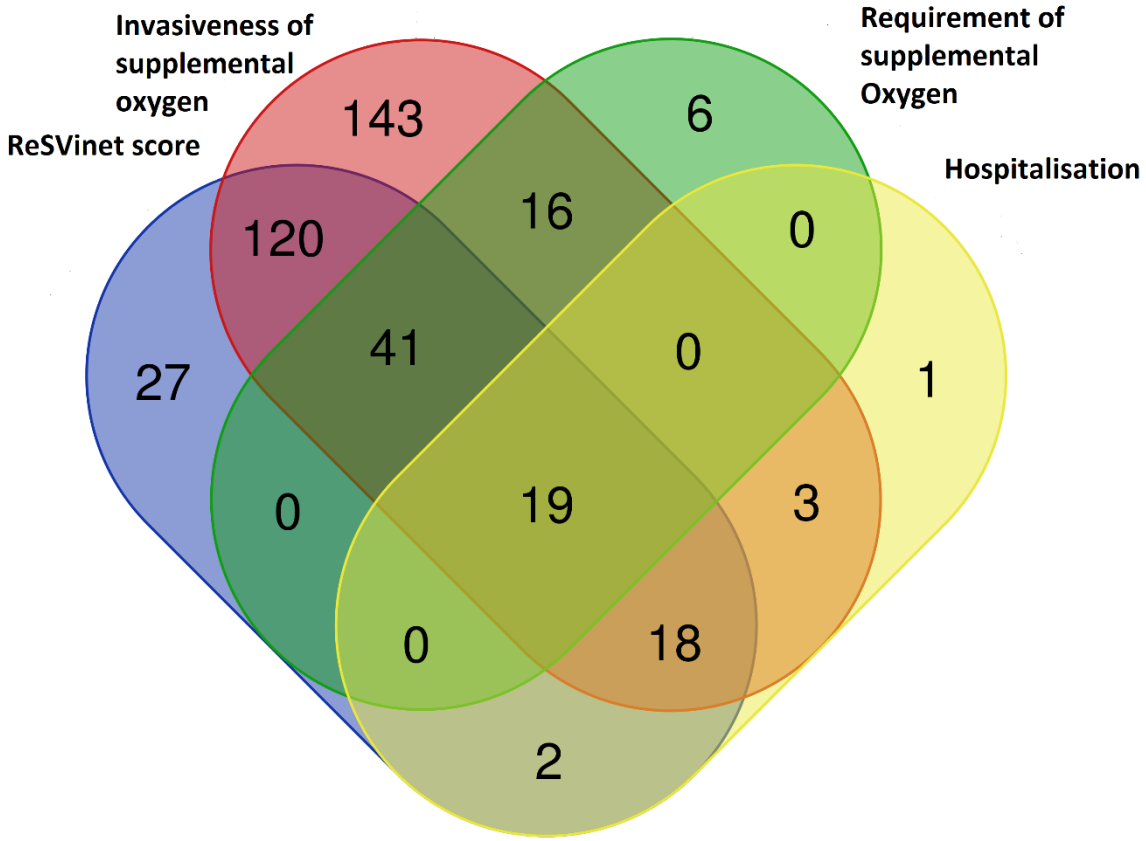


Figure S2.3.2.1 Venn diagram of genes that significantly correlate with disease severity as measured by ReSVinet score, invasiveness of supplemental oxygen treatment requirement for supplemental oxygen and requirement for hospitalisation. Intersections were numbered by shared genes between the intersecting categories

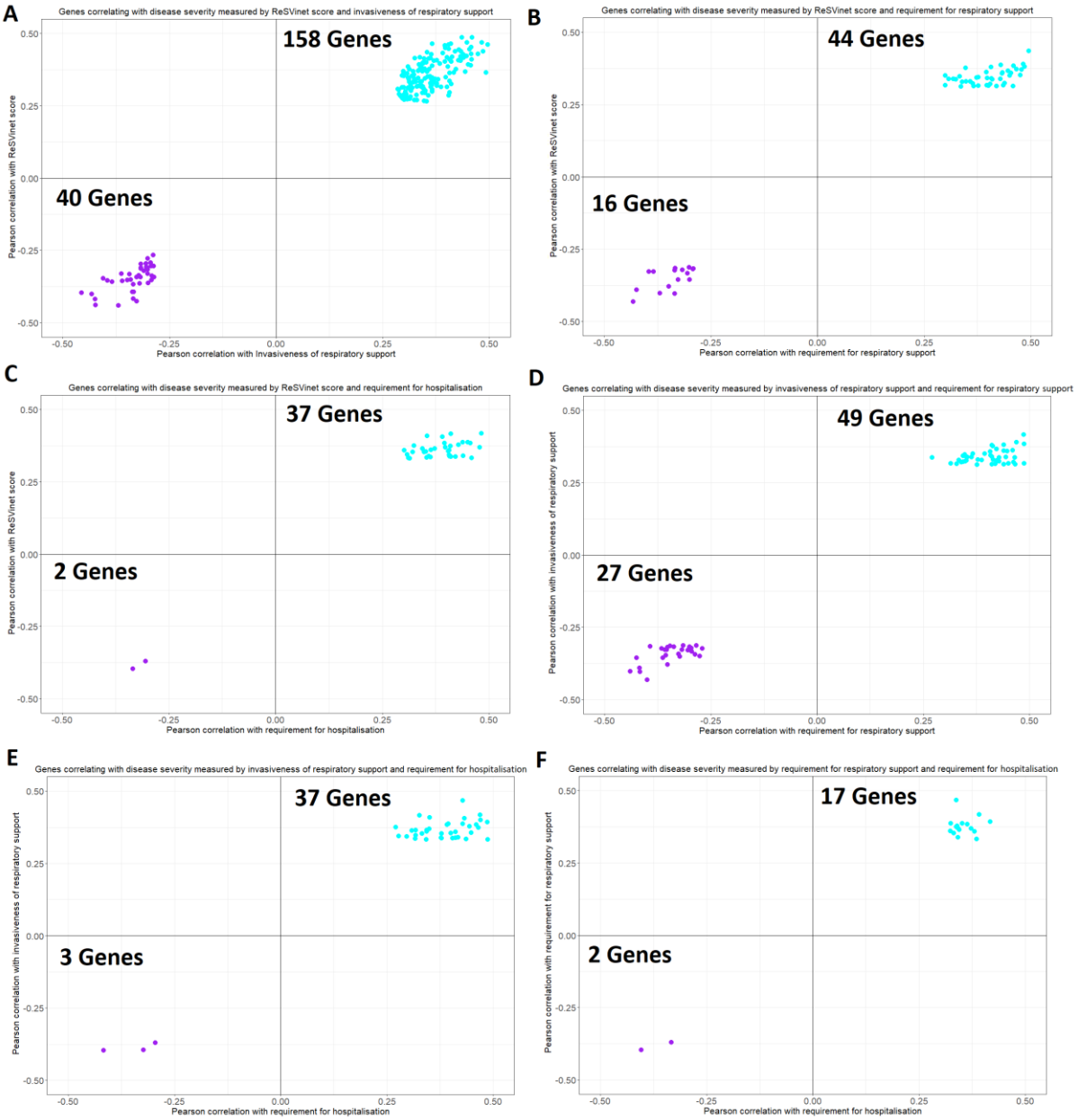


Figure S2.3.2.2 Agreement plots of genes with significant correlation with disease severity by some metric for A: genes commonly correlated with ReSVinet disease severity score and invasiveness of supplemental oxygen treatment B: with ReSVinet disease severity score and supplemental oxygen requirement, C: with ReSVinet disease severity score and hospitalisation requirement, D: invasiveness of supplemental oxygen treatment and requirement for supplemental oxygen, E: invasiveness of supplemental oxygen treatment and hospitalisation requirement and F: requirement for supplemental oxygen and requirement for hospitalisation. Genes were plotted by their correlations with different severity metrics, such that genes that were significantly positively correlated in both were plotted in the top right quadrant and coloured blue, and genes that were significantly correlated with less severe disease by both metrics of disease severity were plotted in the bottom left quadrant and coloured purple. Genes that were positively correlated with one metric and correlated with less severe disease in the other were plotted in the top left and bottom right quadrants.

S2.3.3 Neutrophil and Monocyte gene signatures correlate significantly with ReSVinet score

Ranked gene set enrichment analysis was performed on all 1707 genes found to be differentially expressed between acute disease and convalescence. The genes were ranked by their correlation with disease severity according to the ReSVinet score to determine the most common biological functions among genes that correlate with disease severity. By using the Pearson correlation coefficient value as the ranking value for gene set enrichment analysis, genes and processes that positively correlate with more severe disease and genes and processes that were associated with lower disease severity can be observed separately. Terms found to be positively enriched in the gene set provided can therefore be considered to be reflective of processes that are altered or differently regulated in severe disease, while terms found to be enriched in less severe are reflective of processes that are either protective against or negatively regulated during more severe disease.

A total of 227 genes were significantly correlated with RSV disease severity according to the ReSVinet score, of which 180 were significantly positively correlated and 47 were significantly correlated with less severe disease. The genes with the highest positive correlation with the ReSVinet severity score were *LIN7A*, *MGAM2* and *NSUN7*. The genes with the highest negative correlation with disease severity were *CRIP1*, *SLC30A4* and *ADGRG5* (fig S2.3.3.1).

The most enriched biological process terms for the correlations of genes with ReSVinet score were related to the humoral response to disease, reactive oxygen species, responses to bacteria and myeloid responses. The most highly enriched terms for the correlations of genes with lower ReSVinet score were related to B cells, expression of viral genes and Ribosomes (fig S2.3.3.2).

The most enriched cellular component terms were all related to neutrophil and granulocyte activity, such as specific granule lumen, tertiary granule membrane and tertiary granule, while the most highly enriched terms that were correlated with less severe disease were related to ribosomes (fig 2.3.3.3).

The most highly enriched terms that correlate positively with ReSVinet score disease severity were all relate to either hydrolase, lysase or binding of molecules, such as heparin binding. The most highly enriched terms in less severe disease also related to ribosomes, but with terms relating to transcriptional regulation and transcription (fig S2.3.3.4).

Immune enrichment terms show similarities between the ranked list of genes provided to the algorithm and experiments with similar highly upregulated genes. Immune terms positively correlating with ReSVinet score include terms related to a low ratio of lymphocytes to neutrophils, as well as *Streptococcus pneumoniae* related terms and terms related to viral infection. Immune terms correlating with lower ReSVinet score include an increased B cell phenotype relative to monocytes, and increased expression of lymphocyte related genes relative to NK cell related genes, as well as lower expression of neutrophils relative to B cells and lymphocytes (fig S2.3.3.5).

Positively correlated genes			Negatively correlated genes		
Pearson correlation	FDR	Gene	Pearson correlation	FDR	Gene
0.547475142	4.09E-05	LIN7A	-0.456357854	0.000573703	CRIP1
0.516499589	0.000163	MGAM2	-0.432466629	0.00115239	SLC30A4
0.507072133	0.000165	NSUN7	-0.425189397	0.001423788	ADGRG5
0.505251384	0.000165	SMPDL3A	-0.423280734	0.001517194	HLA.DMB
0.496156776	0.0002	MSRB3	-0.406134055	0.002480863	TSPYL1
0.495355483	0.0002	BPI	-0.39639259	0.003217304	YPEL1
0.492108291	0.000208	ROPN1L	-0.384534041	0.004958266	IL32
0.485565418	0.000267	OMG	-0.370062933	0.007982631	HLA.DPA1
0.482039208	0.000291	ARG1	-0.363112098	0.009075737	TRGV9
0.47778959	0.000333	OLFM4	-0.36120437	0.009270036	ID3
0.475017691	0.000354	PGLYRP1	-0.348984547	0.012882105	CD74
0.473192042	0.000359	CEACAM4	-0.344404721	0.014074881	ANXA2R
0.466664338	0.000475	TDRD9	-0.341173931	0.015193968	ATM
0.459251456	0.000574	S100A8	-0.338183576	0.015868196	PES1
0.458472418	0.000574	GCLM	-0.337613269	0.016066648	HLA.DRA
0.457402135	0.000574	MANSC1	-0.335455955	0.016700042	HLA.DQA1
0.457080325	0.000574	DHRS9	-0.335428934	0.016700042	FCMR
0.456033963	0.000574	GCA	-0.334668746	0.017027332	ZNF831
0.45340783	0.000626	RGL4	-0.329516299	0.019499559	ITGA4
0.450832155	0.000681	MNDA	-0.327618719	0.020530077	HLA.DPB1
0.447028468	0.000784	IL1R2	-0.32725153	0.020530077	CD27
0.446321451	0.000784	SLPI	-0.322628705	0.023253055	RARRES3
0.443361982	0.000874	MRVI1.AS1	-0.319935992	0.024804714	IKZF2
0.439682564	0.001005	TMEM55A	-0.318682073	0.025538004	CD2
0.438270496	0.001005	TLR8	-0.317171489	0.026117043	REXO4
0.438228908	0.001005	APAF1	-0.317017202	0.026117043	DBP
0.435466157	0.001111	CLEC5A	-0.31685642	0.026117043	EOMES
0.433883237	0.001152	PPM1N	-0.312693068	0.028904787	AES
0.432867312	0.001152	S100P	-0.310194097	0.030093114	PPP3CC
0.432005493	0.001152	FAR2	-0.305780866	0.033002018	PIK3C2B

Table S2.3.3.1 Table of the top 30 genes that were most highly associated with more and less severe disease according to the ReSVinet disease severity score. Pearson correlations between gene expression and disease severity were shown along with FDR adjusted p values for each correlation.

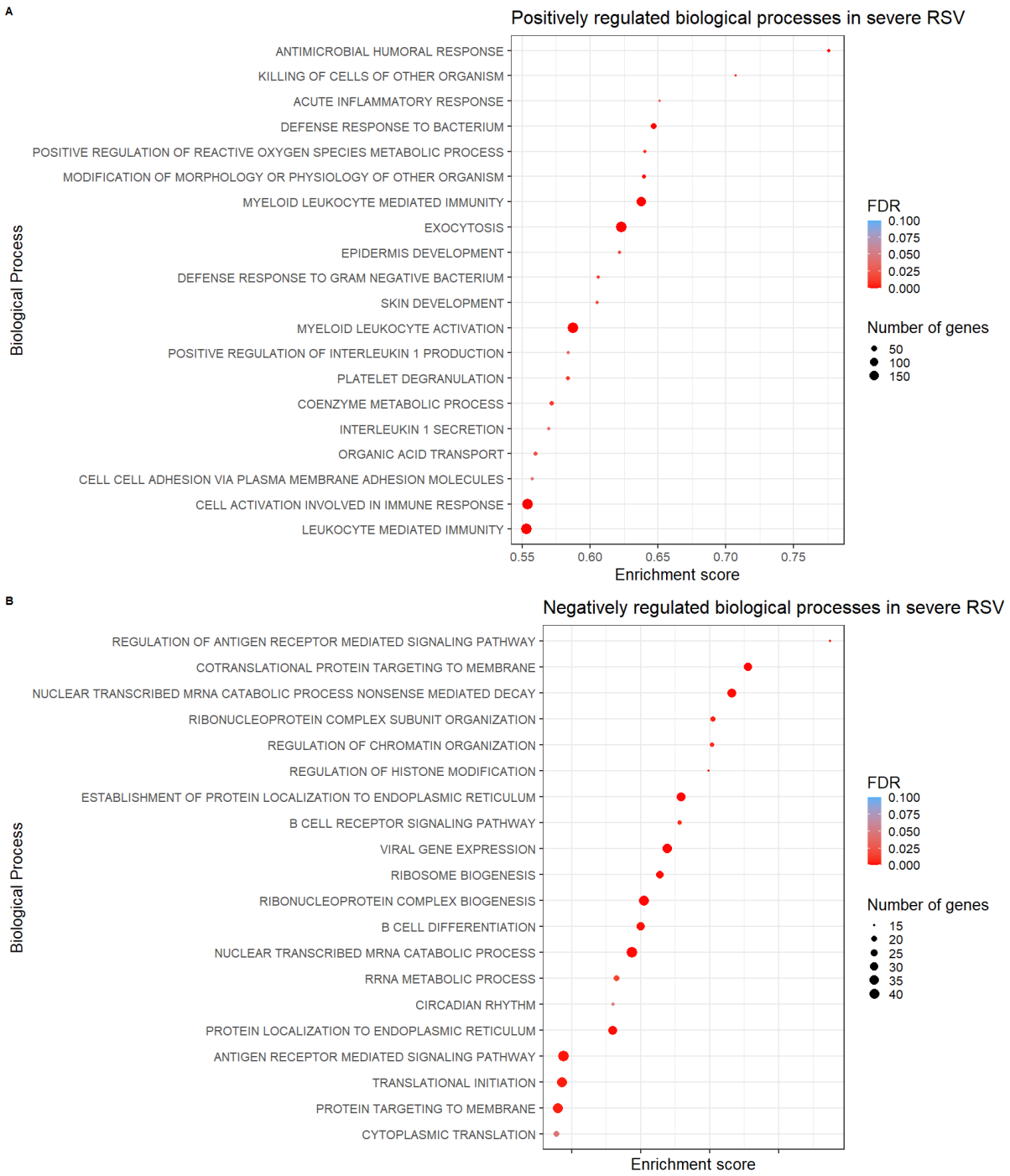


Figure S2.3.3.2 Gene set enrichment analysis of A – biological processes associated with more severe disease and B - biological processes associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with ReSVinet score. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene expression and disease severity measured by ReSVinet score. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

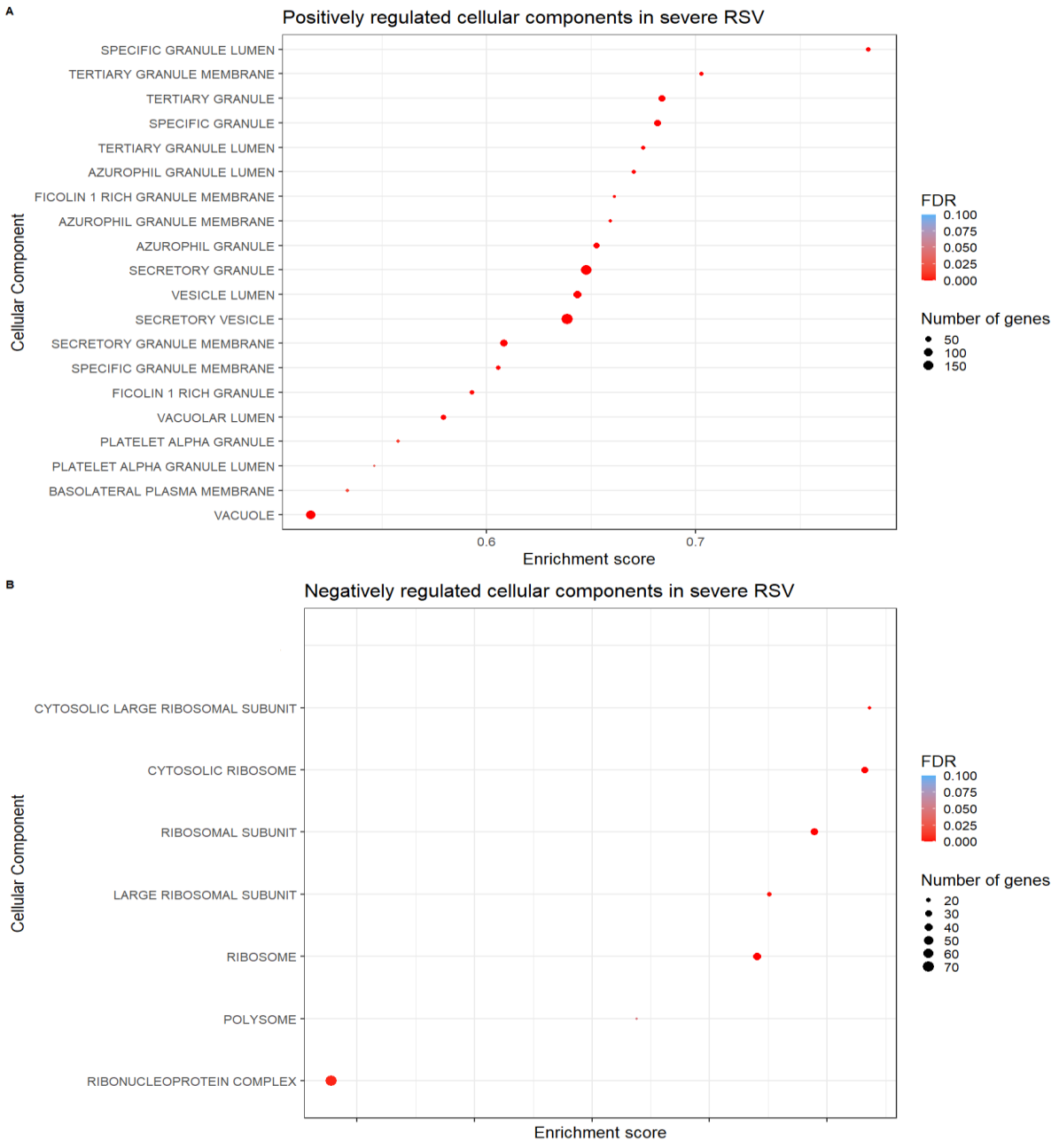


Figure S2.3.3.3 Gene set enrichment analysis of A – cellular components associated with more severe disease and B – cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with ReSVinet score. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene expression and disease severity measured by ReSVinet score. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

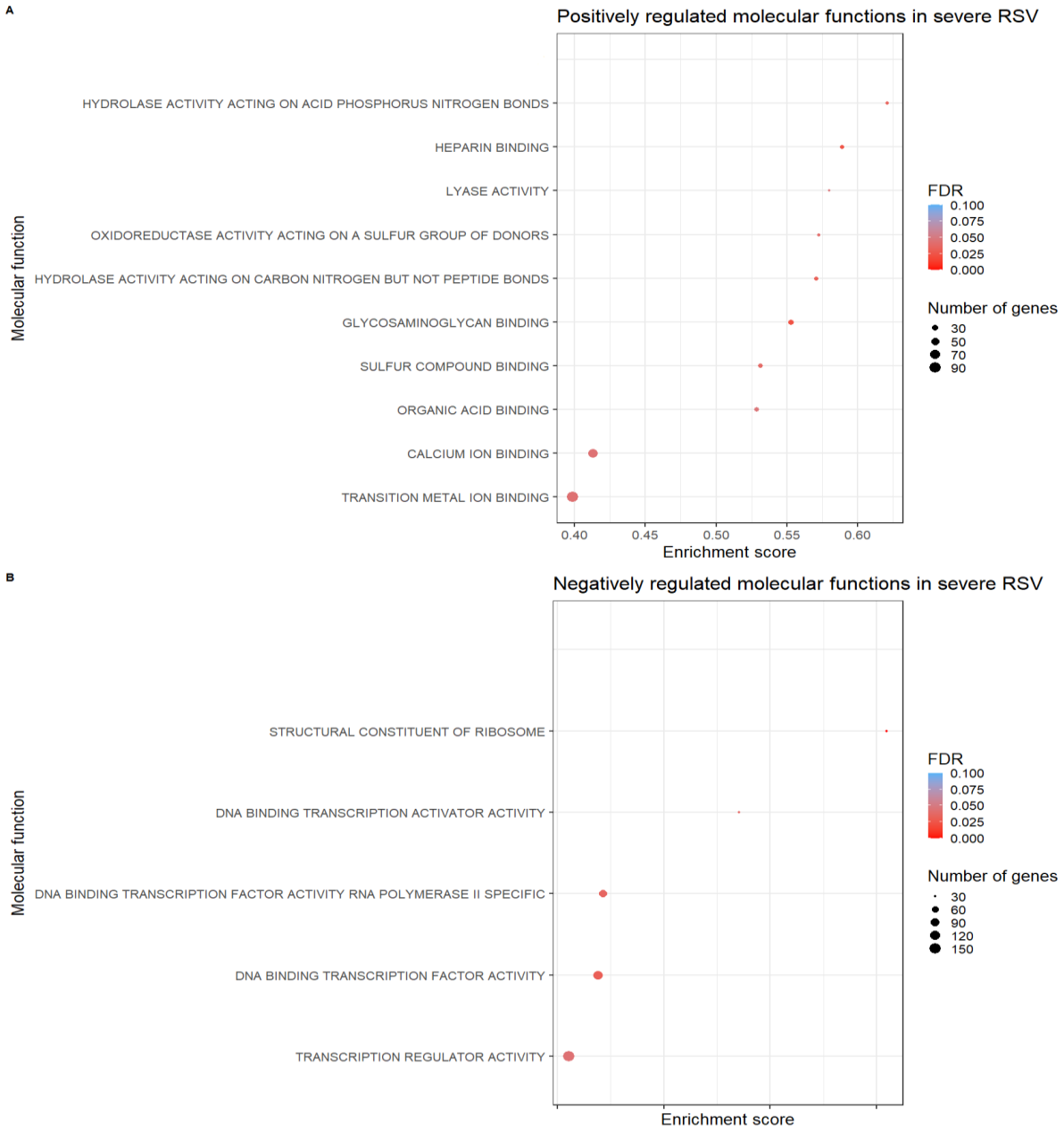


Figure S2.3.3.4 Gene set enrichment analysis of A – molecular functions associated with more severe disease and B – molecular functions associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with ReSVinet score. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene expression and disease severity measured by ReSVinet score. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

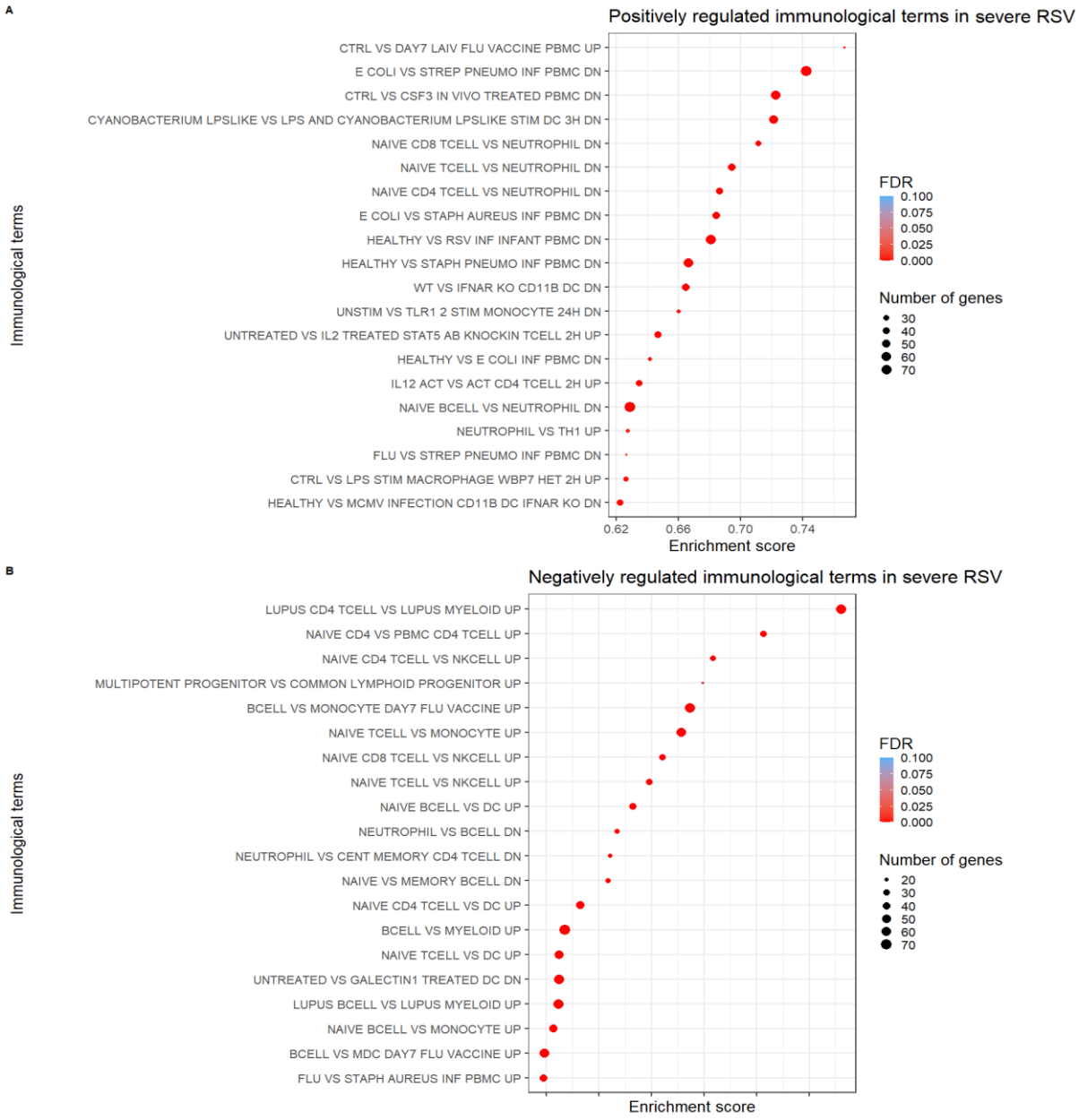


Figure S2.3.3.5 Gene set enrichment analysis of A – immunological terms associated with more severe disease and B – immunological terms associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with ReSVinet score. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene expression and disease severity measured by ReSVinet score. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

S2.3.4 Neutrophil and Monocyte gene signatures correlate significantly with requirement for supplemental Oxygen

A total of 82 genes were significantly correlated with RSV disease severity according to the requirement for supplemental oxygen, of which 50 were significantly positively correlated and 32 were significantly correlated with less severe disease. The genes with the highest correlation with the requirement for supplemental oxygen were *BPI*, *NSUN7* and *ARG1*. The genes with the highest negative correlation with disease severity were *SLC30A4*, *HLA.DQA1* and *HLA.DPA1* (fig S2.3.4.1).

The most highly enriched biological process terms for the correlations of genes with requirement for supplemental oxygen related to antibacterial responses, humoral immune responses, modification of foreign organisms and neutrophils, among general inflammatory responses, such as IL-1 production. The most highly enriched terms for the correlations of genes with less invasive requirement for supplemental oxygen were related to B cells, signalling downstream of antigen receptors, expression of viral genes and Ribosomes (fig S2.3.4.2).

The most highly enriched cellular component terms, similarly to the set of genes correlating with requirement for supplemental oxygen all related to neutrophil and granulocyte activity, such as specific granule lumen, tertiary granule membrane and tertiary granule, while the most highly enriched terms that were correlated with less severe disease were related to ribosomes (fig S2.3.4.3).

Only two molecular function terms were found to be significantly enriched in the set of genes positively correlated with more severe RSV disease. These terms were heparin binding and glycosaminoglycan binding. The terms that were significantly enriched in the set of genes that correlated with requirement for less invasive supplemental oxygen related to binding of transcription factors to DNA and to ribosomes (fig S2.3.4.4).

Immune terms that correlated with requirement for supplemental oxygen include terms related to a low ratio of B cells, NK cells $CD8^+$ and naive $CD4^+$ T cells to neutrophils, as well as *Streptococcus pneumoniae* related terms and terms related to viral infection. Immune terms that correlated with requirement for less invasive supplemental oxygen include an

increased B cell phenotype relative to monocytes and neutrophils, and increased expression of lymphocyte related genes relative to NK cell related genes, as well as lower expression of neutrophils relative to B cells and lymphocytes. A higher expression of memory $CD8^+$ T cell related terms to naïve $CD8^+$ T cells was also associated with decreased disease severity (fig S2.3.4.5).

Positively correlated genes			Negatively correlated genes		
Pearson correlation	FDR	Gene	Pearson correlation	FDR	Gene
0.435891231	0.013864	BPI	-0.431209851	0.013864	SLC30A4
0.41739981	0.018238	NSUN7	-0.403838217	0.023218	HLA.DQA1
0.391172772	0.026131	ARG1	-0.401377667	0.023218	HLA.DPA1
0.387035271	0.026131	DEFA4	-0.389932971	0.026131	ADGRG5
0.384702856	0.026131	S100A8	-0.378876566	0.026131	CD74
0.381477998	0.026131	OMG	-0.355615845	0.037571	HLA.DPB1
0.380354444	0.026131	SERPINB1(	-0.354564325	0.037571	EVL
0.376658881	0.026501	MS4A3	-0.350973008	0.037571	ODF2
0.373508972	0.028095	TDRD9	-0.348844146	0.039291	RFX5
0.371782111	0.028182	OLFM4	-0.346105368	0.03998	FUCA1
0.367538388	0.031487	SLPI	-0.343332638	0.03998	RORA
0.363342806	0.035165	UPP1	-0.341990223	0.03998	TMEM106
0.360880048	0.03587	PPM1N	-0.336617037	0.04013	ZKSCAN8
0.359844063	0.03587	RGL4	-0.333537082	0.043534	PIK3C2B
0.35880674	0.03587	S100P	-0.33013188	0.043534	PDE7B
0.351971339	0.037571	PGLYRP1	-0.328949926	0.043534	TPD52
0.351320372	0.037571	TMEM165	-0.327607779	0.043534	IL32
0.351131927	0.037571	MNDA	-0.327377366	0.043534	YPEL1
0.347626873	0.039709	RNASE2	-0.327003001	0.043534	SLAMF1
0.345438332	0.03998	PFKFB2	-0.324064386	0.045626	CCNI
0.3437707	0.03998	ERO1A	-0.323043537	0.045626	FCRL3
0.342618563	0.03998	CD58	-0.322562438	0.045626	FCMR
0.341387324	0.03998	NECAB1	-0.321891209	0.045626	EOMES
0.339839665	0.03998	HAUS4	-0.318981098	0.048136	FAM214A
0.339070596	0.03998	CLEC4D	-0.317184025	0.048136	MAPRE2
0.338998835	0.03998	ADAM9	-0.316879813	0.048136	PITPNC1
0.338147156	0.03998	MMP8	-0.316601863	0.048136	LINC00612
0.337717796	0.03998	LIN7A	-0.315176689	0.048136	ZNF831
0.337366662	0.03998	CHPT1	-0.314049933	0.048136	ANKRD36
0.330755687	0.043534	IL1R1	-0.313997557	0.048136	PLEKHA1

Table S2.3.4.1 Table of the top 30 genes that were most positively and negatively correlated with the requirement for supplemental oxygen of any kind. Pearson correlations between gene expression and disease severity were shown along with FDR adjusted p values for each correlation.

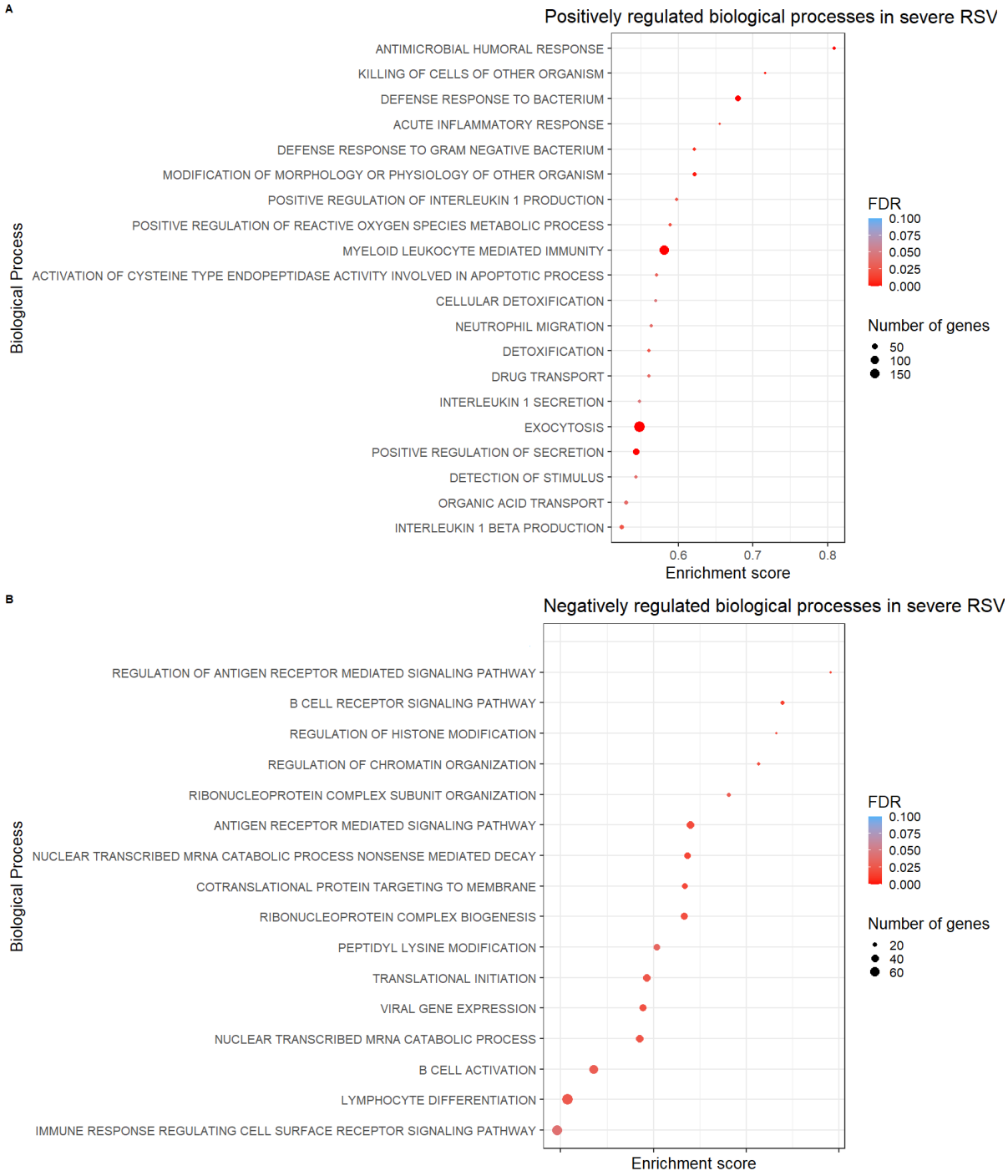


Figure S2.3.4.2 Gene set enrichment analysis of A – biological processes associated with more severe disease and B - biological processes associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with requirement for supplemental oxygen of any kind. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene expression and disease severity measured requirement for supplemental oxygen of any kind. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

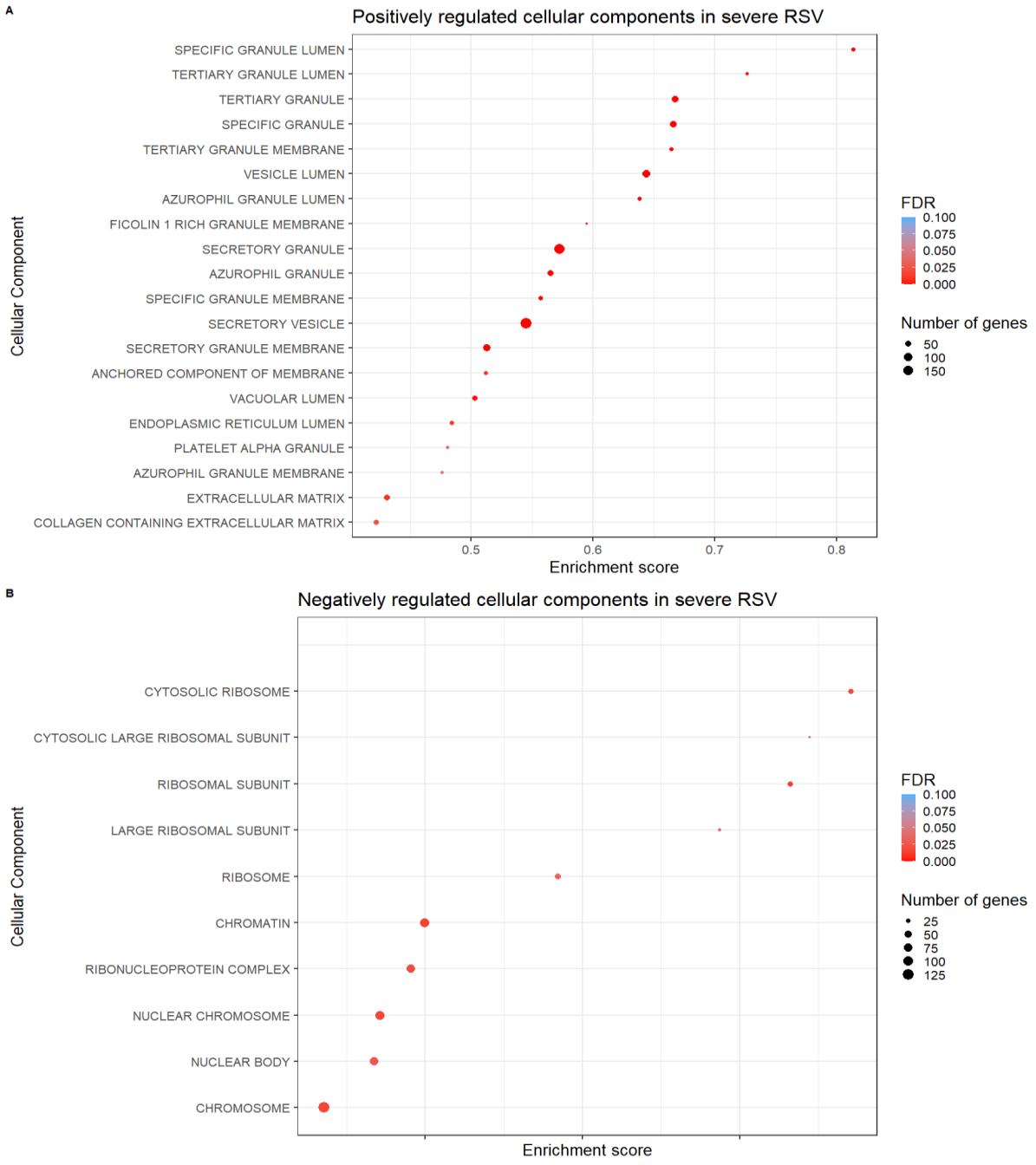


Figure S2.3.4.3 Gene set enrichment analysis of A – cellular components associated with more severe disease and B – cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with requirement for supplemental oxygen of any kind. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene expression and disease severity measured requirement for supplemental oxygen of any kind. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

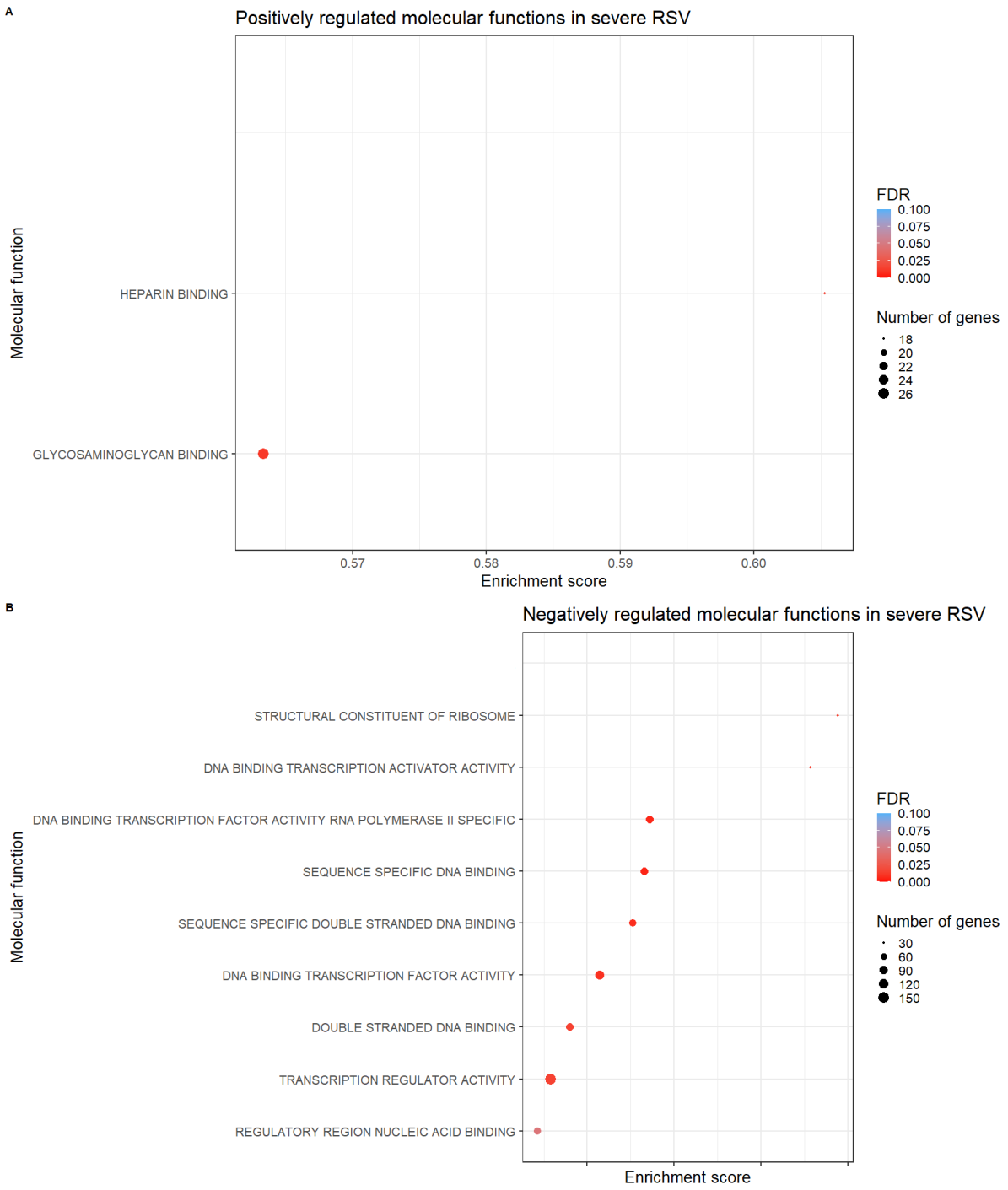


Figure S2.3.4.4 Gene set enrichment analysis of A – molecular functions associated with more severe disease and B – molecular functions associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with requirement for supplemental oxygen of any kind. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene expression and disease severity measured requirement for supplemental oxygen of any kind. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

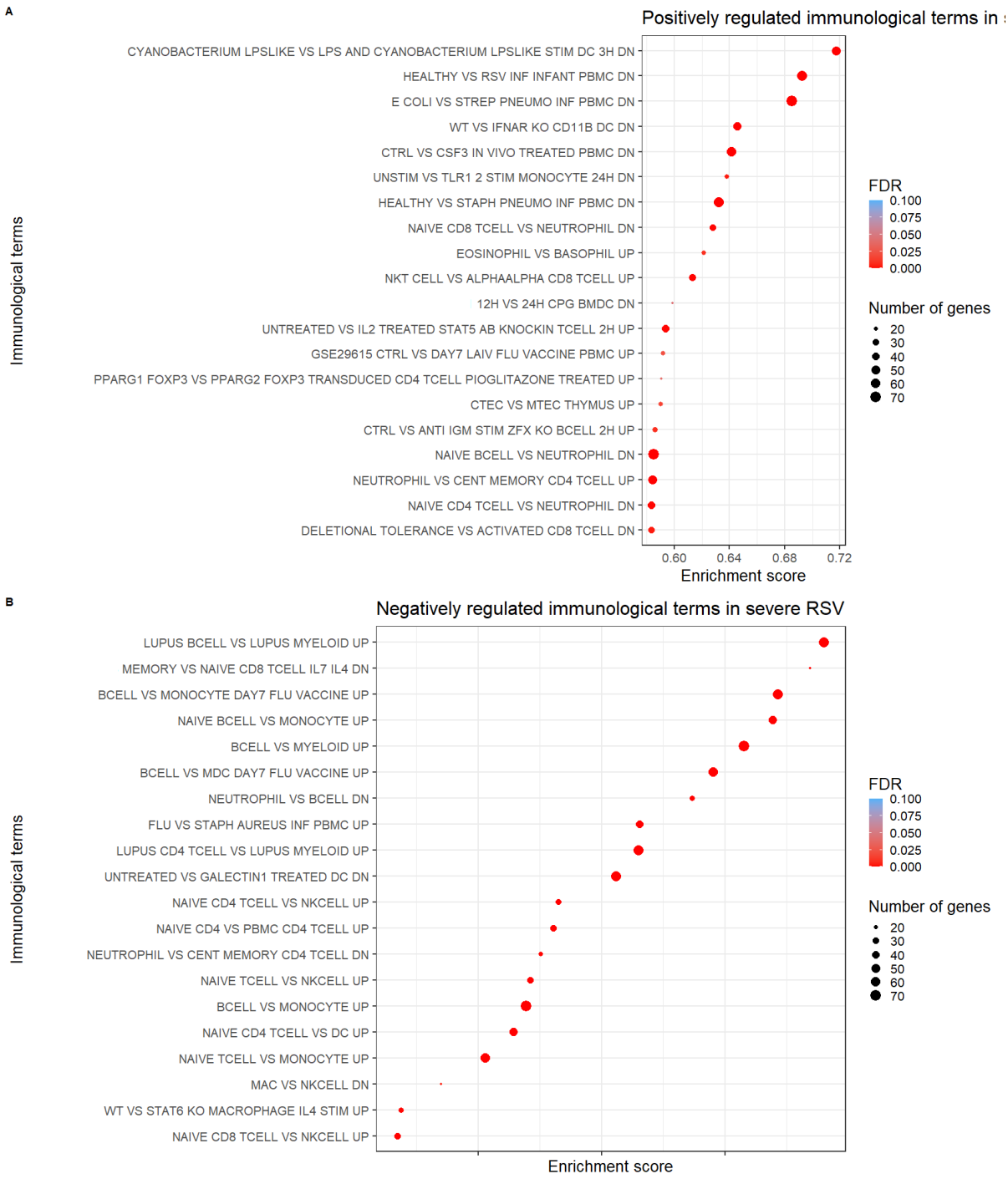


Figure S2.3.4.5 Gene set enrichment analysis of A – immunological tems associated with more severe disease and B – immunological terms associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with requirement for supplemental oxygen of any kind. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene expression and disease severity measured requirement for supplemental oxygen of any kind. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

S2.3.5 Neutrophil and Monocyte gene signature correlates significantly with requirement for hospitalisation

A total of 43 genes were significantly correlated with RSV disease severity defined according to the requirement for hospitalisation, of which 40 were significantly positively correlated and 3 were significantly correlated with no requirement for hospitalisation. This was the smallest number of genes found to be significantly correlated with any disease severity metric used in this study. The genes with the highest correlation with the requirement for hospitalisation were *LIN7A*, *NOG* and *PIK3C2B*. The genes with the highest correlation with less severe disease were *HLA.DQA1*, *HLA.DMB* and *HLA.DPA1* (fig S2.3.5.1).

The most highly enriched biological process terms for the correlations of genes with requirement for hospitalisation were related to reactive oxygen species, IL-1 production and secretion, general inflammatory and myeloid responses. The most highly enriched terms for the correlations of genes with no requirement for hospitalisation were related to B cells, signalling downstream of antigen receptors, expression of viral genes and Ribosomes (fig S2.3.5.2).

The most highly enriched cellular component terms, similarly to other severity metrics were all related to neutrophil and granulocyte activity, such as specific granule lumen, tertiary granule membrane and tertiary granule, while the most highly enriched terms that were correlated with less severe disease were related to ribosomes (fig S2.3.5.3).

No molecular function terms were found to be significantly enriched in the set of genes correlated with more severe RSV disease as measured by requirement for hospitalisation.

The most highly enriched immunological term that positively correlated with requirement for hospitalisation was related to a low ratio of naïve $CD8^+$ T cells to neutrophils. Other immune terms that positively correlated with requirement for hospitalisation required include terms related to a low ratio of B cells, NK cells $CD8^+$ and naive $CD4^+$ T cells to neutrophils, as well as *Streptococcus pneumoniae* related terms and terms related to viral infection. Immune terms correlating with no requirement for hospitalisation included an increased B cell phenotype relative to monocytes and neutrophils, and increased expression

of lymphocyte related genes relative to NK cell related genes, as well as lower expression of neutrophils relative to B cells and lymphocytes (fig S2.3.5.4).

Positively correlated genes			Negatively correlated genes		
Pearson correlation	FDR	Gene	Pearson correlation	FDR	Gene
0.468033081	0.00389	LIN7A	-0.396107487	0.018655	HLA.DQA1
0.418977724	0.018162	ARG1	-0.395381807	0.018655	NOG
0.416462688	0.018162	SULT1B1	-0.369955726	0.024521	PIK3C2B
0.409384092	0.018162	FAM151B			
0.406698459	0.018162	FAM160B1			
0.40061131	0.018655	MGAM2			
0.393209464	0.018655	NSUN7			
0.387434329	0.018871	MNDA			
0.387413365	0.018871	TLR8			
0.384792339	0.018871	GCA			
0.384587613	0.018871	UPP1			
0.379337554	0.021962	ADAM9			
0.376322067	0.023167	CHPT1			
0.37492348	0.023167	MMP8			
0.37048984	0.024521	CD58			
0.369437106	0.024521	OLFM4			
0.366264777	0.025633	COL4A3BP			
0.365920788	0.025633	ERO1A			
0.364683741	0.025733	CD163			
0.360825091	0.027756	ROCK1			
0.360202129	0.027756	SERPINB10			
0.359640071	0.027756	FAR1			
0.356890512	0.029778	RAB27A			
0.354977922	0.02992	IRAK3			
0.354125469	0.02992	IL1R1			
0.354004016	0.02992	PSTPIP2			
0.347960098	0.036589	CYP1B1			
0.344500707	0.040425	RIT1			
0.342890708	0.041632	DNAJC3			
0.340360131	0.043285	TMEM55A			

Table S2.3.5.1 Table of the top 30 genes that were most positively and negatively correlated with the requirement for hospitalisation of participants. Pearson correlations between gene expression and disease severity were shown along with FDR adjusted p values for each correlation.

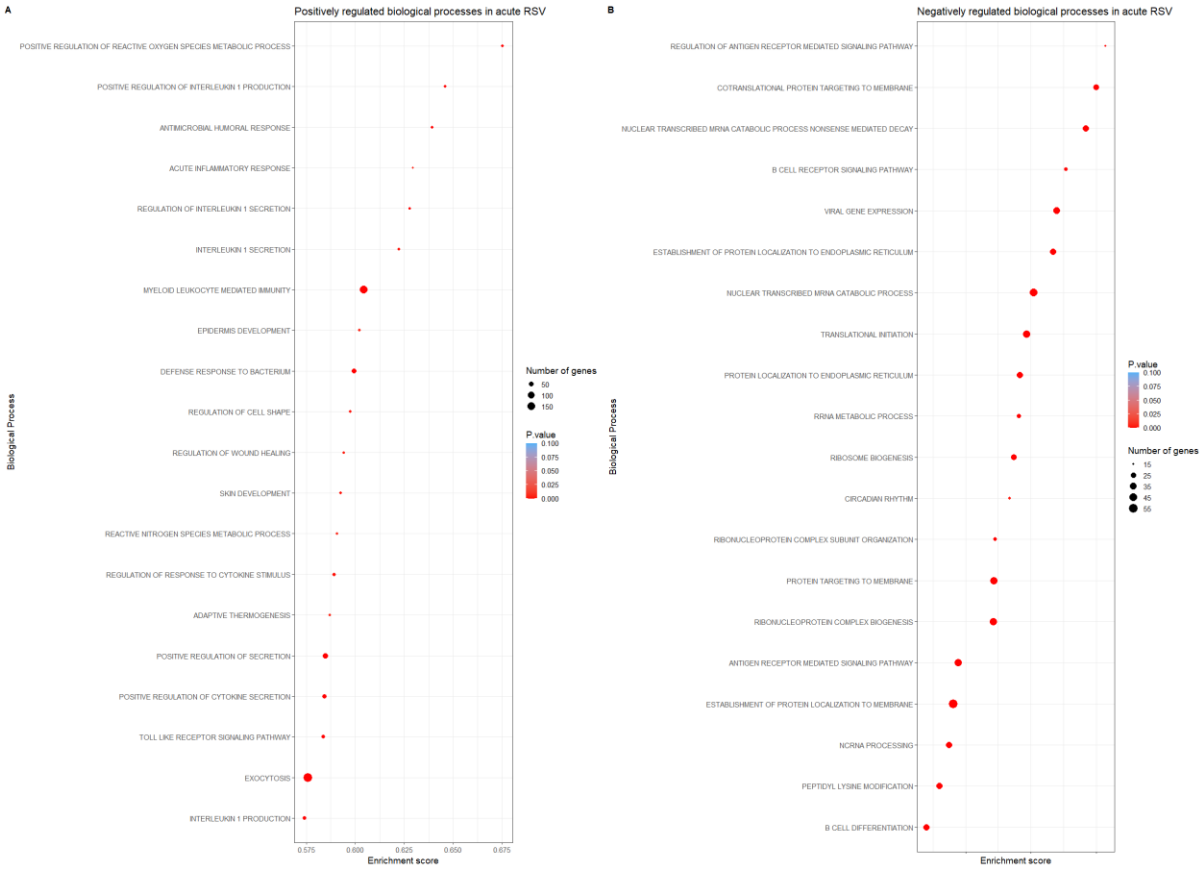


Figure S2.3.5.2 Gene set enrichment analysis of A – biological processes associated with more severe disease and B - biological processes associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with requirement for hospitalisation. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

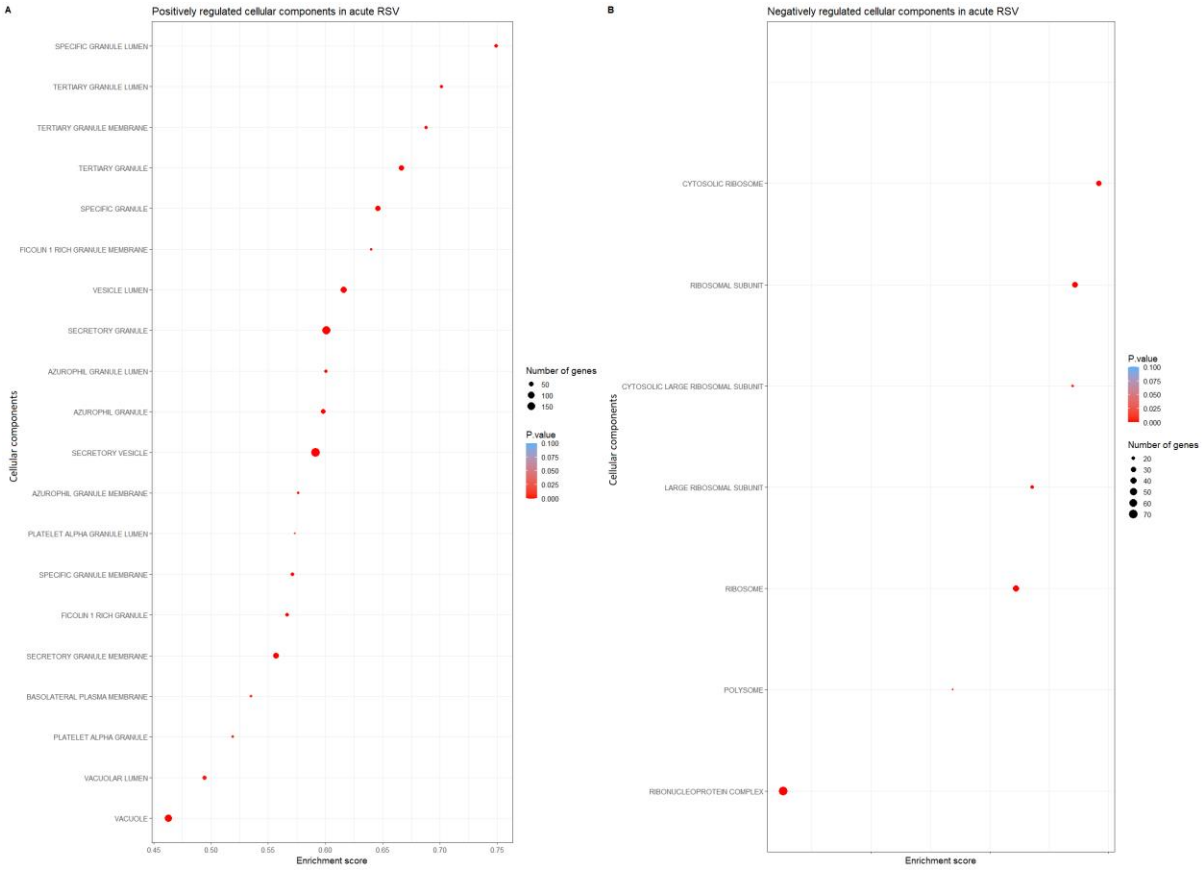


Figure S2.3.5.3 Gene set enrichment analysis of A – cellular components associated with more severe disease and B – cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with requirement for hospitalisation. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

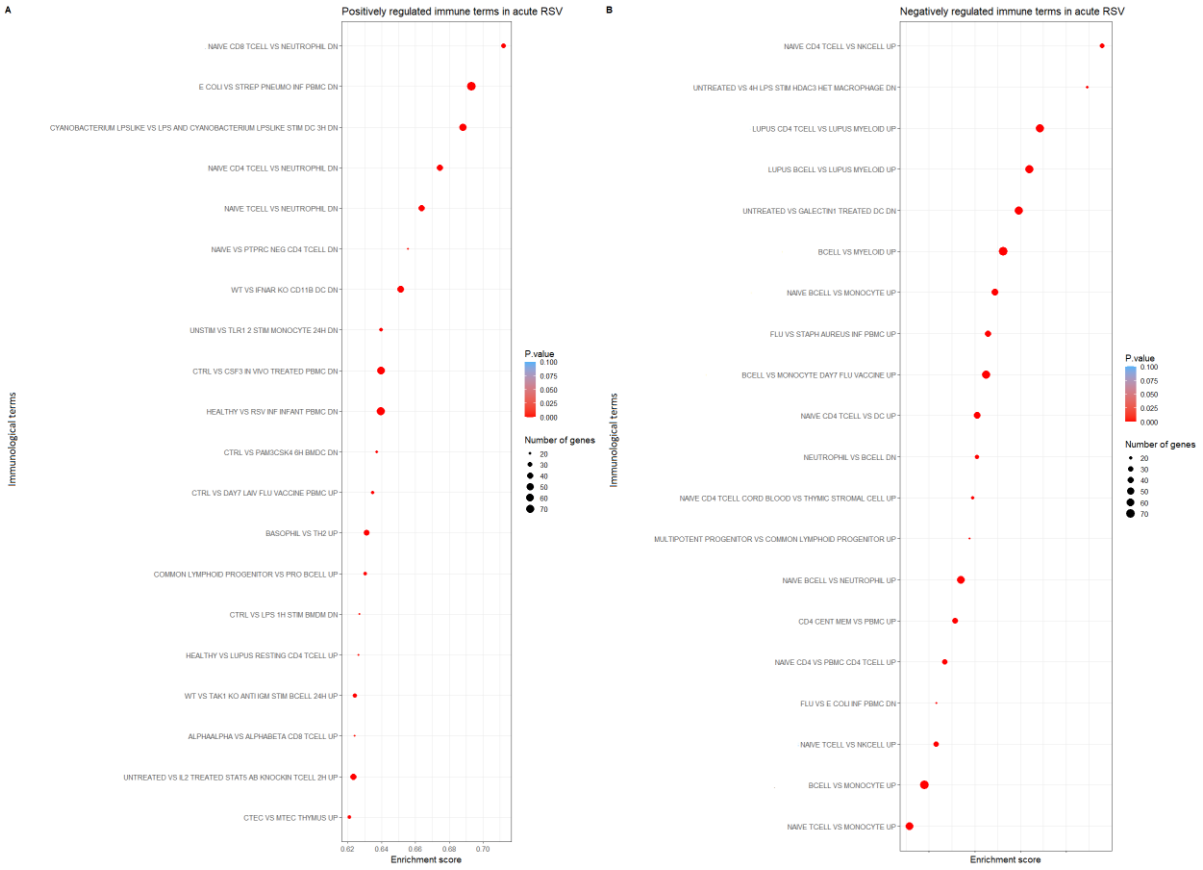


Figure S2.3.5.4 Gene set enrichment analysis of A – molecular functions associated with more severe disease and B – molecular functions associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by correlation with requirement for hospitalisation. Terms were ranked according to enrichment score for the term calculated based on correlation values between the gene and requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

S2.4 Patterns of gene expression are distinct between participants with severe and mild disease

To determine patterns of gene expression that associate most highly with disease severity, PLS-DA was used on the set of 1707 differentially expressed genes found in acute RSV infection. As PLS-DA finds the set of variables that were maximally different between different groups, severity metrics had to be defined prior to application of the algorithm. A ReSVinet score of 14 was used as a threshold for severe disease in these analyses, placing 14 (15%) of the participants in the severe category and 79 in the mild category. Maps of groupings of participants according to PLS-DA groupings were generated, and functional analysis was performed on genes contributing to the first two components of each PLS-DA.

S2.4.1 Machine learning algorithms can robustly separate mild and severe disease as categorised by ReSVinet score

PLS-DA was performed on the set of 1707 genes that were differentially expressed during RSV infection. PLS-DA reduced this set of genes to two components accounting for 13% of the variation in the data set and 8% of the variation in the data set respectively on which two distinct clusters corresponding to mild and severe disease were observable (Fig S2.4.1.1).

A correlation circle plot (fig S2.4.1.2) and a plot of the loadings of all genes contributing to the two components (fig S2.4.1.3A, S2.4.1.3B) were generated to visualise the contribution of the most important genes to these components. The genes that contributed the most to the first PLS-DA component were *MSRB3*, *IL1R1* and *GLIPR2*. The genes that contributed most to the second PLS-DA component were *PKIG*, *C4orf3* and *VSIG1*.

Functional analysis was then performed on the genes contributing to each PLS-DA component using the Pre-ranked GSEA algorithm. Loadings of genes on each individual component were used as ranks for this analysis.

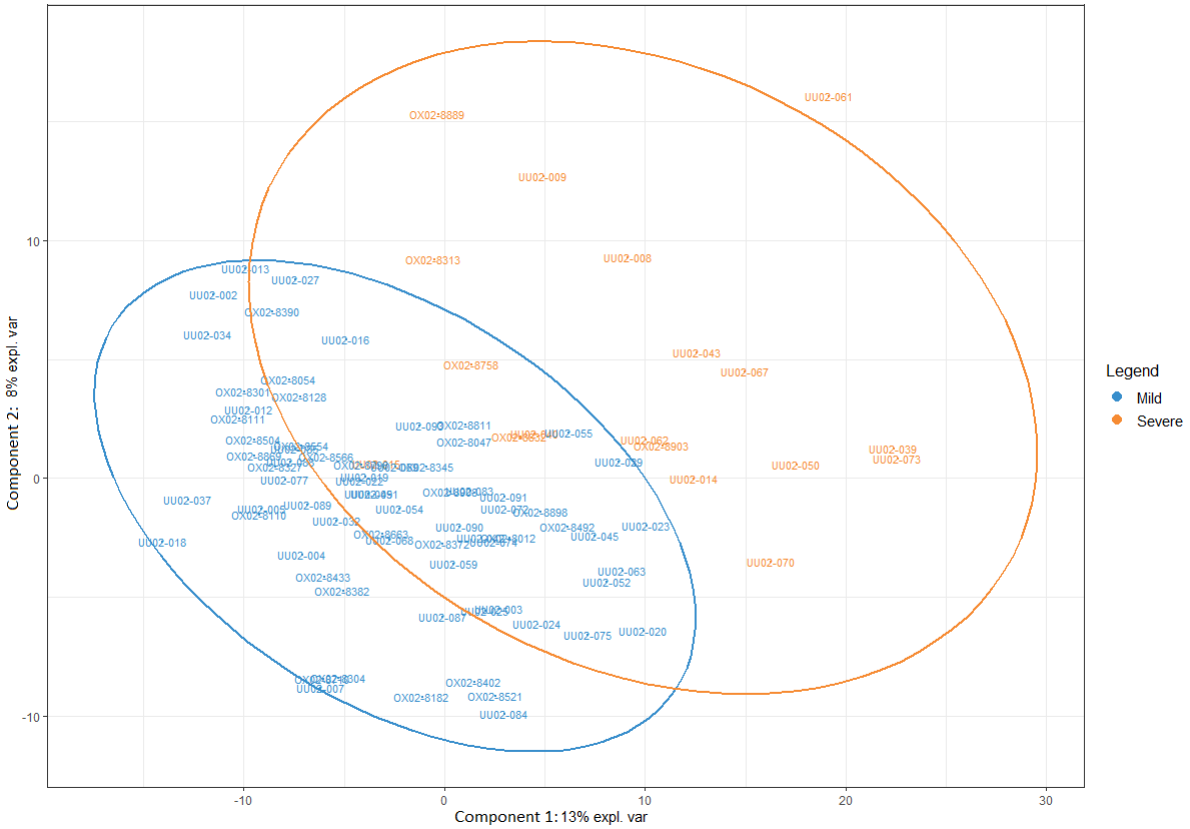


Figure S2.4.1.1 Individual mappings on the first two components of pls-da analysis of disease severity defined by RESViNET score. Participants coloured in orange have a ReSVinet score above 14, and were considered to have severe disease, while those coloured in blue were considered to have mild RSV disease. X and Y axes were labelled with the amount of the total variation in the differentially expressed genes utilised by the algorithm that was used to separate the samples along that axis.

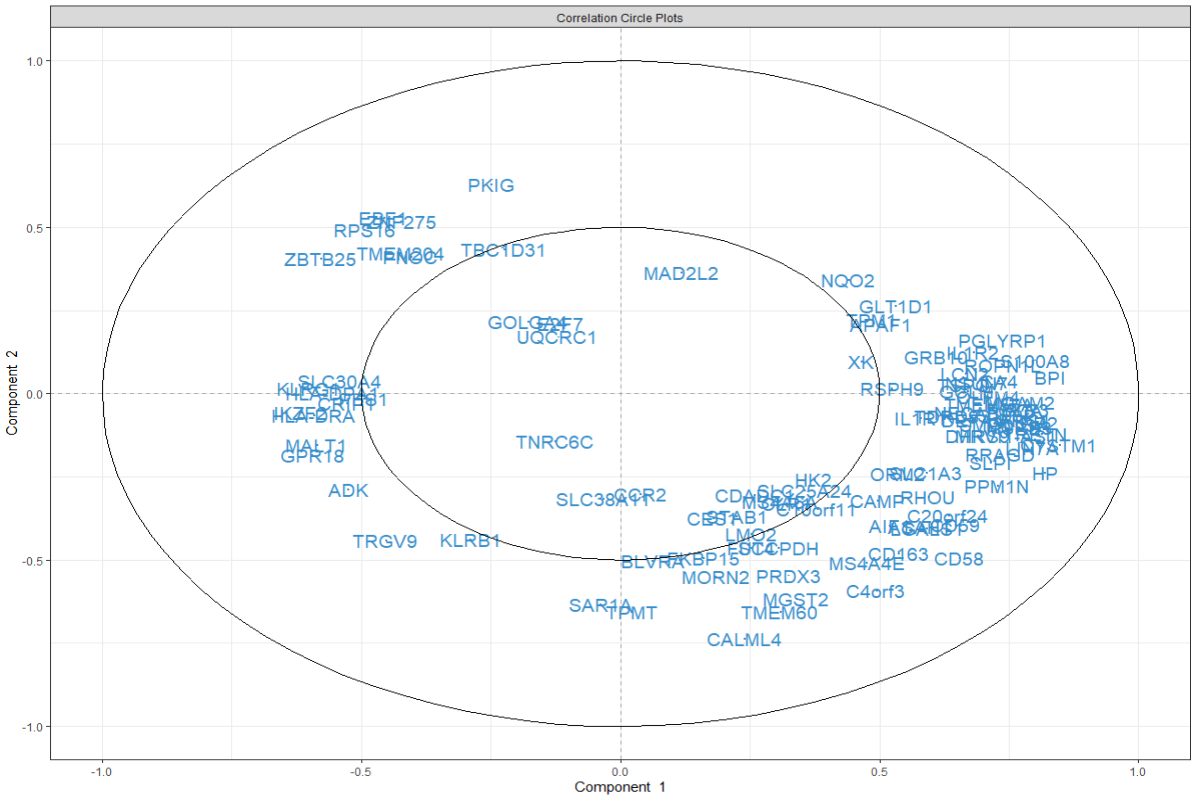


Figure S2.4.1.2 Variable mappings on the first two components of pls-da analysis defined by RESViNET score. Genes were plotted according to extent to which their expression contributes to plotting of participants in the ordinal direction in which they were plotted.

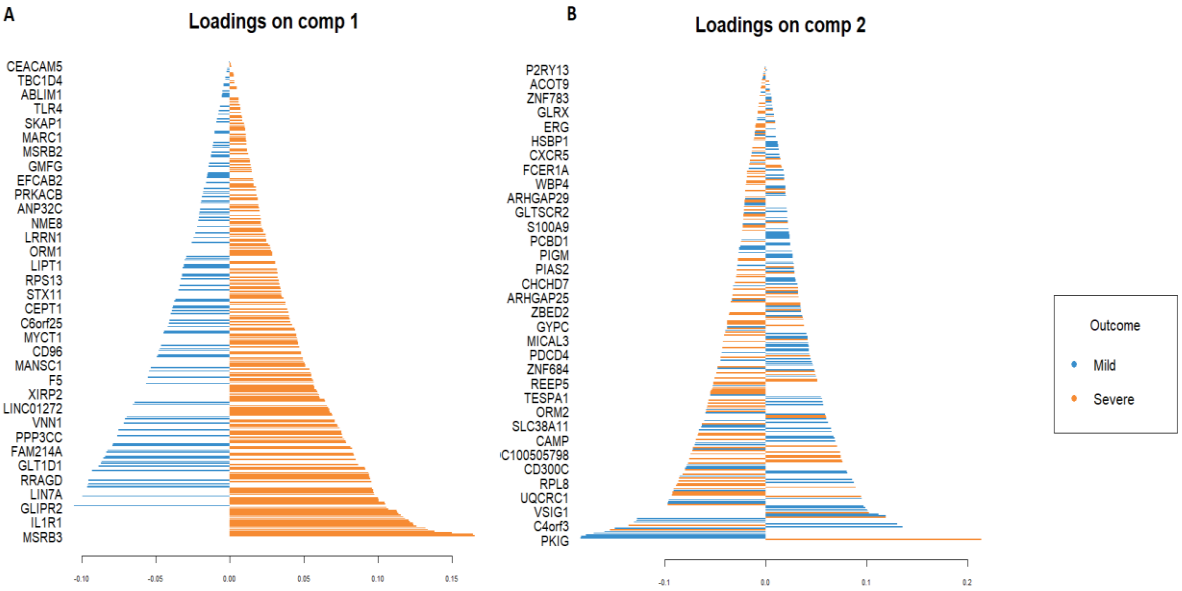


Figure S2.4.1.3 The loadings of genes on A: the first component and B: the second component of PLS-DA plots defined by RESViNET score. Loadings indicate the relative contribution of genes in the development of the PLS-DA model used to maximise distance between groups. Gene loadings were coloured according to whether increased expression of that associates with grouping with participants with severe disease (orange) or mild disease (blue).

S2.4.2 Machine learning algorithms can robustly separate mild and severe disease as categorised by ReSVinet score

PLS-DA was performed on the set of 1707 genes that were differentially expressed during RSV infection. PLS-DA reduced this set of genes to two components accounting for 12% of the variation in the data set and 4% of the variation in the data set respectively on which two distinct clusters corresponding to mild and severe disease were observable (Fig S2.4.2.1).

A correlation circle plot (fig S2.4.2.2) and a plot of the loadings of all genes contributing to the two components (fig S2.4.2.3 A, S2.4.2.3 B) were generated to visualise the contribution of the most important genes to these components. The genes that contributed the most to the first PLS-DA component were *PGLYRP1*, *CLEC5A* and *SMPDL3A*. The genes that contributed most to the second PLS-DA component were *CENPF*, *SDSL* and *LINC01521*.

Functional analysis was then performed on the genes contributing to each PLS-DA component using the Pre-ranked GSEA algorithm. Loadings of genes on each individual component were used as ranks for this analysis.

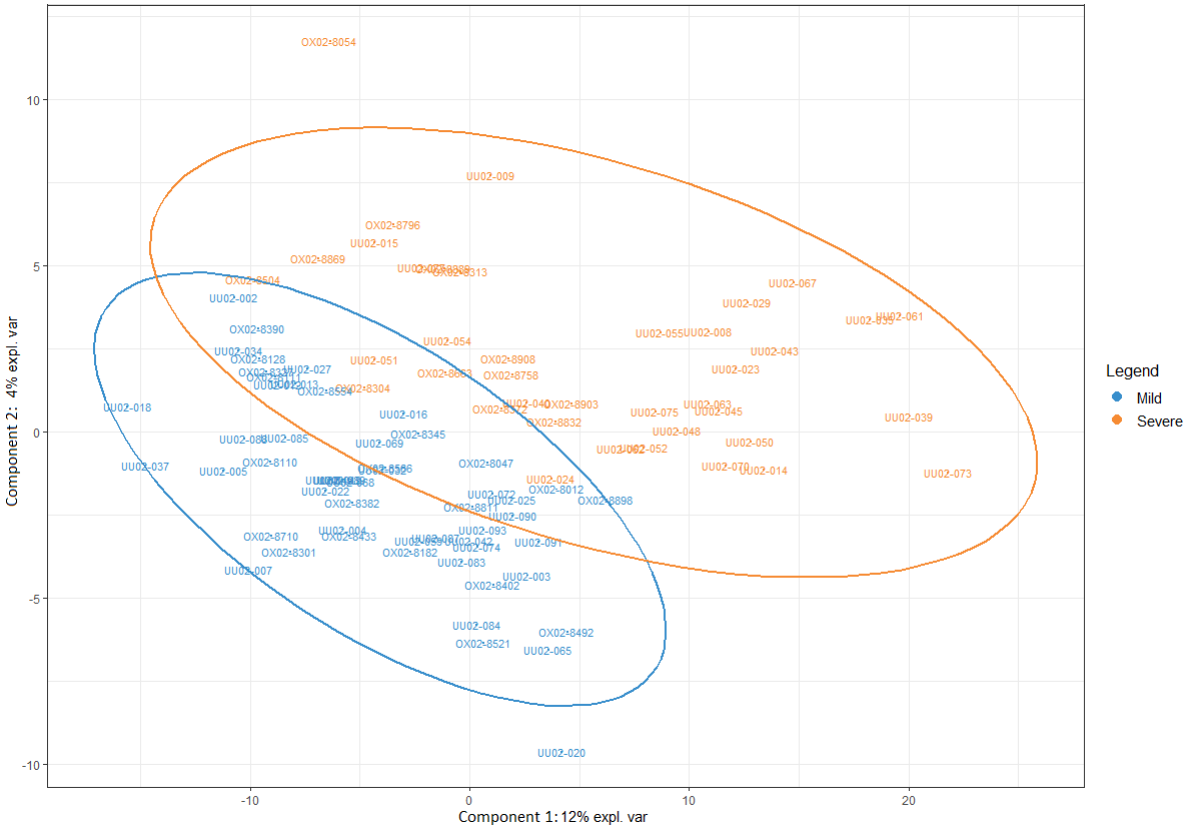


Figure S2.4.2.1 Individual mappings on the first two components of pls-da analysis according to RESViNET score. Participants coloured in orange have a ReSVinet score above 14, and were considered to have severe disease, while those coloured in blue were considered to have mild RSV disease. X and Y axes were labelled with the amount of the total variation in the differentially expressed genes utilised by the algorithm that was used to separate the samples along that axis.



Figure S2.4.2.2 Variable mappings on the first two components of pls-da analysis according to RESViNET score. Genes were plotted according to extent to which their expression contributes to plotting of participants in the ordinal direction in which they were plotted.

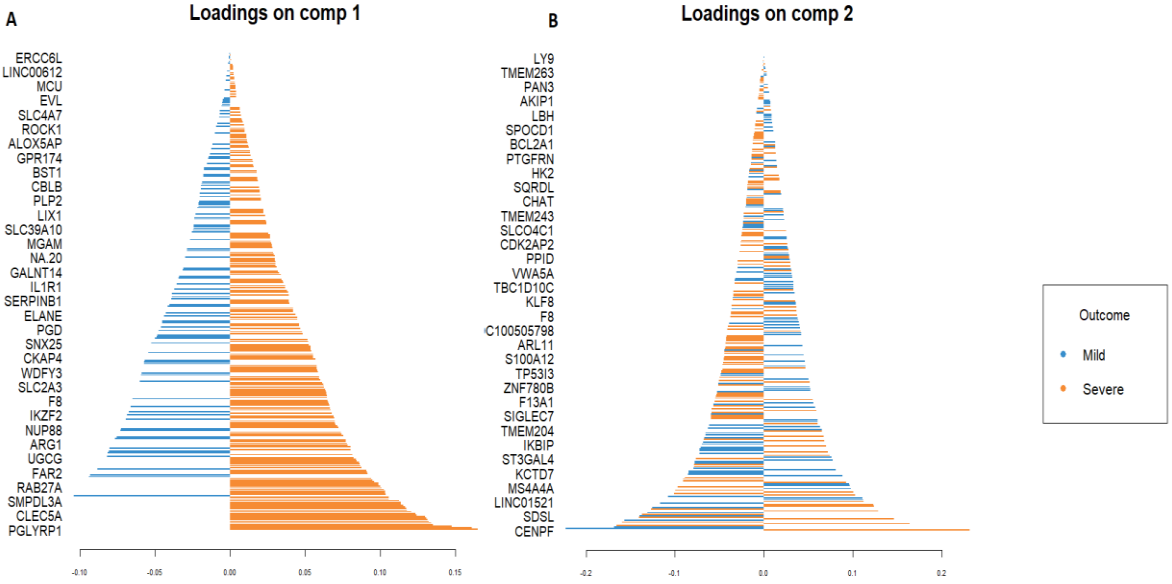


Figure S2.4.2.3 The loadings of genes on A: the first component and B: the second component of PLS-DA plots according to RESViNET score. Loadings indicate the relative contribution of genes in the development of the PLS-DA model used to maximise distance between groups. Gene loadings were coloured according to whether increased expression of that associates with grouping with participants with severe disease (orange) or mild disease (blue).

S2.4.3 First and second components of PLS-DA show distinct patterns of gene expression that correlate with ReSVinet score

The genes set of 1707 differentially expressed genes used in the PLS-DA analysis described in 4.5.1 were ranked according to their loadings on the first and second principal components for gene set enrichment analysis. Each component represents a grouping of highly co-correlated genes that do not correlate with any other components and that were maximally different between the two groups of participants. An FDR adjusted p.value cut-off for significance of .05 was used as a threshold for significance in these analyses.

S2.4.3.1 Genes on first component of ReSVinet score PLS-DA associate with upregulated neutrophils and monocytes, and downregulated lymphocytes

The loadings of all 1707 differentially expressed genes used in PLS-DA analysis were used to rank genes for gene set enrichment analyses. The first component identified by PLS-DA using ReSVinet score as the determinant of severity comprised of 13% of the variation in the set of differentially expressed genes.

The most highly positively enriched biological processes (Fig S2.4.3.1.1 A) were to humoral immune responses, cellular detoxification, regulation of wound healing, antibacterial responses and IL-1 responses, while the most highly enriched biological processes in less severe disease (Fig S2.4.3.1.1 B) were related to ribosomes, B cell differentiation and antigen receptor signalling.

The most highly enriched cellular components in more severe disease (Fig S2.4.3.1.2 A) all related to neutrophil granules and membranes, while the most highly enriched cellular components in less severe disease (Fig S2.4.3.1.2 B) were all related to ribosomes.

The most highly enriched molecular functions in more severe disease (Fig S2.4.3.1.3 A) were hydrolase binding, heparin binding and glycosaminoglycan binding, while the only a single molecular function was significantly enriched that had a negative association with disease severity (Fig S2.4.3.1.3 B), which was structural constituent of ribosome.

The most highly enriched immunological terms in more severe disease (Fig S2.4.3.1.4 A) were related to similarity with *Streptococcus pneumoniae* related signals, a low ratio of lymphocytes to neutrophils and a high ratio of eosinophils to basophils while the most

highly enriched immunological terms in less severe disease (Fig S2.4.3.1.4 B) were related to higher ratios of lymphocytes to neutrophils and monocytes, as well as signatures of healthy controls in vaccine and viral disease experiments.

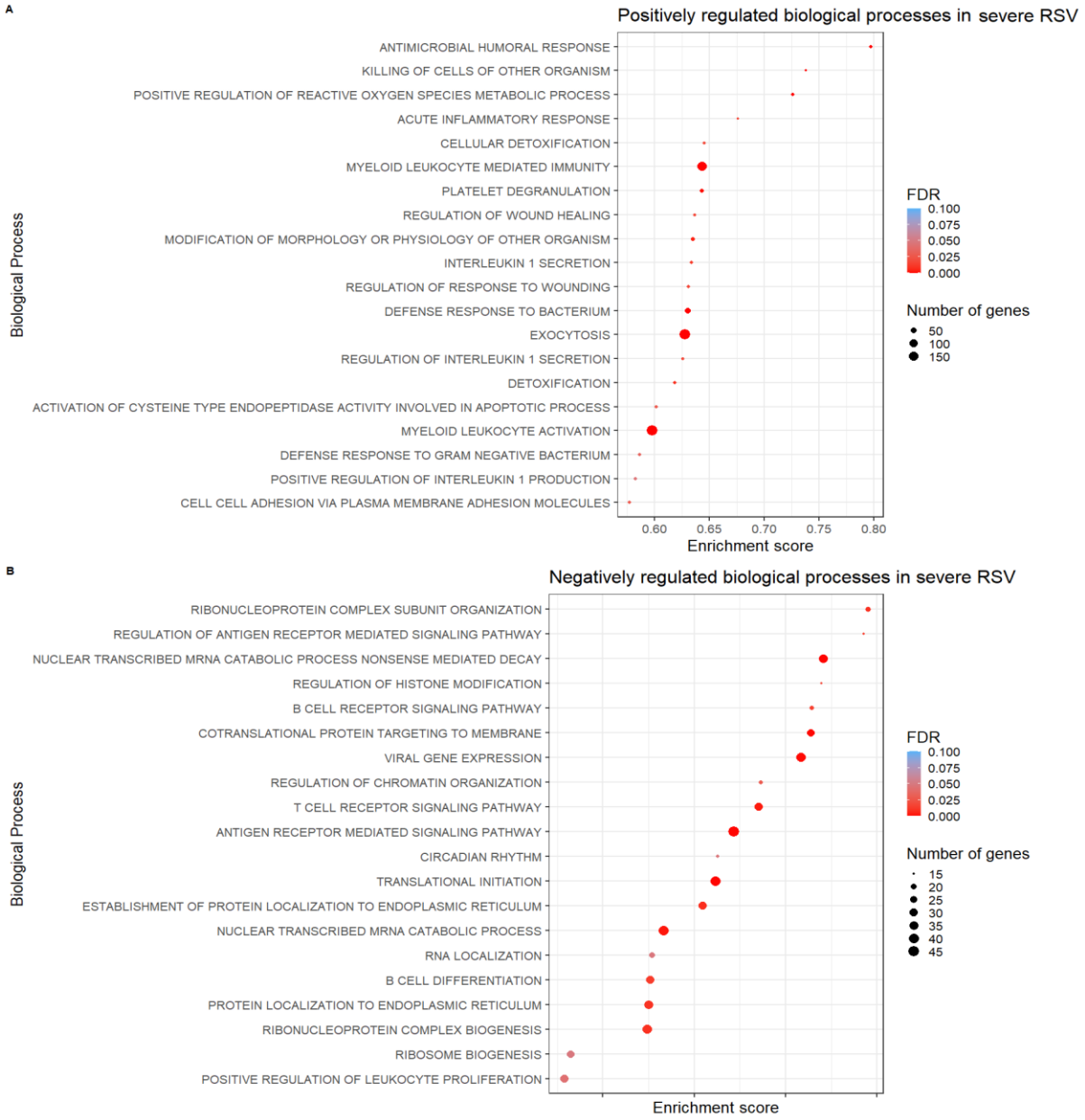


Figure S2.4.3.1.1 Gene set enrichment analysis of A – biological processes associated with more severe disease and B - biological processes associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the first component of the ReSVinet score disease severity PLS-DA. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

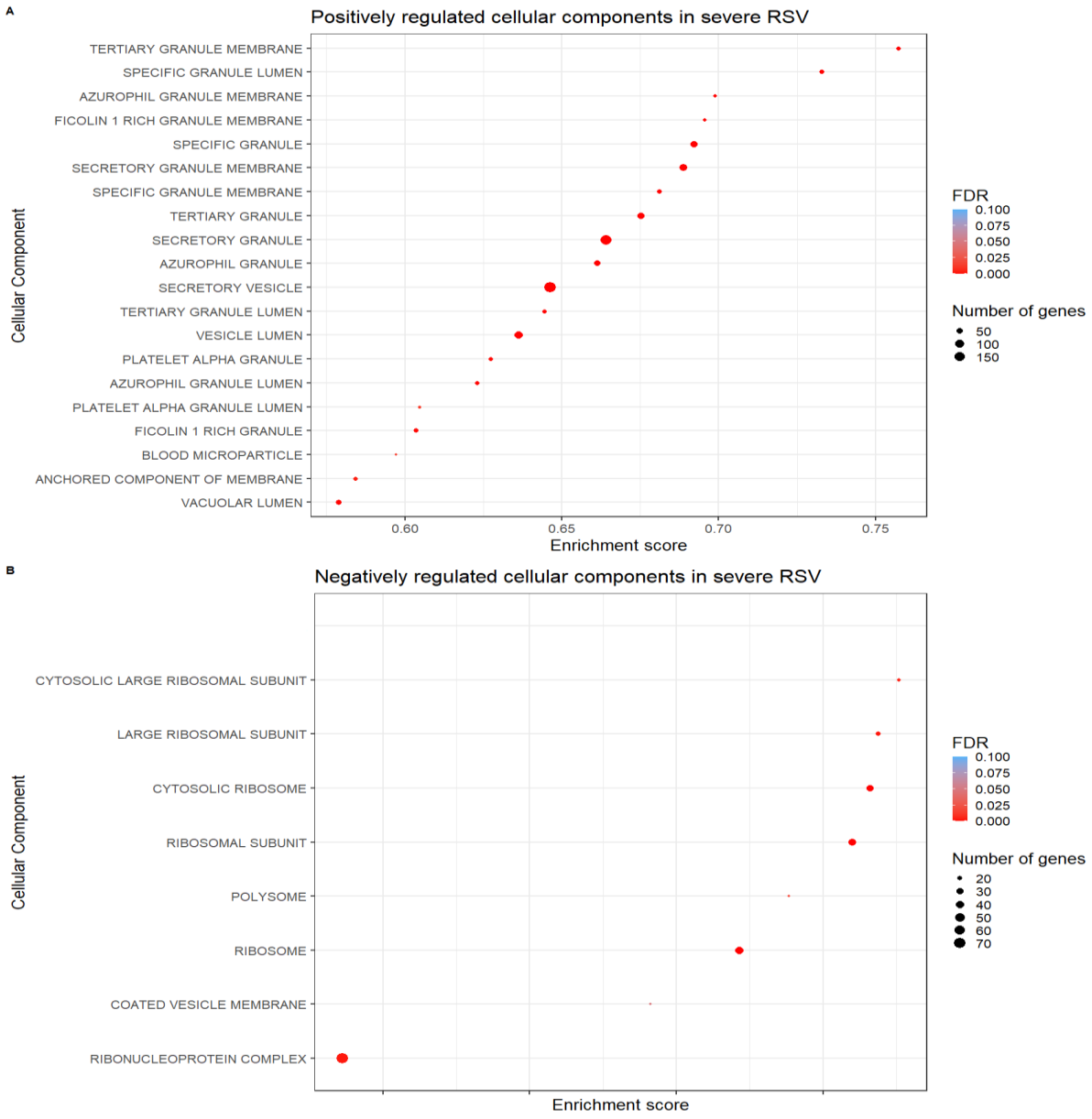


Figure S2.4.3.1.2 Gene set enrichment analysis of A – cellular components associated with more severe disease and B – cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the first component of the ReSVinet score disease severity PLS-DA. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

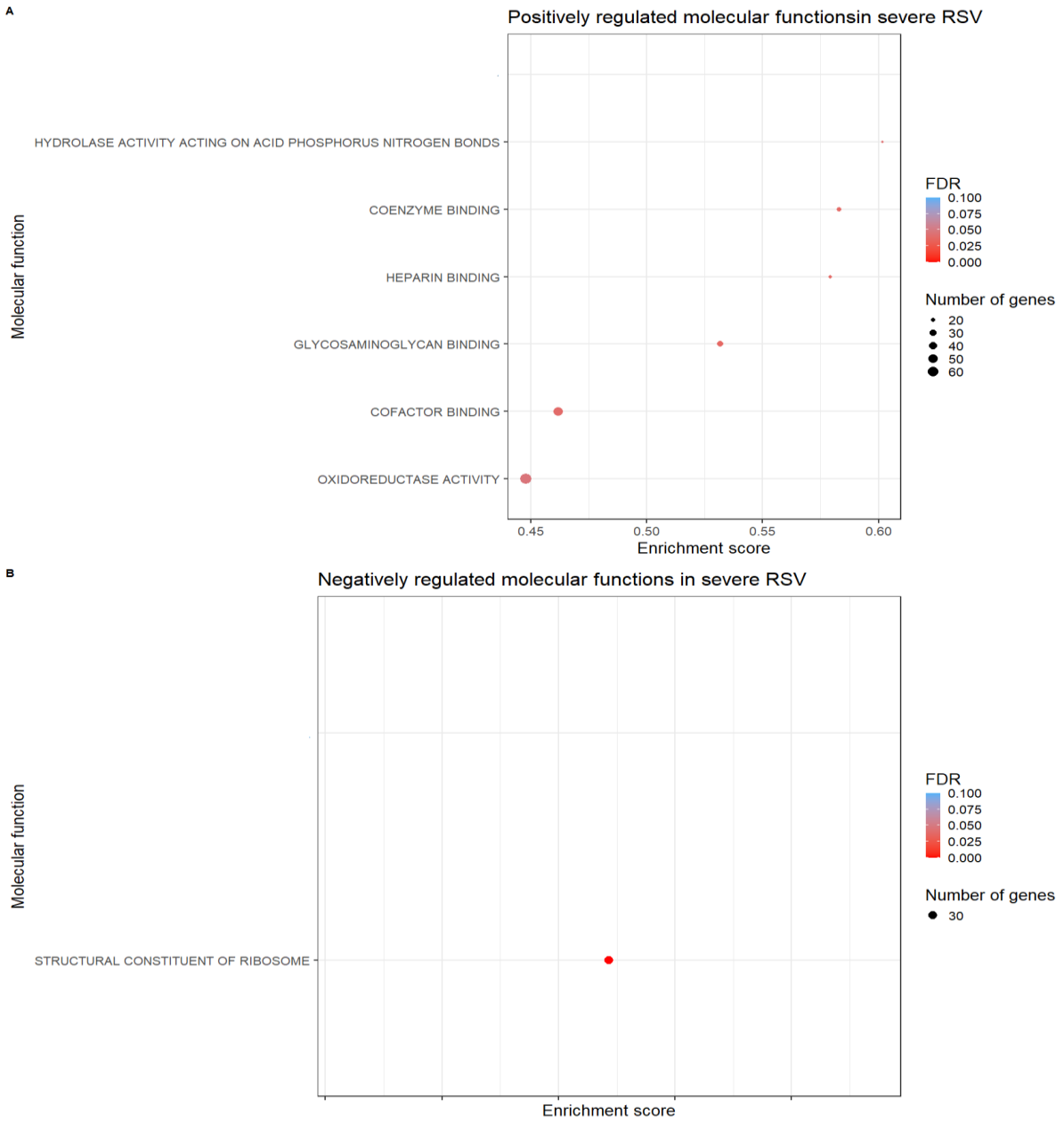


Figure S2.4.3.1.3 Gene set enrichment analysis of A – molecular functions associated with more severe disease and B – molecular functions associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the first component of the ReSVinet score disease severity PLS-DA. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

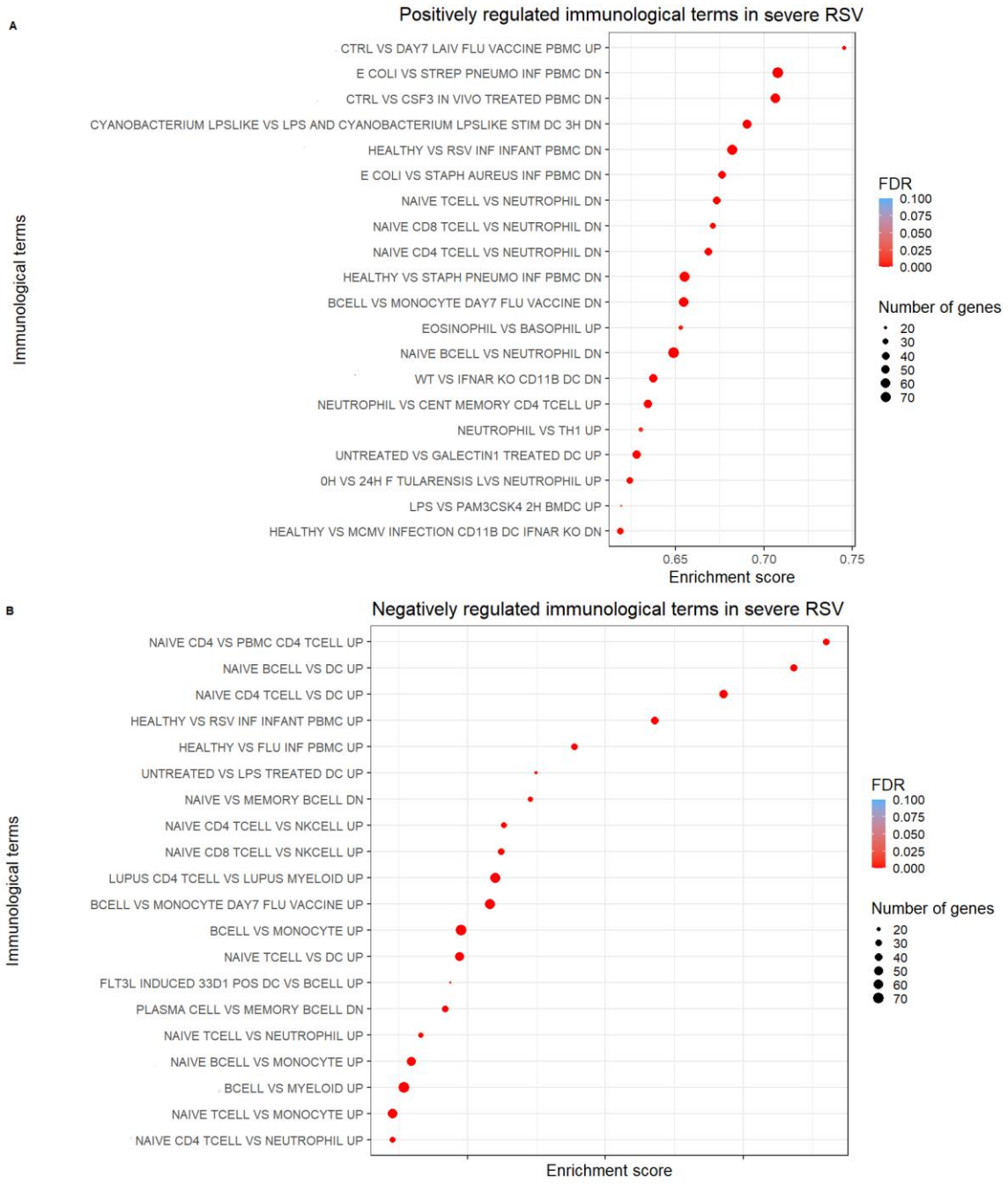


Figure S2.4.3.1.4 Gene set enrichment analysis of A – immunological terms associated with more severe disease and B – immunological terms associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the first component of the ReSVinet score disease severity PLS-DA. Terms were ranked according to enrichment score for the term calculated based on correlation values between the gene and requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

S2.4.3.2 Genes on second component of ReSVinet score PLS-DA associate with downregulation of lymphocytes and upregulation of neutrophils and monocytes

The loadings of all 1707 differentially expressed genes used in PLS-DA analysis were used to rank genes for gene set enrichment analyses. The second component identified by PLS-DA using ReSVinet score as the determinant of severity comprised of 8% of the variation in the set of differentially expressed genes.

There was only a single significantly enriched biological process terms in the second component (Fig S2.4.3.2.1). This term was positively correlated with more severe disease and related to detoxification.

There were also no significantly enriched cellular component terms correlating with more severe disease in the second component. All significantly enriched positively regulated terms, were related to ribosomes (Fig S2.4.3.2.2).

There were also no significantly enriched molecular functions in the second component.

Immunological terms that were associated with more severe disease (Fig S2.4.3.2.3 A) related to upregulation of terms related to days 3 and 7 post YF17D vaccination, and downregulation of genes related to day 1 post vaccination, as well as a high ratio of naïve lymphocyte related genes to memory lymphocyte related genes, and a high ratio of interferon related signalling to baseline. Immunological terms that were associated with less severe disease (Fig S2.4.3.2.3 B) related to a high ratio of antiviral responses to antibacterial responses, as well as a high ratio of lymphocytes to neutrophils and monocytes.

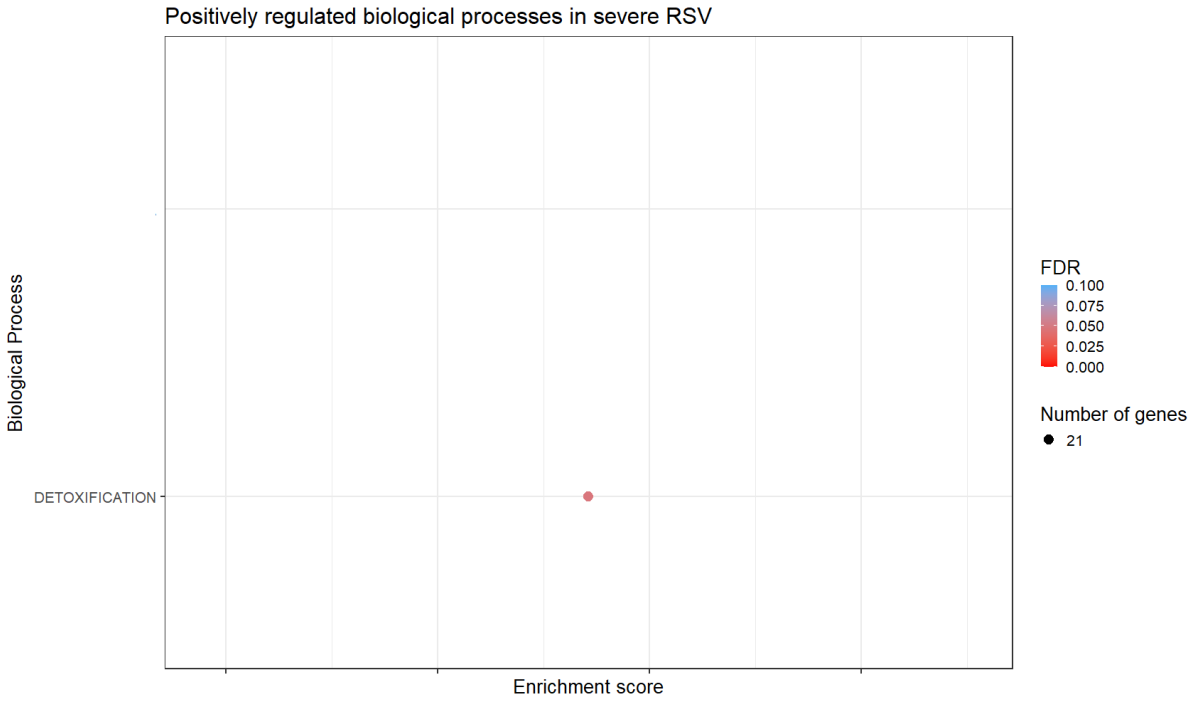


Figure S2.4.3.2.1 Gene set enrichment analysis of biological processes associated with more severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the second component of the ReSVinet score disease severity PLS-DA. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one

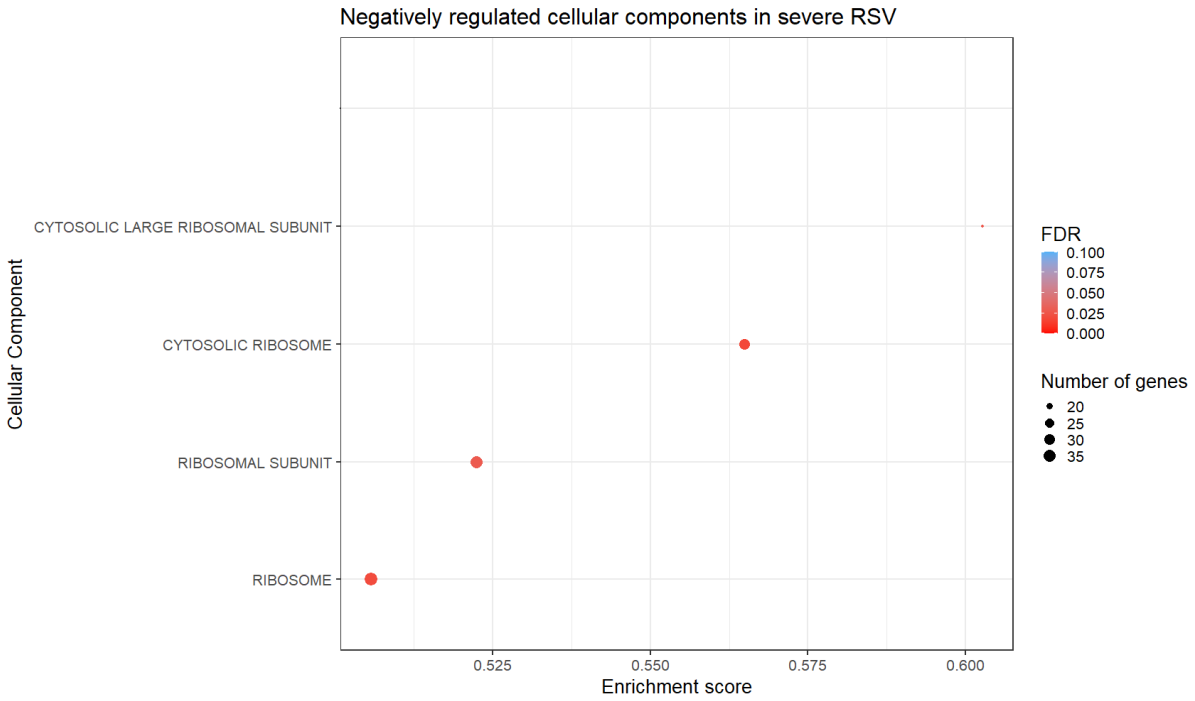


Figure S2.4.3.2.2 Gene set enrichment analysis of cellular components associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the second component of the ReSVinet score disease severity PLS-DA. Terms were ranked according to enrichment score for the term calculated based on correlation values between gene requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

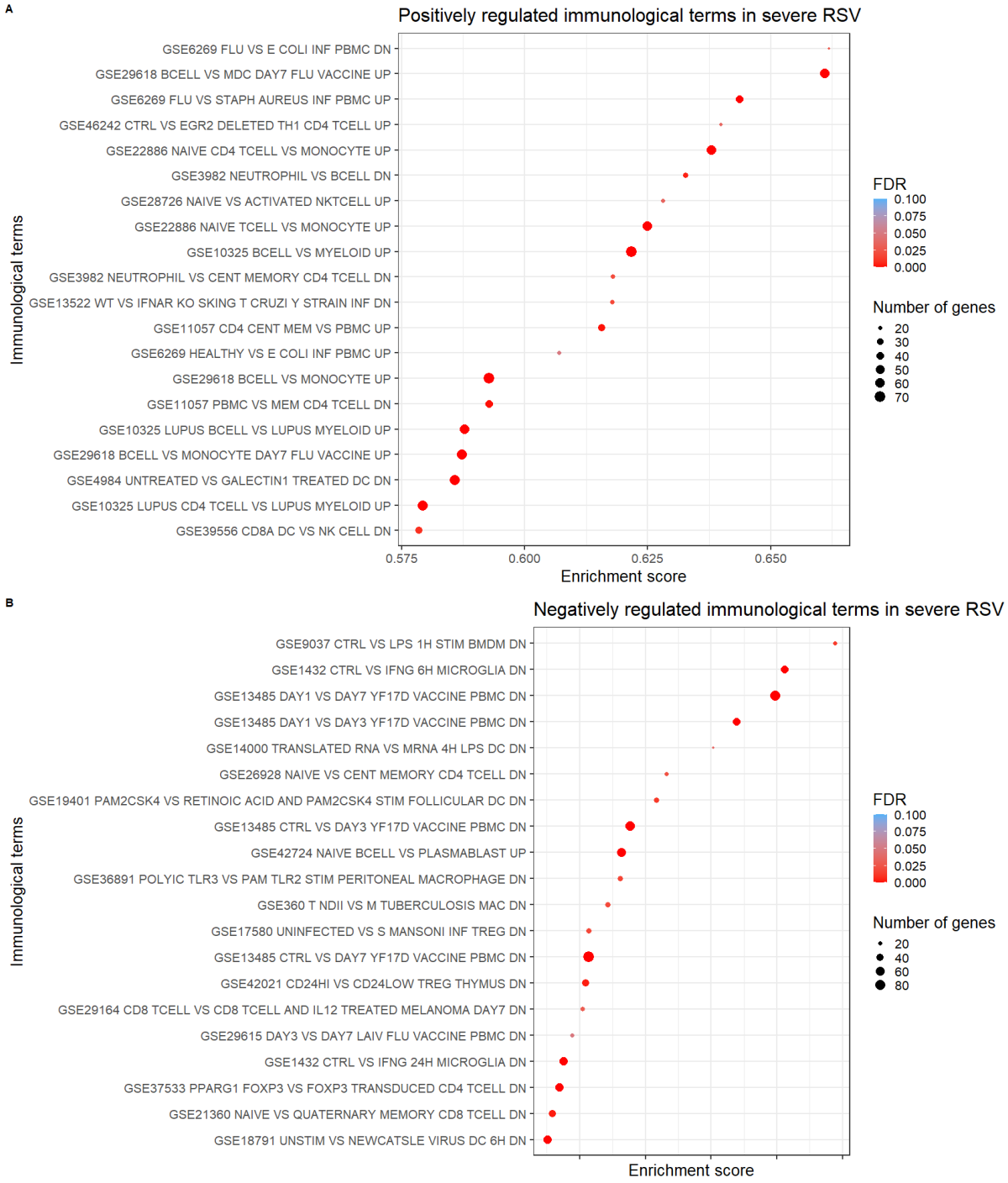


Figure S2.4.3.2.3 Gene set enrichment analysis of A – immunological terms associated with more severe disease and B – immunological terms associated with less severe disease from the set of differentially expressed genes between patients at the time of RSV infection and the same patients in convalescence ranked by PLS-DA loadings on the second component of the ReSVinet score disease severity PLS-DA. Terms were ranked according to enrichment score for the term calculated based on correlation values between the gene and requirement for hospitalisation. Size of dot represents number of genes that contribute towards the enrichment. Colour of dot represents the p.value of the enrichment term, with a blue colour representing a higher p.value and a red colour representing a lower one.

S2.5 Discussion

Genes that significantly correlate with each method of determining disease severity were similar and show a high degree of overlap. Requirement for respiratory support, and requirement for hospitalisation metrics showed a smaller number of genes that significantly correlated with more severe disease than the ReSVinet score or invasiveness of respiratory support metrics, and the majority of genes for these metrics were also significantly correlated with the ReSVinet score or invasiveness of respiratory support metrics. All genes that were significantly correlated with any metric for disease severity were also correlated in the same direction (positive or negative) with any other metric that they significantly correlate with. This indicates that all of these different measurements of disease severity were closely related and were comparable, although there were uniquely correlated genes for some metrics.

The importance of utilising the correct methods of assessing disease severity is important in both the development of clinical biomarkers and in improving understanding of disease. There are competing metrics for disease severity that have been used in previous biomarker-oriented studies(278)(279)(275). Because of this we aimed to determine the comparability of the patterns of gene expression that relate to these metrics, to determine if biomarkers that predict disease severity using different metrics are fundamentally comparable.

Depending on the metric used different proportions of the population of participants were classified as mild or severe. When using the recommended ReSVinet score threshold of 14 for severe disease as a cut-off only 14 participants (15% of the total) were classified as having severe disease. This was different from the proportion that were considered severe if participants were stratified by either requirement for respiratory support (63 participants, or 67% of the total number) or requirement for invasive respiratory support (39 participants, or 41% of the total number). Using requirement for hospitalisation as the metric for determining severity results in 72 (77% of the total) participants being considered as having severe disease. Due to the relatively small proportion of genes correlated with each severity metric that were uniquely correlated to that severity metric and no others, along with the fact that all genes correlated with each metric were correlated in the same direction (positively or negatively) with each other metric.

The second components showed differences depending on the severity metrics used. High interferon responses when compared to IL-4 responses associated with decreased disease severity according to the ReSVinet score, along with some associations between CD4 T cells and increased disease severity, suggesting that this component accounts for the T helper type 2 response to RSV, which has well characterised associations with disease severity(165)(166). The second component of the ReSVinet score PLS-DA also shows that a high ratio of IL-12 to non-IL-12 activated CD4 T cells associates with less severe disease, and that IL-12 (a cytokine commonly associated with Th1 type responses(530)) related signalling was downregulated in more severe disease, further indicating that the second component was likely primarily composed of Th2 response genes. This hypothesis is further supported, as genes related to IFN-γ production and early stage responses to the yellow fever vaccine YF17D (which are closely linked to IFN-γ responses(531)) were also downregulated in more severe disease.

Appendix chapter 3: Supplemental materials for chapter 5

3.1 Acute samples prior to normalization and batch correction

As a biomarker must be able to predict disease based on expression of a single sample at a single time point, only samples corresponding to cases of acute RSV infection were included, while healthy controls and convalescent samples were excluded. Quality control was performed on this subset of samples in the same manner as described in chapters 3.2 to 3.4, including sex chromosome filtration, normalization, and batch correction. Data of participants with acute infection after quality control are shown in chapter 5.3.

Principal component analysis was performed on the data set at this point and showed that the first component accounted for 47% of the variation in the gene expression of acute samples, while the second accounted for 12% and the third accounted for 6% (sfig 3.1.1). There were significant batch effects due to microarray batch (sfig 3.1.2 A) and due to cDNA yields (sfig 3.1.2 B) prior to normalization and batch correction. Principal component analysis also showed that prior to batch correction and normalization no clustering according to disease severity was visible on the first and second principal components (sfig 3.1.2 C).

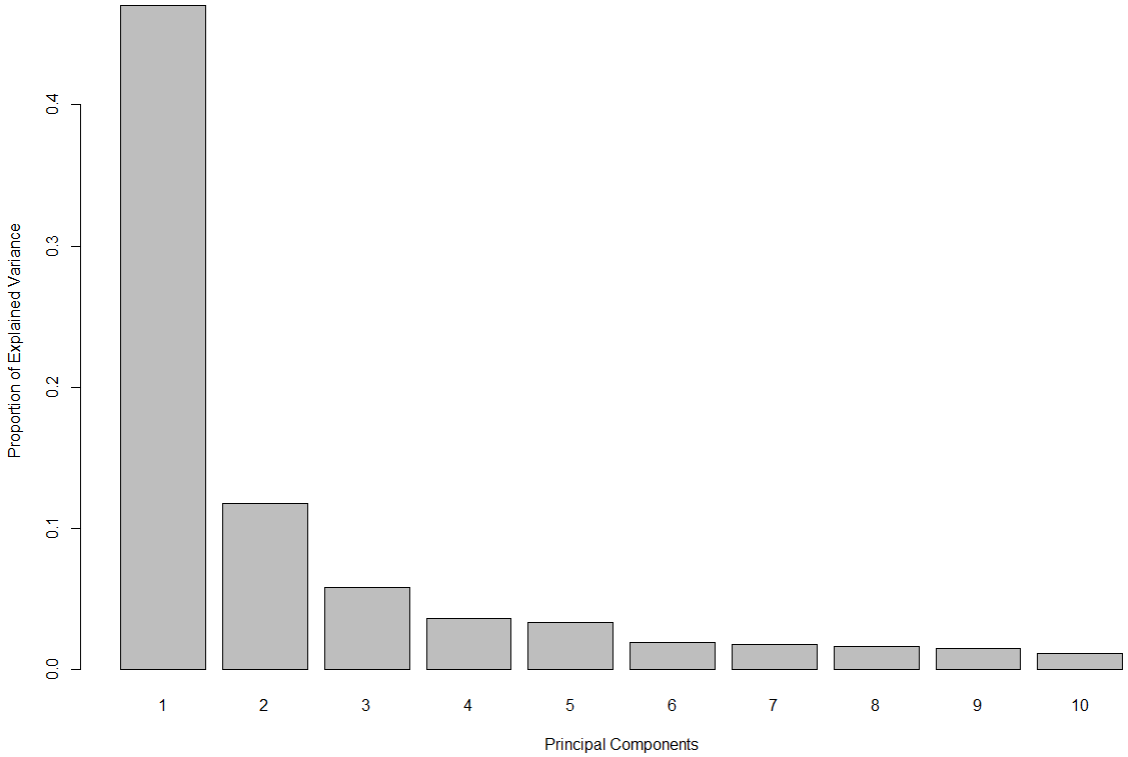


Figure S3.1.1 Scree plot of the first ten principal components of the samples from RSV infected patients at the time of acute disease before normalization and batch correction. Components account for more than 5% of the variation in the data set were used for feature selection

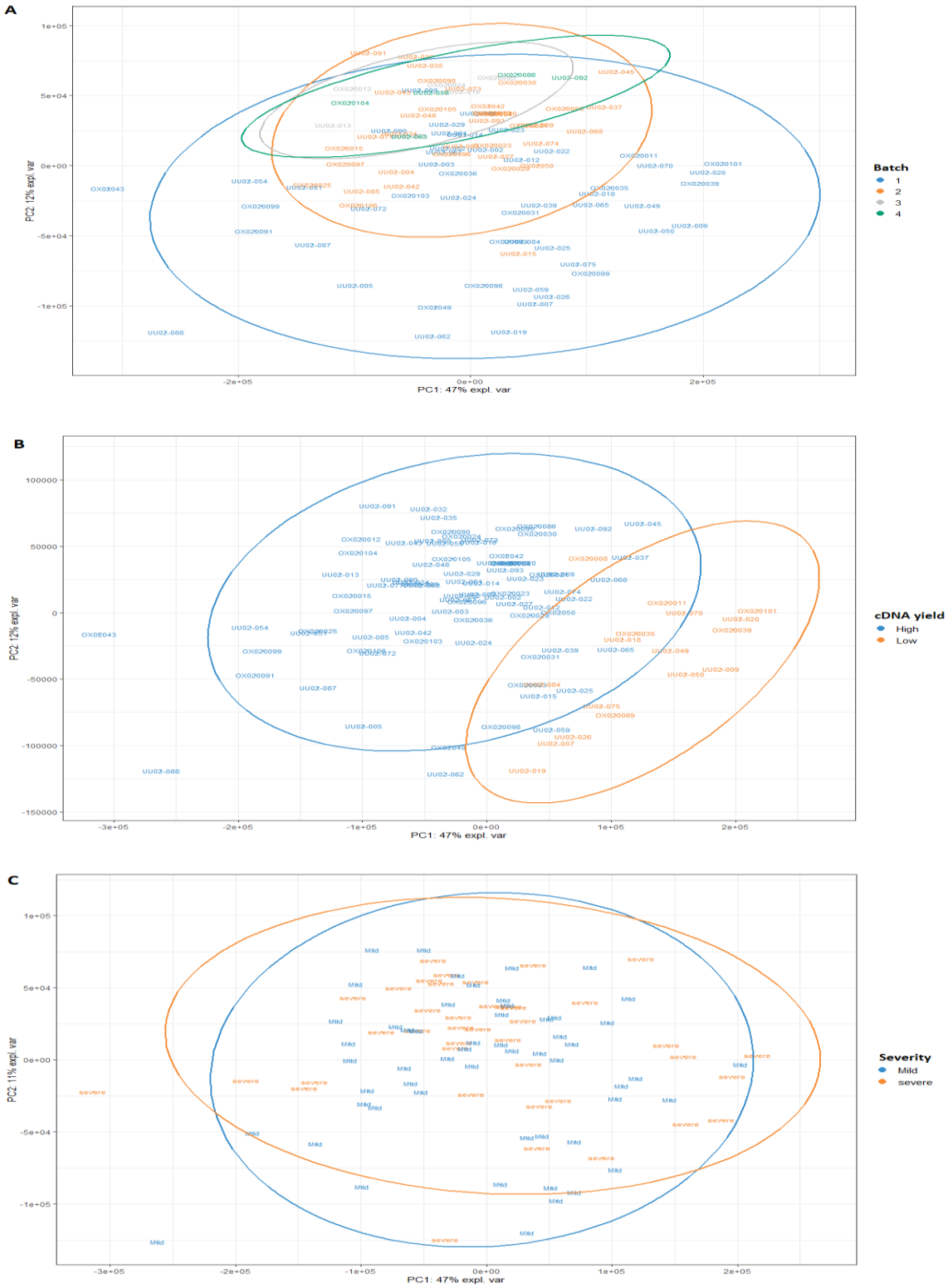


Figure S3.1.2 Mappings of samples at the time of acute disease on the first and second principal components before normalization and batch correction, coloured for A: batch effects according to microarray batch, B: cDNA yield, coloured according to high and low yield, where low yield samples had less than 5ng/µl, C: Disease severity according to requirement for invasive respiratory intervention at the time of recruitment.

3.2 List of models used in biomarker discovery

Algorithm
Linear SVM
Radial SVM
Polynomial SVM
Random forest
Generalized linear model
Neural networks
Linear discriminant analysis
Sparse linear discriminant analysis
Partial least squares
Sparse partial least squares
Neural networks with feature extraction
Nearest shrunken centroids
Boosted logistical regression
Naïve Bayes
High dimensional discriminant analysis
Bayesian generalised linear model
Binary discriminant analysis
Model averaged neural network
K nearest neighbours
heteroscedastic discriminant analysis
Flexible discriminant analysis
Elasticnet
Boosted generalized linear model
Mixture discriminant analysis
Quadratic discriminant analysis
Distance weighted discrimination
Sparse distance weighted discrimination
Nearest neighbour classifier
Stabilised nearest neighbour classifier
Stochastic gradient boosting

Table S3.2.1 List of models used in biomarker discovery

References

1. Jeffree CE, Rixon HWML, Brown G, Aitken J, Sugrue RJ. Distribution of the attachment (G) glycoprotein and GM1 within the envelope of mature respiratory syncytial virus filaments revealed using field emission scanning electron microscopy. Virology. 2003 Feb;306(2):254–67.

2. Ke Z, Dillard RS, Chirkova T, Leon F, Stobart CC, Hampton CM, et al. The morphology and assembly of respiratory syncytial virus revealed by cryo-electron tomography. Viruses. 2018 Aug 20;10(8):446.

3. Zhu C, Fu S, Zhou X, Yu L. Complete genome sequence of human respiratory syncytial virus from Lanzhou, China. Genome Announc. 2017 Aug 24;5(34):e00739-17.

4. Hendricks DA, Mclntosh K, Patterson JL. Further characterization of the soluble form of the G glycoprotein of respiratory syncytial virus. Virus Res. 1988 Jul;11(SUPPL. 1):15.

5. González-Reyes L, Ruiz-Argüello MB, García-Barreno B, Calder L, López JA, Albar JP, et al. Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. Proc Natl Acad Sci U S A. 2001 Aug;98(17):9859–64.

6. Gan SW, Tan E, Lin X, Yu D, Wang J, Tan GMY, et al. The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels. J Biol Chem. 2012 Jul;287(29):24671–89.

7. Webster Marketon JI, Corry J, Teng MN. The respiratory syncytial virus (RSV) nonstructural proteins mediate RSV suppression of glucocorticoid receptor transactivation. Virology. 2014;449:62–9.

8. Castagné N, Barbier A, Bernard J, Rezaei H, Huet JC, Henry C, et al. Biochemical characterization of the respiratory syncytial virus P-P and P-N protein complexes and localization of the P protein oligomerization domain. J Gen Virol. 2004;85(6):1643–53.

9. Cartee TL, Megaw AG, Oomens AGP, Wertz GW. Identification of a Single Amino Acid Change in the Human Respiratory Syncytial Virus L Protein That Affects Transcriptional Termination. J Virol. 2003 Jul;77(13):7352–60.

10. Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, Castagne N, et al. Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. Science (80-). 2009 Nov;326(5957):1279–83.

11. Mitra R, Baviskar P, Duncan-Decocq RR, Patel D, Oomens AGP. The Human Respiratory Syncytial Virus Matrix Protein Is Required for Maturation of Viral Filaments. J Virol. 2012 Apr 15;86(8):4432–43.

12. Bermingham A, Collins PL. The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription. Proc Natl Acad Sci U S A. 1999 Sep;96(20):11259–64.

13. Fearns R, Collins PL. Role of the M2-1 Transcription Antitermination Protein of Respiratory Syncytial Virus in Sequential Transcription. J Virol. 1999 Jul;73(7):5852–64.

14. Sullender WM. Respiratory Syncytial Virus Genetic and Antigenic Diversity. Clin Microbiol Rev. 2000 Jan 1;13(1):1–15.

15. McLellan JS, Ray WC, Peeples ME. Structure and Function of Respiratory Syncytial Virus Surface Glycoproteins. Curr Top Microbiol Immunol. 2013;372:83–104.

16. Zimmer G, Trotz I, Herrler G. N-Glycans of F Protein Differentially Affect Fusion Activity of Human Respiratory Syncytial Virus. J Virol. 2001 May;75(10):4744–51.

17. Tan L, Coenjaerts FEJ, Houspie L, Viveen MC, van Bleek GM, Wiertz EJHJ, et al. The Comparative Genomics of Human Respiratory Syncytial Virus Subgroups A and B: Genetic Variability and Molecular Evolutionary Dynamics. J Virol. 2013/05/22. 2013 Jul;87(14):8213–26.

18. García-Beato R, Martínez I, Francí C, Real FX, García-Barreno B, Melero JA. Host cell effect upon glycosylation and antigenicity of human respiratory syncytial virus G glycoprotein. Virology. 1996 Jul;221(2):301–9.

19. Roberts SR, Lichtenstein D, Ball LA, Wertz GW. The membrane-associated and secreted forms of the respiratory syncytial virus attachment glycoprotein G are synthesized from alternative initiation codons. J Virol. 1994 Jul;68(7):4538–46.

20. Bukreyev A, Yang L, Fricke J, Cheng L, Ward JM, Murphy BR, et al. The Secreted Form of Respiratory Syncytial Virus G Glycoprotein Helps the Virus Evade Antibody-Mediated Restriction of Replication by Acting as an Antigen Decoy and through Effects on Fc Receptor-Bearing Leukocytes. J Virol. 2008 Dec;82(24):12191–204.

21. Eshaghi AR, Duvvuri VR, Lai R, Nadarajah JT, Li A, Patel SN, et al. Genetic variability of human respiratory syncytial virus a strains circulating in Ontario: A novel genotype with a 72 nucleotide G gene duplication. PLoS One. 2012;7(3):e32807.

22. Trento A, Viegas M, Galiano M, Videla C, Carballal G, Mistchenko AS, et al. Natural History of Human Respiratory Syncytial Virus Inferred from Phylogenetic Analysis of the Attachment (G) Glycoprotein with a 60-Nucleotide Duplication. J Virol. 2006 Jan;80(2):975–84.

23. Flynn JA, Durr E, Swoyer R, Cejas PJ, Horton MS, Galli JD, et al. Stability Characterization of a Vaccine Antigen Based on the Respiratory Syncytial Virus Fusion Glycoprotein. PLoS One. 2016 Oct 20;11(10):e0164789.

24. Munywoki PK, Koech DC, Agoti CN, Bett A, Cane PA, Medley GF, et al. Frequent Asymptomatic Respiratory Syncytial Virus Infections during an Epidemic in a Rural Kenyan Household Cohort. J Infect Dis. 2015/05/04. 2015 Dec 1;212(10):1711–8.

25. Shi T, McAllister DA, O'Brien KL, Simoes EAF, Madhi SA, Gessner BD, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. Lancet. 2017;390(10098):946–58.

26. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012 Dec;380(9859):2095–128.

27. Peret TCT, Hall CB, Hammond GW, Piedra PA, Storch GA, Sullender WM, et al. Circulation Patterns of Group A and B Human Respiratory Syncytial Virus Genotypes in 5 Communities in North America. J Infect Dis. 2000 Jun;181(6):1891–6.

28. Janet S, Broad J, Snape MD. Respiratory syncytial virus seasonality and its implications on

prevention strategies. Hum Vaccines Immunother. 2017/12/15. 2018 Jan 2;14(1):234–44.

29. Shi T, Balsells E, Wastnedge E, Singleton R, Rasmussen ZA, Zar HJ, et al. Risk factors for respiratory syncytial virus associated with acute lower respiratory infection in children under five years: Systematic review and meta-analysis. J Glob Health. 2015 Dec;5(2):20416.

30. Checchia PA, Paes B, Bont L, Manzoni P, Simões EAF, Fauroux B, et al. Defining the Risk and Associated Morbidity and Mortality of Severe Respiratory Syncytial Virus Infection Among Infants with Congenital Heart Disease. Infect Dis Ther. 2017/01/09. 2017 Mar;6(1):37–56.

31. Geskey JM, Cyran SE. Managing the Morbidity Associated with Respiratory Viral Infections in Children with Congenital Heart Disease. Fisher R, editor. Int J Pediatr. 2012;2012:1–8.

32. Miller RG, Jackson CE, Kasarskis EJ, England JD, Forshew D, Johnston W, et al. Practice parameter update: The care of the patient with amyotrophic lateral sclerosis: Drug, nutritional, and respiratory therapies (an evidence-based review): Report of the quality standards subcommittee of the American academy of neurology. Neurology. 2009 Oct;73(15):1218–26.

33. Colin AA, McEvoy C, Castile RG. Respiratory morbidity and lung function in preterm infants of 32 to 36 weeks' gestational age. Pediatrics. 2010/06/07. 2010 Jul;126(1):115–28.

34. Taylor S, Taylor RJ, Lustig RL, Schuck-Paim C, Haguinet F, Webb DJ, et al. Modelling estimates of the burden of respiratory syncytial virus infection in children in the UK. BMJ Open. 2016 Jun 1;6(6):e009337.

35. Glezen WP, Taber LH, Frank AL, Kasel JA. Risk of Primary Infection and Reinfection With Respiratory Syncytial Virus. Am J Dis Child. 1986 Jun;140(6):543–6.

36. Houben ML, Bont L, Wilbrink B, Belderbos ME, Kimpen JLL, Visser GHA, et al. Clinical prediction rule for RSV bronchiolitis in healthy newborns: Prognostic birth cohort study. Pediatrics. 2011 Jan 1;127(1):35–41.

37. Henderson FW, Collier AM, Clyde WA, Denny FW. Respiratory-Syncytial-Virus Infections, Reinfections and Immunity: A Prospective, Longitudinal Study in Young Children. N Engl J Med. 1979 Mar;300(10):530–4.

38. Legg JP, Hussain IR, Warner JA, Johnston SL, Warner JO. Type 1 and type 2 cytokine imbalance in acute respiratory syncytial virus bronchiolitis. Am J Respir Crit Care Med. 2003 Sep;168(6 I):633–9.

39. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet. 2010 May;375(9725):1545–55.

40. Broor S, Parveen S, Bharaj P, Prasad VS, Srinivasulu KN, Sumanth KM, et al. A Prospective Three-Year Cohort Study of the Epidemiology and Virology of Acute Respiratory Infections of Children in Rural India. PLoS One. 2007 Jun 6;2(6):e491–e491.

41. Hall CB, Geiman JM, Biggar R, Kotok DI, Hogan PM, Douglas RG. Respiratory Syncytial Virus Infections within Families. N Engl J Med. 1976 Feb;294(8):414–9.

42. Sastre P, Ruiz T, Schildgen O, Schildgen V, Vela C, Rueda P. Seroprevalence of human respiratory syncytial virus and human metapneumovirus in healthy population analyzed by recombinant fusion protein-based enzyme linked immunosorbent assay. Virol J. 2012 Jul 2;9:130.

43. Ebihara T, Endo R, Kikuta H, Ishiguro N, Ishiko H, Kobayashi K. Comparison of the

Seroprevalence of Human Metapneumovirus and Human Respiratory Syncytial Virus. J Med Virol. 2004 Feb;72(2):304–6.

44. Lu G, Gonzalez R, Guo L, Wu C, Wu J, Vernet G, et al. Large-scale seroprevalence analysis of human metapneumovirus and human respiratory syncytial virus infections in Beijing, China. Virol J. 2011 Feb 10;8:62.

45. Kulkarni H, Smith CM, Lee DDH, Hirst RA, Easton AJ, O'Callaghan C. Evidence of respiratory syncytial virus spread by aerosol time to revisit infection control strategies? Am J Respir Crit Care Med. 2016 Aug;194(3):308–16.

46. Hall CB. Respiratory syncytial virus: Its transmission in the hospital environment. Yale J Biol Med. 1982;55(3–4):219–23.

47. Grayson SA, Griffiths PS, Perez MK, Piedimonte G. Detection of airborne respiratory syncytial virus in a pediatric acute care clinic. Pediatr Pulmonol. 2017 May;52(5):684–8.

48. Piedimonte G, Perez MK. Respiratory syncytial virus infection and bronchiolitis. Pediatr Rev. 2014 Dec;35(12):519–30.

49. Smith DK, Seales S, Budzik C. Respiratory syncytial virus bronchiolitis in children. Am Fam Physician. 2017 Jan;95(2):94–9.

50. Lieberthal AS, Bauchner H, Hall CB, Johnson DW, Kotagal U, Light MJ, et al. Diagnosis and management of bronchiolitis. Pediatrics. 2006 Oct;118(4):1774–93.

51. Borchers AT, Chang C, Gershwin ME, Gershwin LJ. Respiratory syncytial virus - A comprehensive review. Clin Rev Allergy Immunol. 2013 Dec;45(3):331–79.

52. Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, Staat MA, et al. The Burden of respiratory syncytial virus infection in young children. N Engl J Med. 2009 Feb;360(6):588–98.

53. Yang Q, Rasmussen SA, Friedman JM. Mortality associated with Down's syndrome in the USA from 1983 to 1997: A population-based study. Lancet. 2002 Mar;359(9311):1019–25.

54. Byard RW. Forensic issues in Down syndrome fatalities. J Forensic Leg Med. 2007 Nov;14(8):475–81.

55. James AL, Pare PD, Hogg JC. The mechanics of airway narrowing in asthma. Am Rev Respir Dis. 1989 Jan;139(1):242–6.

56. Weber MW, Mulholland EK, Greenwood BM. Respiratory syncytial virus infection in tropical and developing countries. Trop Med Int Heal. 1998 Apr 1;3(4):268–80.

57. Ostadabbas S, Bulach C, Ku DN, Anderson LJ, Ghovanloo M. A passive quantitative measurement of airway resistance using depth data. 2014 36th Annu Int Conf IEEE Eng Med Biol Soc EMBC 2014. 2014;2014:5743–7.

58. Okiro EA, Ngama M, Bett A, Nokes DJ. The Incidence and Clinical Burden of Respiratory Syncytial Virus Disease Identified through Hospital Outpatient Presentations in Kenyan Children. PLoS One. 2012 Dec 26;7(12):e52520.

59. Tristram DA, Hicks W, Hard R. Respiratory syncytial virus and human bronchial epithelium. Arch Otolaryngol - Head Neck Surg. 1998 Jul;124(7):777–83.

60. Russell CD, Unger SA, Walton M, Schwarze J. The human immune response to respiratory syncytial virus infection. Clin Microbiol Rev. 2017;30(2):481–502.

61. Everard ML, Swarbrick A, Wrightham M, McIntyre J, Dunkley C, James PD, et al. Analysis of

cells obtained by bronchial lavage of infants with respiratory syncytial virus infection. Arch Dis Child. 1994 Nov;71(5):428–32.

62. Johnson JE, Gonzales RA, Olson SJ, Wright PF, Graham BS. The histopathology of fatal untreated human respiratory syncytial virus infection. Mod Pathol. 2007 Jan;20(1):108–19.

63. El Saleeby CM, Bush AJ, Harrison LM, Aitken JA, DeVincenzo JP. Respiratory syncytial virus load, viral dynamics, and disease severity in previously healthy naturally infected children. J Infect Dis. 2011 Oct;204(7):996–1002.

64. Hasegawa K, Jartti T, Mansbach JM, Laham FR, Jewell AM, Espinola JA, et al. Respiratory syncytial virus genomic load and disease severity among children hospitalized with bronchiolitis: Multicenter cohort studies in the United States and Finland. J Infect Dis. 2015 May;211(10):1550–9.

65. Houben ML, Coenjaerts FEJ, Rossen JWA, Belderbos ME, Hofland RW, Kimpen JLL, et al. Disease severity and viral load are correlated in infants with primary respiratory syncytial virus infection in the community. J Med Virol. 2010 Jul;82(7):1266–71.

66. Fodha I, Vabret A, Ghedira L, Seboui H, Chouchane S, Dewar J, et al. Respiratory syncytial virus infections in hospitalized infants: Association between viral load, virus subgroup, and disease severity. J Med Virol. 2007 Dec;79(12):1951–8.

67. Mazur NI, Bont L, Cohen AL, Cohen C, Von Gottberg A, Groome MJ, et al. Severity of respiratory syncytial virus lower respiratory tract infection with viral coinfection in HIV-uninfected children. Clin Infect Dis. 2017 Feb;64(4):443–50.

68. Wright PF, Gruber WC, Peters M, Reed G, Zhu Y, Robinson F, et al. Illness Severity, Viral Shedding, and Antibody Responses in Infants Hospitalized with Bronchiolitis Caused by Respiratory Syncytial Virus. J Infect Dis. 2002 Apr;185(8):1011–8.

69. Garcia-Mauriño C, Moore-Clingenpeel M, Thomas J, Mertz S, Cohen DM, Ramilo O, et al. Viral load dynamics and clinical disease severity in infants with respiratory syncytial virus infection. J Infect Dis. 2019 Nov 12;219(8):1207–15.

70. Piedra FA, Mei M, Avadhanula V, Mehta R, Aideyan L, Garofalo RP, et al. The interdependencies of viral load, the innate immune response, and clinical outcome in children presenting to the emergency department with respiratory syncytial virus-Associated bronchiolitis. PLoS One. 2017;12(3):e0172953.

71. Nicholson EG, Schlegel C, Garofalo RP, Mehta R, Scheffler M, Mei M, et al. Robust cytokine and chemokine response in nasopharyngeal secretions: Association with decreased severity in children with physician diagnosed bronchiolitis. J Infect Dis. 2016 Aug;214(4):649–55.

72. Thwaites RS, Coates M, Ito K, Ghazaly M, Feather C, Abdulla F, et al. Reduced nasal viral load and IFN responses in infants with respiratory syncytial virus bronchiolitis and respiratory failure. Am J Respir Crit Care Med. 2018 Oct;198(8):1074–84.

73. Krusat T, Streckert HJ. Heparin-dependent attachment of respiratory syncytial virus (RSV) to host cells. Arch Virol. 1997;142(6):1247–54.

74. Hallak LK, Spillmann D, Collins PL, Peeples ME. Glycosaminoglycan Sulfation Requirements for Respiratory Syncytial Virus Infection. J Virol. 2000 Nov 15;74(22):10508–13.

75. Hallak LK, Collins PL, Knudson W, Peeples ME. Iduronic acid-containing glycosaminoglycans on target cells are required for efficient respiratory syncytial virus infection. Virology. 2000 Jun;271(2):264–75.

76. Zhang L, Bukreyev A, Thompson CI, Watson B, Peeples ME, Collins PL, et al. Infection of Ciliated Cells by Human Parainfluenza Virus Type 3 in an In Vitro Model of Human Airway Epithelium. J Virol. 2005 Jan;79(2):1113–24.

77. Chirkova T, Lin S, Oomens AGP, Gaston KA, Boyoglu-Barnum S, Meng J, et al. CX3CR1 is an important surface molecule for respiratory syncytial virus infection in human airway epithelial cells. J Gen Virol. 2015/06/25. 2015 Sep;96(9):2543–56.

78. Escribano-Romero E, Rawling J, García-Barreno B, Melero JA. The Soluble Form of Human Respiratory Syncytial Virus Attachment Protein Differs from the Membrane-Bound Form in Its Oligomeric State but Is Still Capable of Binding to Cell Surface Proteoglycans. J Virol. 2004 Apr;78(7):3524–32.

79. Tripp RA, Jones LP, Haynes LM, Zheng HQ, Murphy PM, Anderson LJ. CX3C chemokine mimicry by respiratory syncytial virus G glycoprotein. Nat Immunol. 2001 Aug;2(8):732–8.

80. Johnson SM, McNally BA, Ioannidis I, Flano E, Teng MN, Oomens AG, et al. Respiratory Syncytial Virus Uses CX3CR1 as a Receptor on Primary Human Airway Epithelial Cultures. PLoS Pathog. 2015 Dec;11(12):e1005318.

81. Jeong K Il, Piepenhagen PA, Kishko M, DiNapoli JM, Groppo RP, Zhang L, et al. CX3CR1 is expressed in differentiated human ciliated airway cells and co-localizes with respiratory syncytial virus on cilia in a G protein-dependent manner. PLoS One. 2015 Jun 24;10(6):e0130517.

82. Zhivaki D, Lemoine S, Lim A, Morva A, Vidalain PO, Schandene L, et al. Respiratory Syncytial Virus Infects Regulatory B Cells in Human Neonates via Chemokine Receptor CX3CR1 and Promotes Lung Disease Severity. Immunity. 2017 Feb;46(2):301–14.

83. Karron RA, Buonagurio DA, Georgiu AF, Whitehead SS, Adamus JE, Clements-Mann M Lou, et al. Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vitro: Clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant. Proc Natl Acad Sci U S A. 1997 Dec;94(25):13961–6.

84. Feldman SA, Audet S, Beeler JA. The Fusion Glycoprotein of Human Respiratory Syncytial Virus Facilitates Virus Attachment and Infectivity via an Interaction with Cellular Heparan Sulfate. J Virol. 2000 Jul;74(14):6442–7.

85. Battles MB, McLellan JS. Respiratory syncytial virus entry and how to block it. Nat Rev Microbiol. 2019;17(4):233–45.

86. Tayyari F, Marchant D, Moraes TJ, Duan W, Mastrangelo P, Hegele RG. Identification of nucleolin as a cellular receptor for human respiratory syncytial virus. Nat Med. 2011 Aug;17(9):1132–5.

87. Currier MG, Lee S, Stobart CC, Hotard AL, Villenave R, Meng J, et al. EGFR Interacts with the Fusion Protein of Respiratory Syncytial Virus Strain 2-20 and Mediates Infection and Mucin Expression. PLoS Pathog. 2016 May 6;12(5):e1005622.

88. Behera AK, Matsuse H, Kumar M, Kong X, Lockey RF, Mohapatra SS. Blocking intercellular adhesion molecule-1 on human epithelial cells decreases respiratory syncytial virus infection. Biochem Biophys Res Commun. 2001 Jan;280(1):188–95.

89. Krzyzaniak MA, Zumstein MT, Gerez JA, Picotti P, Helenius A. Host Cell Entry of Respiratory Syncytial Virus Involves Macropinocytosis Followed by Proteolytic Activation of the F Protein. PLoS Pathog. 2013;9(4):e1003309.

90. San-Juan-Vergara H, Sampayo-Escobar V, Reyes N, Cha B, Pacheco-Lugo L, Wong T, et al. Cholesterol-Rich Microdomains as Docking Platforms for Respiratory Syncytial Virus in Normal Human Bronchial Epithelial Cells. J Virol. 2011/11/16. 2012 Feb;86(3):1832–43.

91. White JM, Delos SE, Brecher M, Schornberg K. Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme. Crit Rev Biochem Mol Biol. 2008;43(3):189–219.

92. Schlender J, Zimmer G, Herrler G, Conzelmann K-K. Respiratory Syncytial Virus (RSV) Fusion Protein Subunit F2, Not Attachment Protein G, Determines the Specificity of RSV Infection. J Virol. 2003 Apr;77(8):4609–16.

93. Liljeroos L, Krzyzaniak MA, Helenius A, Butcher SJ. Architecture of respiratory syncytial virus revealed by electron cryotomography. Proc Natl Acad Sci U S A. 2013 Jul 2;110(27):11133–8.

94. Yunus AS, Jackson TP, Crisafi K, Burimski I, Kilgore NR, Zoumplis D, et al. Elevated temperature triggers human respiratory syncytial virus F protein six-helix bundle formation. Virology. 2009/11/18. 2010 Jan 20;396(2):226–37.

95. Bagga B, Cehelsky JE, Vaishnaw A, Tomwilkinson T, Meyers R, Harrison LM, et al. Effect of Preexisting Serum and Mucosal Antibody on Experimental Respiratory Syncytial Virus (RSV) Challenge and Infection of Adults. J Infect Dis. 2015 Dec;212(10):1719–25.

96. Lee FEH, Walsh EE, Falsey AR, Betts RF, Treanor JJ. Experimental infection of humans with A2 respiratory syncytial virus. Antiviral Res. 2004 Sep;63(3):191–6.

97. Raes M, Peeters V, Alliet P, Gillis P, Kortleven J, Magerman K, et al. Peripheral blood T and B lymphocyte subpopulations in infants with acute respiratory syncytial virus brochiolitis. Pediatr Allergy Immunol. 1997 May;8(2):97–102.

98. Reed JL, Welliver TP, Sims GP, McKinney L, Velozo L, Avendano L, et al. Innate Immune Signals Modulate Antiviral and Polyreactive Antibody Responses during Severe Respiratory Syncytial Virus Infection. J Infect Dis. 2009 Apr;199(8):1128–38.

99. McNamara PS, Fonceca AM, Howarth D, Correia JB, Slupsky JR, Trinick RE, et al. Respiratory syncytial virus infection of airway epithelial cells, in vivo and in vitro, supports pulmonary antibody responses by inducing expression of the B cell differentiation factor BAFF. Thorax. 2013 Jan;68(1):76–81.

100. Belnoue E, Pihlgren M, McGaha TL, Tougne C, Rochat AF, Bossen C, et al. APRIL is critical for plasmablast survival in the bone marrow and poorly expressed by early-life bone marrow stromal cells. Blood. 2008 Mar;111(5):2755–64.

101. Thai LH, Le Gallou S, Robbins A, Crickx E, Fadeev T, Zhou Z, et al. BAFF and CD4 1 T cells are major survival factors for long-lived splenic plasma cells in a B-cell–depletion context. Blood. 2018 Apr;131(14):1545–55.

102. Shinoff JJ, O'Brien KL, Thumar B, Shaw JB, Reid R, Hua W, et al. Young Infants Can Develop Protective Levels of Neutralizing Antibody after Infection with Respiratory Syncytial Virus. J Infect Dis. 2008 Oct 1;198(7):1007–15.

103. Habibi MS, Jozwik A, Makris S, Dunning J, Paras A, DeVincenzo JP, et al. Impaired antibody-mediated protection and defective iga b-cell memory in experimental infection of adults with respiratory syncytial virus. Am J Respir Crit Care Med. 2015 May;191(9):1040–9.

104. TSUTSUMI H, MATSUDA K, YAMAZAKI H, OGRA PL, CHIBA S. Different kinetics of antibody responses between IgA and IgG classes in nasopharyngeal secretion in infants and children

during primary respiratory syncytial virus infection. Pediatr Int. 1995 Aug;37(4):464–8.

105. Freitas GRO, Silva DAO, Yokosawa J, Paula NT, Costa LF, Carneiro BM, et al. Antibody response and avidity of respiratory syncytial virus-specific total IgG, IgG1, and IgG3 in young children. J Med Virol. 2011 Oct;83(10):1826–33.

106. Chu HY, Steinhoff MC, Magaret A, Zaman K, Roy E, Langdon G, et al. Respiratory Syncytial Virus Transplacental Antibody Transfer and Kinetics in Mother-Infant Pairs in Bangladesh. J Infect Dis. 2014 Nov;210(10):1582–9.

107. Glezen WP, Paredes A, Allison JE, Taber LH, Frank AL. Risk of respiratory syncytial virus infection for infants from low-income families in relationship to age, sex, ethnic group, and maternal antibody level. J Pediatr. 1981 May;98(5):708–15.

108. Jans J, Wicht O, Widjaja I, Ahout IML, De Groot R, Guichelaar T, et al. Characteristics of RSV-specific maternal antibodies in plasma of hospitalized, acute RSV patients under three months of age. PLoS One. 2017 Jan 30;12(1):e0170877–e0170877.

109. Murphy BR, Alling DW, Snyder MH, Walsh EE, Prince GA, Chanock RM, et al. Effect of age and preexisting antibody on serum antibody response of infants and children to the F and G glycoproteins during respiratory syncytial virus infection. J Clin Microbiol. 1986 Jun;24(5):894–8.

110. Yamazaki H, Tsutsumi H, Matsuda K, Nagai K, Ogra PL, Chiba S. Effect of maternal antibody on IgA antibody response in nasopharyngeal secretion in infants and children during primary respiratory syncytial virus infection. J Gen Virol. 1994 Aug;75(8):2115–9.

111. Bont L, Heijnen CJ, Kavelaars A, van Aalderen WMC, Brus F, Draaisma JMT, et al. Local Interferon-γ Levels during Respiratory Syncytial Virus Lower Respiratory Tract Infection Are Associated with Disease Severity. J Infect Dis. 2001 Jun 26;184(3):355–8.

112. Eichinger KM, Egaña L, Orend JG, Resetar E, Anderson KB, Patel R, et al. Alveolar macrophages support interferon gamma-mediated viral clearance in RSVinfected neonatal mice. Respir Res. 2015;16(1):122.

113. Semple MG, Dankert HM, Ebrahimi B, Correia JB, Booth JA, Stewart JP, et al. Severe respiratory syncytial virus bronchiolitis in infants is associated with reduced airway interferon gamma and substance P. PLoS One. 2007 Oct;2(10):e1038.

114. Smeltz RB, Chen J, Ehrhardt R, Shevach EM. Role of IFN-γ in Th1 Differentiation: IFN-γ Regulates IL-18Rα Expression by Preventing the Negative Effects of IL-4 and by Inducing/Maintaining IL-12 Receptor β2 Expression. J Immunol. 2002 Jun;168(12):6165–72.

115. Pinto RA, Arredondo SM, Bono MR, Gaggero AA, Díaz P V. T helper 1/T helper 2 cytokine imbalance in respiratory syncytial virus infection is associated with increased endogenous plasma cortisol. Pediatrics. 2006 May;117(5):e878-86.

116. Chung HL, Park HJ, Kim SY, Kim SG. Age-related difference in immune responses to respiratory syncytial virus infection in young children. Pediatr Allergy Immunol. 2007 Mar;18(2):94–9.

117. Hassan MA, Eldin AME, Ahmed MM. T - helper2 /T - helper1 imbalance in respiratory syncytial virus bronchiolitis in relation to disease severity and outcome. Egypt J Immunol. 2008;15(2):153–60.

118. Bont L, Heijnen CJ, Kavelaars A, Van Aalderen WMC, Brus F, Draaisma JTM, et al. Peripheral blood cytokine responses and disease severity in respiratory syncytial virus bronchiolitis. Eur

Respir J. 1999 Jul;14(1):144–9.

119. Hattori S, Shimojo N, Mashimo Y, Inoue Y, Ono Y, Kohno Y, et al. Relationship between RANTES polymorphisms and respiratory syncytial virus bronchiolitis in a Japanese infant population. Jpn J Infect Dis. 2011;64(3):242–5.

120. Fernández JA, Roine I, Vasquez A, Cáneo M. Soluble interleukin-2 receptor (sCD25) and interleukin-10 plasma concentrations are associated with severity of primary respiratory syncytial virus (RSV) infection. Eur Cytokine Netw. 2005;16(1):81–90.

121. Aberle JH, Aberle SW, Dworzak MN, Mandl CW, Rebhandl W, Vollnhofer G, et al. Reduced interferon-γ expression in peripheral blood mononuclear cells of infants with severe respiratory syncytial virus disease. Am J Respir Crit Care Med. 1999 Oct;160(4):1263–8.

122. Bermejo-Martin JF, Garcia-Arevalo MC, Alonso A, De Lejarazu RO, Pino M, Resino S, et al. Persistence of proinflammatory response after severe respiratory syncytial virus disease in children. Vol. 119, Journal of Allergy and Clinical Immunology. United States; 2007. p. 1547–50.

123. Oshansky CM, Krunkosky TM, Barber J, Jones LP, Tripp RA. Respiratory syncytial virus proteins modulate suppressors of cytokine signaling 1 and 3 and the type I interferon response to infection by a toll-like receptor pathway. Viral Immunol. 2009 Jun;22(3):147–61.

124. Lo MS, Brazas RM, Holtzman MJ. Respiratory Syncytial Virus Nonstructural Proteins NS1 and NS2 Mediate Inhibition of Stat2 Expression and Alpha/Beta Interferon Responsiveness. J Virol. 2005 Jul;79(14):9315–9.

125. Johansson C. Respiratory syncytial virus infection: An innate perspective. F1000Research. 2016 Dec 21;5:2898.

126. Goritzka M, Makris S, Kausar F, Durant LR, Pereira C, Kumagai Y, et al. Alveolar macrophage-derived type I interferons orchestrate innate immunity to RSV through recruitment of antiviral monocytes. J Exp Med. 2015 May 4;212(5):699–714.

127. Bertrand P, Lay MK, Piedimonte G, Brockmann PE, Palavecino CE, Hernández J, et al. Elevated IL-3 and IL-12p40 levels in the lower airway of infants with RSV-induced bronchiolitis correlate with recurrent wheezing. Cytokine. 2015 Dec;76(2):417–23.

128. McNamara PS, Flanagan BF, Hart CA, Smyth RL. Production of Chemokines in the Lungs of Infants with Severe Respiratory Syncytial Virus Bronchiolitis. J Infect Dis. 2005 Apr 15;191(8):1225–32.

129. Davidson S, Crotta S, McCabe TM, Wack A. Pathogenic potential of interferon αβ in acute influenza infection. Nat Commun. 2014 May;5:3864.

130. Boelen A, Kwakkel J, Barends M, De Rond L, Dormans J, Kimman T. Effect of lack of Interleukin-4, Interleukin-12, Interleukin-18, or the Interferon-γ receptor on virus replication, cytokine response, and lung pathology during respiratory syncytial virus infection in mice. J Med Virol. 2002;66(4):552–60.

131. Goritzka M, Durant LR, Pereira C, Salek-Ardakani S, Openshaw PJM, Johansson C. Alpha/Beta Interferon Receptor Signaling Amplifies Early Proinflammatory Cytokine Production in the Lung during Respiratory Syncytial Virus Infection. J Virol. 2014;88(11):6128–36.

132. Lee YM, Miyahara N, Takeda K, Prpich J, Oh A, Balhorn A, et al. IFN-γ production during initial infection determines the outcome of reinfection with respiratory syncytial virus. Am J Respir Crit Care Med. 2008;177(2):208–18.

133. Mosquera RA, Stark JM, Atkins CL, Colasurdo GN, Chevalier J, Samuels CL, et al. Functional and immune response to respiratory syncytial virus infection in aged BALB/c mice: A search for genes determining disease severity. Exp Lung Res. 2013/12/19. 2014;40(1):40–9.

134. Pennings JLA, Mariman R, Hodemaekers HM, Reemers SSN, Janssen R, Guichelaar T. Transcriptomics in lung tissue upon respiratory syncytial virus infection reveals aging as important modulator of immune activation and matrix maintenance. Sci Rep. 2018/11/09. 2018;8(1):16653.

135. Kannangai R, Joseph Finny G, Jacob John T, Sridharan G, Gopal R. Role of interferon gamma in the pathogenesis of primary respiratory syncytial virus infection in BALB/c mice. J Med Virol. 2000;62(2):257–66.

136. Smit JJ, Rudd BD, Lukacs NW. Plasmacytoid dendritic cells inhibit pulmonary immunopathology and promote clearance of respiratory syncytial virus. J Exp Med. 2006;203(5):1153–9.

137. Cormier SA, Shrestha B, Saravia J, Lee GI, Shen L, DeVincenzo JP, et al. Limited Type I Interferons and Plasmacytoid Dendritic Cells during Neonatal Respiratory Syncytial Virus Infection Permit Immunopathogenesis upon Reinfection. J Virol. 2014 Aug;88(16):9350–60.

138. Pulendran B, Banchereau J, Burkeholder S, Kraus E, Guinet E, Chalouni C, et al. Flt3-Ligand and Granulocyte Colony-Stimulating Factor Mobilize Distinct Human Dendritic Cell Subsets In Vivo. J Immunol. 2000 Jul 1;165(1):566–72.

139. Remot A, Descamps D, Jouneau L, Laubreton D, Dubuquoy C, Bouet S, et al. Flt3 ligand improves the innate response to respiratory syncytial virus and limits lung disease upon RSV reexposure in neonate mice. Eur J Immunol. 2016 Apr 1;46(4):874–84.

140. Hijano DR, Siefker DT, Shrestha B, Jaligama S, Vu LD, Tillman H, et al. Type I Interferon Potentiates IgA Immunity to Respiratory Syncytial Virus Infection During Infancy. Sci Rep. 2018;8(1):11034.

141. Van Benten IJ, Van Drunen CM, Koopman LP, KleinJan A, Van Middelkoop BC, De Waal L, et al. RSV-induced bronchiolitis but not upper respiratory tract infection is accompanied by an increased nasal IL-18 response. J Med Virol. 2003 Oct;71(2):290–7.

142. Tabarani CM, Bonville CA, Suryadevara M, Branigan P, Wang D, Huang D, et al. Novel inflammatory markers, clinical risk factors and virus type associated with severe respiratory syncytial virus infection. Pediatr Infect Dis J. 2013 Dec;32(12):e437-42.

143. Choi J, Callaway Z, Kim H Bin, Fujisawa T, Kim CK. The role of TNF-α in eosinophilic inflammation associated with RSV bronchiolitis. Pediatr Allergy Immunol. 2010 May;21(3):474–9.

144. Mbawuike IN, Wells J, Byrd R, Cron SG, Glezen WP, Piedra PA. HLA-Restricted CD8 + Cytotoxic T Lymphocyte, Interferon-γ, and Interleukin-4 Responses to Respiratory Syncytial Virus Infection in Infants and Children . J Infect Dis. 2001 Mar;183(5):687–96.

145. De Weerd W, Twilhaar WN, Kimpen JLL. T cell subset analysis in peripheral blood of children with RSV bronchiolitis. Scand J Infect Dis. 1998;30(1):77–80.

146. Lukens M V., van de Pol AC, Coenjaerts FEJ, Jansen NJG, Kamp VM, Kimpen JLL, et al. A Systemic Neutrophil Response Precedes Robust CD8+ T-Cell Activation during Natural Respiratory Syncytial Virus Infection in Infants. J Virol. 2010 Mar;84(5):2374–83.

147. Bem RA, Bos AP, Bots M, Wolbink AM, Van Ham SM, Medema JP, et al. Activation of the

granzyme pathway in children with severe respiratory syncytial virus infection. Pediatr Res. 2008 Jun;63(6):650–5.

148. Isaacs D, Bangham CRM, Mcmichael AJ. Cell-Mediated Cytotoxic Response To Respiratory Syncytial Virus in Infants With Bronchiolitis. Lancet. 1987 Oct;330(8562):769–71.

149. Jozwik A, Habibi MS, Paras A, Zhu J, Guvenel A, Dhariwal J, et al. RSV-specific airway resident memory CD8+ T cells and differential disease severity after experimental human infection. Nat Commun. 2015 Dec;6(1):10224.

150. Braciale TJ, Sun J, Kim TS. Regulating the adaptive immune response to respiratory virus infection. Nat Rev Immunol. 2012;12(4):295–305.

151. Zhou J, Yang XQ, Fu Z, Zhao XD, Jiang LP, Wang LJ, et al. Increased pathogenesis and inflammation of airways from respiratory syncytial virus infection in T cell deficient nude mice. Med Microbiol Immunol. 2008;197(4):345–51.

152. ALWAN WH, RECORD FM, OPENSHAW PJM. CD4+ T cells clear virus but augment disease in mice infected with respiratory syncytial virus. Comparison with the effects of CD8+ T cells. Clin Exp Immunol. 2008;88(3):527–36.

153. You D, Marr N, Saravia J, Shrestha B, Lee GI, Turvey SE, et al. IL-4Rα on CD4 + T cells plays a pathogenic role in respiratory syncytial virus reinfection in mice infected initially as neonates . J Leukoc Biol. 2013;93(6):933–42.

154. Mcinnes E, Sopp P, Howard CJ, Taylor G. Phenotypic analysis of local cellular responses in calves infected with bovine respiratory syncytial virus. Immunology. 1999;96(3):396–403.

155. Taylor G, Thomas LH, Wyld SG, Furze J, Sopp P, Howard CJ. Role of T-lymphocyte subsets in recovery from respiratory syncytial virus infection in calves. J Virol. 1995;69(11):6658–64.

156. Graham BS, Bunton LA, Wright PF, Karzon DT. Role of T lymphocyte subsets in the pathogenesis of primary infection and rechallenge with respiratory syncytial virus in mice. J Clin Invest. 1991;88(3):1026–33.

157. Schwarze J, Cieslewicz G, Joetham A, Ikemura T, Hamelmann E, Gelfand EW. CD8 T cells are essential in the development of respiratory syncytial virus-induced lung eosinophilia and airway hyperresponsiveness. J Immunol. 1999;162(7):4207–11.

158. Tregoning JS, Yamaguchi Y, Harker J, Wang B, Openshaw PJM. The Role of T Cells in the Enhancement of Respiratory Syncytial Virus Infection Severity during Adult Reinfection of Neonatally Sensitized Mice. J Virol. 2008;82(8):4115–24.

159. Schmidt ME, Knudson CJ, Hartwig SM, Pewe LL, Meyerholz DK, Langlois RA, et al. Memory CD8 T cells mediate severe immunopathology following respiratory syncytial virus infection. PLoS Pathog. 2018 Jan 2;14(1):e1006810–e1006810.

160. Cannon MJ, Openshaw PJM, Askonas BA. Cytotoxic T cells clear virus but augment lung pathology in mice infected with respiratory syncytial virus. J Exp Med. 1988 Sep 1;168(3):1163–8.

161. Thomas LH, Cook RS, Howard CJ, Gaddum RM, Taylor G. Influence of selective T-lymphocyte depletion on the lung pathology of gnotobiotic calves and the distribution of different T-lymphocyte subsets following challenge with bovine respiratory syncytial virus. Res Vet Sci. 1996;61(1):38–44.

162. Woolums AR, Gunther RA, McArthur-Vaughan K, Anderson ML, Omlor A, Boyle GA, et al. Cytotoxic T lymphocyte activity and cytokine expression in calves vaccinated with formalin-

inactivated bovine respiratory syncytial virus prior to challenge. Comp Immunol Microbiol Infect Dis. 2004;27(1):57–74.

163. Gaddum RM, Cook RS, Thomas LH, Taylor G. Primary cytotoxic T-cell responses to bovine respiratory syncytial virus in calves. Immunology. 1996;88(3):421–7.

164. Kinnear E, Lambert L, McDonald JU, Cheeseman HM, Caproni LJ, Tregoning JS. Airway T cells protect against RSV infection in the absence of antibody. Mucosal Immunol. 2018 Jan;11(1):249–56.

165. Bermejo-Martin JF, Garcia-Arevalo MC, De Lejarazu RO, Ardura J, Eiros JM, Alonso A, et al. Predominance of Th2 cytokines, CXC chemokines and innate immunity mediators at the mucosal level during severe respiratory syncytial virus infection in children. Eur Cytokine Netw. 2007 Sep;18(3):162–7.

166. Caballero MT, Serra ME, Acosta PL, Marzec J, Gibbons L, Salim M, et al. TLR4 genotype and environmental LPS mediate RSV bronchiolitis through Th2 polarization. J Clin Invest. 2015 Feb;125(2):571–82.

167. Pino M, Kelvin DJ, Bermejo-Martin JF, Alonso A, Matías V, Tenorio A, et al. Nasopharyngeal aspirate cytokine levels 1 yr after severe respiratory syncytial virus infection. Pediatr Allergy Immunol. 2009 Dec;20(8):791–5.

168. Tawill S, Le Goff L, Ali F, Blaxter M, Allen JE. Both Free-Living and Parasitic Nematodes Induce a Characteristic Th2 Response That Is Dependent on the Presence of Intact Glycans. Infect Immun. 2004 Jan;72(1):398–407.

169. Deo SS, Mistry KJ, Kakade AM, Niphadkar P V. Role played by Th2 type cytokines in IgE mediated allergy and asthma. Lung India. 2010 Apr;27(2):66–71.

170. Mori Y, Iwasaki H, Kohno K, Yoshimoto G, Kikushige Y, Okeda A, et al. Identification of the human eosinophil lineage-committed progenitor: Revision of phenotypic definition of the human common myeloid progenitor. J Exp Med. 2009 Jan;206(1):183–93.

171. Hassan MA, Ezz Eldin AM, Ahmed MM. T-{Helper}2 / {T}-{Helper}1 {Imbalance} in {Respiratory} {Syncytial} {Virus} {Bronchiolitis} in {Relation} {To} {Disease} {Severity} and {Outcome}. 2008.

172. Mella C, Suarez-Arrabal MC, Lopez S, Stephens J, Fernandez S, Hall MW, et al. Innate immune dysfunction is associated with enhanced disease severity in infants with severe respiratory syncytial virus bronchiolitis. J Infect Dis. 2013 Feb;207(4):564–73.

173. Garofalo RP, Patti J, Hintz KA, Hill V, Ogra PL, Welliver RC. Macrophage Inflammatory Protein–1α (Not T Helper Type 2 Cytokines) Is Associated with Severe Forms of Respiratory Syncytial Virus Bronchiolitis. J Infect Dis. 2001 Aug;184(4):393–9.

174. Roe MFE, Bloxham DM, Cowburn AS, O'Donnell DR. Changes in helper lymphocyte chemokine receptor expression and elevation of IP-10 during acute respiratory syncytial virus infection in infants. Pediatr Allergy Immunol. 2011 Mar;22(2):229–34.

175. Chen ZM, Mao JH, Du LZ, Tang YM. Association of cytokine responses with disease severity in infants with respiratory syncytial virus infection. Acta Paediatr Int J Paediatr. 2002;91(9):914–22.

176. Christiaansen AF, Syed MA, Ten Eyck PP, Hartwig SM, Durairaj L, Kamath SS, et al. Altered Treg and cytokine responses in RSV-infected infants. Pediatr Res. 2016 Nov;80(5):702–9.

177. Raiden S, Pandolfi J, Payasliàn F, Anderson M, Rivarola N, Ferrero F, et al. Depletion of

circulating regulatory T cells during severe respiratory syncytial virus infection in young children. Am J Respir Crit Care Med. 2014 Apr;189(7):865–8.

178. Schiering C, Krausgruber T, Chomka A, Fröhlich A, Adelmann K, Wohlfert EA, et al. The alarmin IL-33 promotes regulatory T-cell function in the intestine. Nature. 2014 Sep;513(7519):564–8.

179. Bacharier LB, Coverstone A, Schweiger T, Gregory G, Yin-DeClue H, Sajol G, et al. Regulatory T Cells In Acute Severe Rsv Bronchiolitis. Am J Respir Crit Care Med. 2013;187:no pagination.

180. Fulton RB, Meyerholz DK, Varga SM. Foxp3 + CD4 Regulatory T Cells Limit Pulmonary Immunopathology by Modulating the CD8 T Cell Response during Respiratory Syncytial Virus Infection . J Immunol. 2010;185(4):2382–92.

181. Loebbermann J, Thornton H, Durant L, Sparwasser T, Webster KE, Sprent J, et al. Regulatory T cells expressing granzyme B play a critical role in controlling lung inflammation during acute viral infection. Mucosal Immunol. 2012;5(2):161–72.

182. Ruckwardt TJ, Bonaparte KL, Nason MC, Graham BS. Regulatory T Cells Promote Early Influx of CD8+ T Cells in the Lungs of Respiratory Syncytial Virus-Infected Mice and Diminish Immunodominance Disparities. J Virol. 2009;83(7):3019–28.

183. Weiss KA, Christiaansen AF, Fulton RB, Meyerholz DK, Varga SM. Multiple CD4 + T Cell Subsets Produce Immunomodulatory IL-10 During Respiratory Syncytial Virus Infection . J Immunol. 2011;187(6):3145–54.

184. Durant LR, Makris S, Voorburg CM, Loebbermann J, Johansson C, Openshaw PJM. Regulatory T Cells Prevent Th2 Immune Responses and Pulmonary Eosinophilia during Respiratory Syncytial Virus Infection in Mice. J Virol. 2013;87(20):10946–54.

185. Zhou W, Hashimoto K, Moore ML, Elias JA, Zhu Z, Durbin J, et al. IL-13 is associated with reduced illness and replication in primary respiratory syncytial virus infection in the mouse. Microbes Infect. 2006;8(14–15):2880–9.

186. Lukacs NW, Tekkanat KK, Berlin A, Hogaboam CM, Miller A, Evanoff H, et al. Respiratory syncytial virus predisposes mice to augmented allergic airway responses via IL-13-mediated mechanisms. J Immunol. 2005;175(12):8442.2-8442.

187. Tekkanat KK, Maassab HF, Cho DS, Lai JJ, John A, Berlin A, et al. IL-13-induced airway hyperreactivity during respiratory syncytial virus infection is STAT6 dependent. J Immunol. 2005;175(12):8442.1-8442.

188. Graham BS, Henderson GS, Tang YW, Lu X, Neuzil KM, Colley DG. Priming immunization determines T helper cytokine mRNA expression patterns in lungs of mice challenged with respiratory syncytial virus. J Immunol. 1993 Aug;151(4):2032–40.

189. Connors M, Giese NA, Kulkarni AB, Firestone CY, Morse HC, Murphy BR. Enhanced pulmonary histopathology induced by respiratory syncytial virus (RSV) challenge of formalin-inactivated RSV-immunized BALB/c mice is abrogated by depletion of interleukin-4 (IL-4) and IL-10. J Virol. 1994 Aug;68(8):5321–5.

190. Connors M, Kulkarni AB, Firestone CY, Holmes KL, Morse HC, Sotnikov A V, et al. Pulmonary histopathology induced by respiratory syncytial virus (RSV) challenge of formalin-inactivated RSV-immunized BALB/c mice is abrogated by depletion of CD4+ T cells. J Virol. 1992 Dec;66(12):7444–51.

191. Waris ME, Tsou C, Erdman DD, Zaki SR, Anderson LJ. Respiratory synctial virus infection in

BALB/c mice previously immunized with formalin-inactivated virus induces enhanced pulmonary inflammatory response with a predominant Th2-like cytokine pattern. J Virol. 1996 May;70(5):2852–60.

192. Johnson TR, Parker RA, Johnson JE, Graham BS. IL-13 Is Sufficient for Respiratory Syncytial Virus G Glycoprotein-Induced Eosinophilia After Respiratory Syncytial Virus Challenge. J Immunol. 2003 Feb;170(4):2037–45.

193. Li B, Wu FL, Feng X Bin, Sun DK, Cui QQ, Zhao ZX. [Changes and the clinical significance of CD4+ CD25+ regulatory T cells and Th17 cells in peripheral blood of infants with respiratory syncytial virus bronchiolitis]. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi. 2012 Apr;28(4):426–8.

194. Faber TE, Groen H, Welfing M, Jansen KJG, Bont LJ. Specific increase in local IL-17 production during recovery from primary RSV bronchiolitis. J Med Virol. 2012 Jul;84(7):1084–8.

195. Stoppelenburg AJ, Salimi V, Hennus M, Plantinga M, Huis in 't Veld R, Walk J, et al. Local IL-17A Potentiates Early Neutrophil Recruitment to the Respiratory Tract during Severe RSV Infection. PLoS One. 2013;8(10):e78461.

196. Kallal LE, Hartigan AJ, Hogaboam CM, Schaller MA, Lukacs NW. Inefficient Lymph Node Sensitization during Respiratory Viral Infection Promotes IL-17–Mediated Lung Pathology. J Immunol. 2010;185(7):4137–47.

197. Mukherjee S, Lindell DM, Berlin AA, Morris SB, Shanley TP, Hershenson MB, et al. IL-17Induced pulmonary pathogenesis during respiratory viral infection and exacerbation of allergic disease. Am J Pathol. 2011;179(1):248–58.

198. Pyle CJ, Uwadiae FI, Swieboda DP, Harker JA. Early IL-6 signalling promotes IL-27 dependent maturation of regulatory T cells in the lungs and resolution of viral immunopathology. PLoS Pathog. 2017;13(9).

199. De Almeida Nagata DE, Demoor T, Ptaschinski C, Ting HA, Jang S, Reed M, et al. IL-27R-mediated regulation of IL-17 controls the development of respiratory syncytial virus-associated pathogenesis. Am J Pathol. 2014 Jun;184(6):1807–18.

200. Hashimoto K, Durbin JE, Zhou W, Collins RD, Ho SB, Kolls JK, et al. Respiratory syncytial virus infection in the absence of STAT1 results in airway dysfunction, airway mucus, and augmented IL-17 levels. J Allergy Clin Immunol. 2005;116(3):550–7.

201. Hayday A, Theodoridis E, Ramsburg E, Shires J. Intraepithelial lymphocytes: Exploring the Third Way in immunology. Nat Immunol. 2001 Nov;2(11):997–1003.

202. Aoyagi M, Shimojo N, Sekine K, Nishimuta T, Kohno Y. Respiratory syncytial virus infection suppresses IFN-γ production of γδ T cells. Clin Exp Immunol. 2003 Feb;131(2):312–7.

203. Dodd J, Riffault S, Kodituwakku JS, Hayday AC, Openshaw PJM. Pulmonary Vγ4 + γδ T Cells Have Proinflammatory and Antiviral Effects in Viral Lung Disease . J Immunol. 2009 Jan;182(2):1174–81.

204. Welte T, Aronson J, Gong B, Rachamallu A, Mendell N, Tesh R, et al. Vγ4+ T cells regulate host immune response to West Nile virus infection. FEMS Immunol Med Microbiol. 2011 Nov;63(2):183–92.

205. Suryawanshi A, Veiga-Parga T, Rajasagi NK, Reddy PBJ, Sehrawat S, Sharma S, et al. Role of IL-17 and Th17 Cells in Herpes Simplex Virus-Induced Corneal Immunopathology. J Immunol. 2011 Aug;187(4):1919–30.

206. Mcgill JL, Nonnecke BJ, Lippolis JD, Reinhardt TA, Sacco RE. Differential chemokine and

cytokine production by neonatal bovine γδ T-cell subsets in response to viral toll-like receptor agonists and in vivo respiratory syncytial virus infection. Immunology. 2013 Jun;139(2):227–44.

207. Pawankar RU, Okuda M, Suzuki K, Okumura K, Ra C. Phenotypic and molecular characteristics of nasal mucosal γδ T cells in allergic and infectious rhinitis. Am J Respir Crit Care Med. 1996;153(5):1655–65.

208. Krug N, Erpenbeck VJ, Balke K, Petschallies J, Tschernig T, Hohlfeld JM, et al. Cytokine profile of bronchoalveolar lavage-derived CD4+, CD8+, and γδ T cells in people with asthma after segmental allergen challenge. Am J Respir Cell Mol Biol. 2001 Jul;25(1):125–31.

209. Welliver TP, Garofalo RP, Hosakote Y, Hintz KH, Avendano L, Sanchez K, et al. Severe Human Lower Respiratory Tract Illness Caused by Respiratory Syncytial Virus and Influenza Virus Is Characterized by the Absence of Pulmonary Cytotoxic Lymphocyte Responses. J Infect Dis. 2007 Apr;195(8):1126–36.

210. Abu-Harb M, Bell F, Finn A, Rao WH, Nixon L, Shale D, et al. IL-8 and neutrophil elastase levels in the respiratory tract of infants with RSV bronchiolitis. Eur Respir J. 1999 Jul;14(1):139–43.

211. Wang SZ, Smith PK, Lovejoy M, Bowden JJ, Alpers JH, Forsyth KD. The apoptosis of neutrophils is accelerated in respiratory syncytial virus (RSV)-induced bronchiolitis. Clin Exp Immunol. 1998 Oct;114(1):49–54.

212. Wang SZE, Smith PK, Lovejoy M, Bowden JJ, Alpers JH, Forsyth KD. Shedding of L-selectin and PECAM-1 and upregulation of Mac-1 and ICAM-1 on neutrophils in RSV bronchiolitis. Am J Physiol - Lung Cell Mol Physiol. 1998 Nov 1;275(5 19-5):L983–9.

213. Yui I, Hoshi A, Shigeta Y, Takami T, Nakayama T. Detection of human respiratory syncytial virus sequences in peripheral blood mononuclear cells. J Med Virol. 2003 Jul;70(3):481–9.

214. Halfhide CP, Flanagan BF, Brearey SP, Hunt JA, Fonceca AM, McNamara PS, et al. Respiratory syncytial virus binds and undergoes transcription in neutrophils from the blood and airways of infants with severe bronchiolitis. J Infect Dis. 2011 Aug 1;204(3):451–8.

215. Saitoh T, Komano J, Saitoh Y, Misawa T, Takahama M, Kozaki T, et al. Neutrophil extracellular traps mediate a host defense response to human immunodeficiency virus-1. Cell Host Microbe. 2012 Jul;12(1):109–16.

216. Li RHL, Tablin F. A comparative review of neutrophil extracellular traps in sepsis. Vol. 5, Frontiers in Veterinary Science. 2018. p. 291.

217. Hoeksema M, Tripathi S, White M, Qi L, Taubenberger J, Van Eijk M, et al. Arginine-rich histones have strong antiviral activity for influenza A viruses. Innate Immun. 2015 Oct;21(7):736–45.

218. Tamura M, Natori K, Kobayashi M, Miyamura T, Takeda N. Inhibition of attachment of virions of Norwalk virus to mammalian cells by soluble histone molecules. Arch Virol. 2003 Sep;148(9):1659–70.

219. Kirsebom FCM, Kausar F, Nuriev R, Makris S, Johansson C. Neutrophil recruitment and activation are differentially dependent on MyD88/TRIF and MAVS signaling during RSV infection. Mucosal Immunol. 2019;12(5):1244–55.

220. Bhoj VG, Sun Q, Bhoj EJ, Somers C, Chen X, Torres JP, et al. MAVS and MyD88 are essential for innate immunity but not cytotoxic T lymphocyte response against respiratory syncytial virus. Proc Natl Acad Sci U S A. 2008;105(37):14046–51.

221. Hsu DH, De Waal Malefyt R, Fiorentino DF, Dang MN, Vieira P, Devries J, et al. Expression of interleukin-10 activity by Epstein-Barr virus protein BCRF1. Science (80-). 1990 Nov;250(4982):830–2.

222. Kotenko S V., Saccani S, Izotova LS, Mirochnitchenko O V., Pestka S. Human cytomegalovirus harbors its own unique IL-10 homolog (cmvIL-10). Proc Natl Acad Sci U S A. 2000 Feb;97(4):1695–700.

223. Urban CF, Ermert D, Schmid M, Abu-Abed U, Goosmann C, Nacken W, et al. Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against Candida albicans. PLoS Pathog. 2009 Oct;5(10):e1000639.

224. Currie SM, Gwyer Findlay E, McFarlane AJ, Fitch PM, Böttcher B, Colegrave N, et al. Cathelicidins Have Direct Antiviral Activity against Respiratory Syncytial Virus In Vitro and Protective Function In Vivo in Mice and Humans. J Immunol. 2016 Mar 15;196(6):2699–710.

225. Lindemans CA, Kimpen JLL, Luijk B, Heidema J, Kanters D, Van Der Ent CK, et al. Systemic eosinophil response induced by respiratory syncytial virus. Clin Exp Immunol. 2006 Jun;144(3):409–17.

226. Kawasaki Y, Hosoya M, Kanno H, Suzuki H. Serum regulated upon activation, normal T cell expressed and presumably secreted concentrations and eosinophils in respiratory syncytial virus infection. Pediatr Int. 2006 Jun;48(3):257–60.

227. Hyun HK, Mi HL, Joon SL. Eosinophil cationic protein and chemokines in nasopharyngeal secretions of infants with respiratory syncytial virus (RSV) bronchiolitis and non-RSV bronchiolitis. J Korean Med Sci. 2007 Feb;22(1):37–42.

228. Harrison AM, Bonville CA, Rosenberg HF, Domachowske JB. Respiratory syncytical virus-induced chemokine expression in the lower airways: Eosinophil recruitment and degranulation. Am J Respir Crit Care Med. 1999 Jun;159(6):1918–24.

229. Dimova-Yaneva D, Russell D, Main M, Brooker RJ, Helms PJ. Eosinophil activation and cysteinyl leukotriene production in infants with respiratory syncytial virus bronchiolitis. Clin Exp Allergy. 2004 Apr;34(4):555–8.

230. Chung HL, Kim SG. RANTES may be predictive of later recurrent wheezing after respiratory syncytial virus bronchiolitis in infants. Ann Allergy, Asthma Immunol. 2002 May;88(5):463–7.

231. Schwarze J, Cieslewicz G, Hamelmann E, Joetham A, Shultz LD, Lamers MC, et al. IL-5 and eosinophils are essential for the development of airway hyperresponsiveness following acute respiratory syncytial virus infection. J Immunol. 1999;162(5):2997–3004.

232. Schwarze J, Mäkelä M, Cieslewicz G, Dakhama A, Lahn M, Ikemura T, et al. Transfer of the enhancing effect of respiratory syncytial virus infection on subsequent allergic airway sensitization by T lymphocytes. J Immunol. 1999;163(10):5729–34.

233. Welliver TP, Garofalo RP, Hosakote Y, Hintz KH, Avendano L, Sanchez K, et al. Severe Human Lower Respiratory Tract Illness Caused by Respiratory Syncytial Virus and Influenza Virus Is Characterized by the Absence of Pulmonary Cytotoxic Lymphocyte Responses. J Infect Dis. 2007 Apr;195(8):1126–36.

234. Larrañaga CL, Ampuero SL, Luchsinger VF, Carrión FA, Aguilar N V., Morales PR, et al. Impaired immune response in severe human lower tract respiratory infection by respiratory syncytial virus. Pediatr Infect Dis J. 2009 Oct;28(10):867–73.

235. Noyola DE, Juárez-Vega G, Monjarás-Ávila C, Escalante-Padrón F, Rangel-Ramírez V, Cadena-

Mota S, et al. NK cell immunophenotypic and genotypic analysis of infants with severe respiratory syncytial virus infection. Microbiol Immunol. 2015 Jul;59(7):389–97.

236. Barkal AA, Weiskopf K, Kao KS, Gordon SR, Rosental B, Yiu YY, et al. Engagement of MHC class i by the inhibitory receptor LILRB1 suppresses macrophages and is a target of cancer immunotherapy article. Nat Immunol. 2018;19(1):76–84.

237. Ederveen THA, Ferwerda G, Ahout IM, Vissers M, de Groot R, Boekhorst J, et al. Haemophilus is overrepresented in the nasopharynx of infants hospitalized with RSV infection and associated with increased viral load and enhanced mucosal CXCL8 responses. Microbiome. 2018 Jan;6(1):10.

238. De Steenhuijsen Piters WAA, Heinonen S, Hasrat R, Bunsow E, Smith B, Suarez-Arrabal MC, et al. Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. Am J Respir Crit Care Med. 2016 Nov 1;194(9):1104–15.

239. Sande CJ, Njunge JM, Mwongeli Ngoi J, Mutunga MN, Chege T, Gicheru ET, et al. Airway response to respiratory syncytial virus has incidental antibacterial effects. Nat Commun. 2019;10(1):2218.

240. Hijano DR, Vu LD, Kauvar LM, Tripp RA, Polack FP, Cormier SA. Role of type I interferon (IFN) in the respiratory syncytial virus (RSV) immune response and disease severity. Front Immunol. 2019;10(MAR):566.

241. Hillyer P, Mane VP, Chen A, dos Santos MB, Schramm LM, Shepard RE, et al. Respiratory syncytial virus infection induces a subset of types I and III interferons in human dendritic cells. Virology. 2017 Apr;504:63–72.

242. Hall CB, Douglas RG, Simons RL, Geiman JM. Interferon production in children with respiratory syncytial, influenza, and parainfluenza virus infections. J Pediatr. 1978 Jul;93(1):28–32.

243. McIntosh K. Interferon in nasal secretions from infants with viral respiratory tract infections. J Pediatr. 1978 Jul;93(1):33–6.

244. Taylor CE, Webb MSC, Milner AD, Milner PD, Morgan LA, Scott R, et al. Interferon alfa-infectious virus, and virus antigen secretion in respiratory syncytial virus infections of graded severity. Arch Dis Child. 1989 Dec;64(12):1656–60.

245. Marr N, Wang T-I, Kam SHY, Hu YS, Sharma AA, Lam A, et al. Attenuation of Respiratory Syncytial Virus–Induced and RIG-I–Dependent Type I IFN Responses in Human Neonates and Very Young Children. J Immunol. 2014 Feb;192(3):948–57.

246. Sung RYT, Yin J, Oppenheimer SJ, Tam JS, Lau J. Treatment of respiratory sycytial virus infection with recombinant interferon alfa-2A. Arch Dis Child. 1993 Oct;69(4):440–2.

247. Sriskandan K, Garner P, Watkinson J, Pettingale KW, Brinkley D, Calman FMB, et al. A toxicity study of recombinant interferon-gamma given by intravenous infusion to patients with advanced cancer. Cancer Chemother Pharmacol. 1986;18(1):63–8.

248. Ye Q, Shao WX, Shang SQ, Pan YX, Shen HQ, Chen XJ. Epidemiological characteristics and immune status of children with Respiratory Syncytial Virus. J Med Virol. 2015 Feb;87(2):323–9.

249. Vu LD, Siefker D, Jones TL, You D, Taylor R, DeVincenzo J, et al. Elevated levels of type 2 respiratory innate lymphoid cells in human infants with severe respiratory syncytial virus bronchiolitis. Am J Respir Crit Care Med. 2019 Jun;200(11):1414–23.

250. Stier MT, Bloodworth MH, Toki S, Newcomb DC, Goleniewska K, Boyd KL, et al. Respiratory syncytial virus infection activates IL-13–producing group 2 innate lymphoid cells through thymic stromal lymphopoietin. J Allergy Clin Immunol. 2016 Sep;138(3):814-824.e11.

251. Salter BM, Aw M, Sehmi R. The role of type 2 innate lymphoid cells in eosinophilic asthma. J Leukoc Biol. 2019 May 1;106(4):889–901.

252. Bera MM, Lu B, Martin TR, Cui S, Rhein LM, Gerard C, et al. Th17 Cytokines Are Critical for Respiratory Syncytial Virus-Associated Airway Hyperreponsiveness through Regulation by Complement C3a and Tachykinins. J Immunol. 2011 Oct 15;187(8):4245–55.

253. Kolli D, Yueqing Z, Palkowetz K, Garofalo R, Casola A. Critical role of Neutrophils in Respiratory Syncytial Virus (RSV) induced disease pathogenesis (INC8P.440). J Immunol. 2014 May 1;192(1 Supplement):187.13 LP-187.13.

254. Makris S, Bajorek M, Culley FJ, Goritzka M, Johansson C. Alveolar macrophages can control respiratory syncytial virus infection in the absence of Type I interferons. J Innate Immun. 2016/07/16. 2016;8(5):452–63.

255. Loo YM, Gale M. Immune Signaling by RIG-I-like Receptors. Immunity. 2011 May 27;34(5):680–92.

256. Marr N, Turvey SE, Grandvaux N. Pathogen recognition receptor crosstalk in respiratory syncytial virus sensing: A host and cell type perspective. Trends Microbiol. 2013/10/09. 2013 Nov;21(11):568–74.

257. Liu P, Jamaluddin M, Li K, Garofalo RP, Casola A, Brasier AR. Retinoic Acid-Inducible Gene I Mediates Early Antiviral Response and Toll-Like Receptor 3 Expression in Respiratory Syncytial Virus-Infected Airway Epithelial Cells. J Virol. 2007 Feb;81(3):1401–11.

258. Kurt-Jones EA, Popova L, Kwinn L, Haynes LM, Jones LP, Tripp RA, et al. Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus. Nat Immunol. 2000;1(5):398–401.

259. Shirey KA, Pletneva LM, Puche AC, Keegan AD, Prince GA, Blanco JCG, et al. Control of RSV-induced lung injury by alternatively activated macrophages is IL-4Rα-, TLR4-, and IFN-B-dependent. Mucosal Immunol. 2010;3(3):291–300.

260. Tsutsumi H, Matsuda K, Sone S, Takeuchi R, Chiba S. Respiratory syncytial virus-induced cytokine production by neonatal macrophages. Clin Exp Immunol. 1996 Dec;106(3):442–6.

261. Veremeyko T, Siddiqui S, Sotnikov I, Yung A, Ponomarev ED. IL-4/IL-13-dependent and independent expression of miR-124 and its contribution to M2 phenotype of monocytic cells in normal conditions and during allergic inflammation. PLoS One. 2013;8(12):e81774.

262. Orecchioni M, Ghosheh Y, Pramod AB, Ley K. Corrigendum: Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS–) vs. Alternatively Activated Macrophages (Frontiers in Immunology, (2019), 10, (1084), 10.3389/fimmu.2019.01084). Vol. 11, Frontiers in Immunology. 2020. p. 1084.

263. Mantovani A, Sica A, Sozzani S, Allavena P, Vecchi A, Locati M. The chemokine system in diverse forms of macrophage activation and polarization. Trends Immunol. 2004 Dec;25(12):677–86.

264. Mills CD, Kincaid K, Alt JM, Heilman MJ, Hill AM. M-1/M-2 Macrophages and the Th1/Th2 Paradigm. J Immunol. 2000 Jun;164(12):6166–73.

265. Hong JE, Kye YC, Park SM, Cheon IS, Chu H, Park BC, et al. Alveolar macrophages treated with

Bacillus subtilis spore protect mice infected with respiratory syncytial virus A2. Front Microbiol. 2019 Mar 12;10(MAR):447.

266. Fernando MR, Reyes JL, Iannuzzi J, Leung G, McKay DM. The pro-inflammatory cytokine, interleukin-6, enhances the polarization of alternatively activated macrophages. PLoS One. 2014;9(4):e94188.

267. Rallabhandi P, Phillips RL, Boukhvalova MS, Pletneva LM, Shirey KA, Gioannini TL, et al. Respiratory syncytial virus fusion protein-induced toll-like receptor 4 (TLR4) signaling is inhibited by the TLR4 antagonists rhodobacter sphaeroides lipopolysaccharide and eritoran (E5564) and requires direct interaction with MD-2. MBio. 2012;3(4).

268. Marr N, Turvey SE. Role of human TLR4 in respiratory syncytial virus-induced NF-κB activation, viral entry and replication. Innate Immun. 2012 Dec;18(6):856–65.

269. Kawai T, Akira S. Signaling to NF-κB by Toll-like receptors. Trends Mol Med. 2007 Nov;13(11):460–9.

270. Hosakote YM, Brasier AR, Casola A, Garofalo RP, Kurosky A. Respiratory Syncytial Virus Infection Triggers Epithelial HMGB1 Release as a Damage-Associated Molecular Pattern Promoting a Monocytic Inflammatory Response. J Virol. 2016 Nov;90(21):9618–31.

271. Foronjy RF, Ochieng PO, Salathe MA, Dabo AJ, Eden E, Baumlin N, et al. Protein tyrosine phosphatase 1B negatively regulates S100A9-mediated lung damage during respiratory syncytial virus exacerbations. Mucosal Immunol. 2016/01/27. 2016 Sep;9(5):1317–29.

272. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: Update on toll-like receptors. Nat Immunol. 2010 May;11(5):373–84.

273. Rudd BD, Schaller MA, Smit JJ, Kunkel SL, Neupane R, Kelley L, et al. MyD88-Mediated Instructive Signals in Dendritic Cells Regulate Pulmonary Immune Responses during Respiratory Virus Infection. J Immunol. 2007;178(9):5820–7.

274. Wang H, Peters N, Schwarze J. Plasmacytoid Dendritic Cells Limit Viral Replication, Pulmonary Inflammation, and Airway Hyperresponsiveness in Respiratory Syncytial Virus Infection. J Immunol. 2006;177(9):6263–70.

275. Brand HK, Ahout IML, De Ridder D, Van Diepen A, Li Y, Zaalberg M, et al. Olfactomedin 4 serves as a marker for disease severity in pediatric respiratory syncytial virus (RSV) infection. PLoS One. 2015 Jul 10;10(7):e0131927–e0131927.

276. Mejias A, Dimo B, Suarez NM, Garcia C, Suarez-Arrabal MC, Jartti T, et al. Whole Blood Gene Expression Profiles to Assess Pathogenesis and Disease Severity in Infants with Respiratory Syncytial Virus Infection. PLoS Med. 2013/11/12. 2013 Nov;10(11):e1001549–e1001549.

277. Sonawane AR, Tian L, Chu CY, Qiu X, Wang L, Holden-Wiltse J, et al. Microbiome-Transcriptome Interactions Related to Severity of Respiratory Syncytial Virus Infection. Sci Rep. 2019;9(1):13824.

278. Jong VL, Ahout IML, Van Den Ham HJ, Jans J, Zaaraoui-Boutahar F, Zomer A, et al. Transcriptome assists prognosis of disease severity in respiratory syncytial virus infected infants. Sci Rep. 2016;6(1):36603.

279. Caserta MT, Wang L, Chu C-Y, Slaunwhite C, Holden-Wiltse J, Corbett A, et al. 415. Airway Gene-Expression Classifiers for Respiratory Syncytial Virus (RSV) Disease Severity in Infants. Open Forum Infect Dis. 2019 Oct 23;6(Supplement_2):S210–S210.

280. Fjaerli H, Bukholm G, Skjaeret C, Holden M, Nakstad B. Cord Blood Gene Expression in Infants

Hospitalized with Respiratory Syncytial Virus Bronchiolitis. J Infect Dis. 2007 Aug;196(3):394–404.

281. Fjaerli HO, Bukholm G, Krog A, Skjaeret C, Holden M, Nakstad B. Whole blood gene expression in infants with respiratory syncytial virus bronchiolitis. BMC Infect Dis. 2006 Dec;6:175.

282. Herberg JA, Kaforou M, Gormley S, Sumner ER, Patel S, Jones KDJ, et al. Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. J Infect Dis. 2013 Jul;208(10):1664–8.

283. Van Den Kieboom CH, Ahout IML, Zomer A, Brand KH, De Groot R, Ferwerda G, et al. Nasopharyngeal gene expression, a novel approach to study the course of respiratory syncytial virus infection. Eur Respir J. 2015 Mar 1;45(3):718–25.

284. Do LAH, Pellet J, Rogier Van Doorn H, Tran AT, Nguyen BH, Tran TTL, et al. Host transcription profle in nasal epithelium and whole blood of hospitalized children under 2 years of age with respiratory syncytial virus infection. J Infect Dis. 2018 Dec;217(1):134–46.

285. Heinonen S, Velazquez VM, Ye F, Mertz S, Acero-Bedoya S, Smith B, et al. Immune profiles provide insights into respiratory syncytial virus disease severity in young children. Sci Transl Med. 2020 Apr 22;12(540):eaaw0268.

286. Caserta MT, Qiu X, Tesini B, Wang L, Murphy A, Corbett A, et al. Development of a global respiratory severity score for respiratory syncytial virus infection in infants. J Infect Dis. 2017 Mar;215(5):750–6.

287. Zuccolo J, Deng L, Unruh TL, Sanyal R, Bau JA, Storek J, et al. Expression of MS4A and TMEM176 genes in human B lymphocyte. Vol. 4, Frontiers in Immunology. 2013. p. 195.

288. Schölkopf B. SVMs - A practical consequence of learning theory. IEEE Intell Syst Their Appl. 1998 Jul;13(4):18–21.

289. Henrickson SE, Manne S, Dolfi D V., Mansfield KD, Parkhouse K, Mistry RD, et al. Genomic Circuitry Underlying Immunological Response to Pediatric Acute Respiratory Infection. Cell Rep. 2018 Jan;22(2):411–26.

290. Zhai Y, Franco LM, Atmar RL, Quarles JM, Arden N, Bucasas KL, et al. Host Transcriptional Response to Influenza and Other Acute Respiratory Viral Infections – A Prospective Cohort Study. PLoS Pathog. 2015 Jun;11(6):e1004869.

291. Justicia-Grande AJ, Pardo-Seco J, Cebey-López M, Vilanova-Trillo L, Gómez-Carballa A, Rivero-Calle I, et al. Development and validation of a new clinical scale for infants with acute respiratory infection: The resvinet scale. PLoS One. 2016;11(6):e0157665.

292. Liu W, Yan M, Liu Y, Wang R, Li C, Deng C, et al. Olfactomedin 4 down-regulates innate immunity against Helicobacter pylori infection. Proc Natl Acad Sci. 2010 Jun 15;107(24):11056 LP – 11061.

293. Gauthier M, Lebel MH. Use of ribavirin in the treatment of respiratory syncytial virus infection. Can J Infect Dis. 1994 Sep 1;5(5):239–40.

294. Smart KA, Lanctôt KL, Paes BA. Rebuttal: Palivizumab for the prevention of respiratory syncytial virus infection. Can Fam Physician. 2010 Aug;56(10):988–90.

295. Sanchez-Luna M, Burgos-Pol R, Oyagüez I, Figueras-Aloy J, Sánchez-Solís M, Martinón-Torres F, et al. Cost-utility analysis of Palivizumab for Respiratory Syncytial Virus infection prophylaxis in preterm infants: Update based on the clinical evidence in Spain. BMC Infect

Dis. 2017;17(1):687.

296. Wegzyn C, Toh LK, Notario G, Biguenet S, Unnebrink K, Park C, et al. Safety and Effectiveness of Palivizumab in Children at High Risk of Serious Disease Due to Respiratory Syncytial Virus Infection: A Systematic Review. Infect Dis Ther. 2014/10/09. 2014 Dec;3(2):133–58.

297. Jung JW. Respiratory syncytial virus infection in children with congenital heart disease: Global data and interim results of Korean RSV-CHD survey. Korean J Pediatr. 2011/05/31. 2011 May;54(5):192–6.

298. Kim HW, Canchola JG, Brandt CD, Pyles G, Chanock RM, Jensen K, et al. Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. Am J Epidemiol. 1969 Apr;89(4):422–34.

299. Prince GA, Curtis SJ, Yim KC, Porter DD. Vaccine-enhanced respiratory syncytial virus disease in cotton rats following immunization with Lot 100 or a newly prepared reference vaccine. J Gen Virol. 2001 Dec;82(12):2881–8.

300. Delgado MF, Coviello S, Monsalvo AC, Melendi GA, Hernandez JZ, Batalle JP, et al. Lack of antibody affinity maturation due to poor Toll-like receptor stimulation leads to enhanced respiratory syncytial virus disease. Nat Med. 2009;15(1):34–41.

301. Srikiatkhachorn A, Braciale TJ. Virus-specific CD8+ T lymphocytes downregulate T helper cell type 2 cytokine secretion and pulmonary eosinophilia during experimental murine respiratory syncytial virus infection. J Exp Med. 1997 Aug 4;186(3):421–32.

302. Polack FP, Hoffman SJ, Crujeiras G, Griffin DE. A role for nonprotective complement-fixing antibodies with low avidity for measles virus in atypical measles. Nat Med. 2003;9(9):1209–13.

303. Polack FP, Teng MN, Collins PL, Prince GA, Exner M, Regele H, et al. A role for immune complexes in enhanced respiratory syncytial virus disease. J Exp Med. 2002 Sep 16;196(6):859–65.

304. Devey ME, Steward MW. The induction of chronic antigen-antibody complex disease in selectively bred mice producing either high or low affinity antibody to protein antigens. Immunology. 1980 Oct;41(2):303–11.

305. Acosta PL, Caballero MT, Polack FP. Brief History and Characterization of Enhanced Respiratory Syncytial Virus Disease. Papasian CJ, editor. Clin Vaccine Immunol. 2016 Mar 1;23(3):189–95.

306. Amirthalingam G, Andrews N, Campbell H, Ribeiro S, Kara E, Donegan K, et al. Effectiveness of maternal pertussis vaccination in England: An observational study. Lancet. 2014 Oct 25;384(9953):1521–8.

307. Jepsen MT, Trebbien R, Emborg HD, Krause TG, Schønning K, Voldstedlund M, et al. Incidence and seasonality of respiratory syncytial virus hospitalisations in young children in Denmark, 2010 to 2015. Eurosurveillance. 2018 Jan;23(3):17–163.

308. Eberhardt CS, Blanchard-Rohner G, Lemaître B, Combescure C, Othenin-Girard V, Chilin A, et al. Pertussis antibody transfer to preterm neonates after second-versus third-trimester maternal immunization. Clin Infect Dis. 2017 Apr;64(8):1129–32.

309. Muňoz FM, Swamy GK, Hickman SP, Agrawal S, Piedra PA, Glenn GM, et al. Safety and Immunogenicity of a Respiratory Syncytial Virus Fusion (F) Protein Nanoparticle Vaccine in Healthy Third-Trimester Pregnant Women and Their Infants. J Infect Dis. 2019 Aug

12;220(11):1802–15.

310. Ventola CL. Immunization in the United States: Recommendations, barriers, and measures to improve compliance. P T. 2016 Jul;41(8):426–36.

311. Barrios C, Brawand P, Berney M, Brandt C, Lambert PH, Siegrist CA. Neonatal and early life immune responses to various forms of vaccine antigens qualitatively differ from adult responses: Predominance of a Th2-biased pattern which persists after adult boosting. Eur J Immunol. 1996 Jul;26(7):1489–96.

312. Fries L, Shinde V, Stoddard JJ, Thomas DN, Kpamegan E, Lu H, et al. Immunogenicity and safety of a respiratory syncytial virus fusion protein (RSV F) nanoparticle vaccine in older adults. Immun Ageing. 2017;14(1):8.

313. Beran J, Lickliter JD, Schwarz TF, Johnson C, Chu L, Domachowske JB, et al. Safety and Immunogenicity of 3 Formulations of an Investigational Respiratory Syncytial Virus Vaccine in Nonpregnant Women: Results from 2 Phase 2 Trials. J Infect Dis. 2018 Apr 23;217(10):1616–25.

314. Schmoele-Thoma B, Falsey AR, Walsh EE, Swanson K, Zareba A, Cooper D, et al. 2755. Phase 1/2, First-in-Human Study of the Safety, Tolerability, and Immunogenicity of an RSV Prefusion F-Based Subunit Vaccine Candidate. Open Forum Infect Dis. 2019 Oct 23;6(Supplement_2):S970–S970.

315. Leroux-Roels G, De Boever F, Maes C, Nguyen TLA, Baker S, Gonzalez Lopez A. Safety and immunogenicity of a respiratory syncytial virus fusion glycoprotein F subunit vaccine in healthy adults: Results of a phase 1, randomized, observer-blind, controlled, dosage-escalation study. Vaccine. 2019 May;37(20):2694–703.

316. Zhang S, Zhao G, Su C, Li C, Zhou X, Zhao W, et al. Neonatal priming and infancy boosting with a novel respiratory syncytial virus vaccine induces protective immune responses without concomitant respiratory disease upon RSV challenge. Hum Vaccines Immunother. 2020 Mar;16(3):664–72.

317. Langley JM, Macdonald LD, Weir GM, Mackinnon-Cameron D, Ye L, Mcneil S, et al. A respiratory syncytial virus vaccine based on the small hydrophobic protein ectodomain presented with a novel lipid-based formulation is highly immunogenic and safe in adults: A first-in-humans study. J Infect Dis. 2018 Jul 2;218(3):378–87.

318. Vartak A, Sucheck SJ. Recent advances in subunit vaccine carriers. Vaccines. 2016 Apr 19;4(2):12.

319. Lauer KB, Borrow R, Blanchard TJ. Multivalent and multipathogen viral vector vaccines. Papasian CJ, editor. Clin Vaccine Immunol. 2017 Jan 1;24(1):e00298-16.

320. DeVincenzo J, Gymnopoulou E, De Paepe E, Murray B, Rosemary Bastian A, Haazen W, et al. 902. A Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy of a Single Immunization of Ad26.RSV.preF against RSV Infection in a Viral Challenge Model in Healthy Adults. Open Forum Infect Dis. 2019 Oct 23;6(Supplement_2):S27–8.

321. Samy N, Reichhardt D, Schmidt D, Chen LM, Silbernagl G, Vidojkovic S, et al. Safety and immunogenicity of novel modified vaccinia Ankara-vectored RSV vaccine: A randomized phase I clinical trial. Vaccine. 2020;38(11):2608–19.

322. Cicconi P, Jones C, Sarkar E, Silva-Reyes L, Klenerman P, de Lara C, et al. First-in-Human Randomized Study to Assess the Safety and Immunogenicity of an Investigational Respiratory Syncytial Virus (RSV) Vaccine Based on Chimpanzee-Adenovirus-155 Viral Vector-Expressing

RSV Fusion, Nucleocapsid, and Antitermination Viral Proteins. Clin Infect Dis. 2020 May;70(10):2073–81.

323. Geall AJ, Verma A, Otten GR, Shaw CA, Hekele A, Banerjee K, et al. Nonviral delivery of self-amplifying RNA vaccines. Proc Natl Acad Sci U S A. 2012 Sep 4;109(36):14604–9.

324. Espeseth AS, Cejas PJ, Citron MP, Wang D, DiStefano DJ, Callahan C, et al. Modified mRNA/lipid nanoparticle-based vaccines expressing respiratory syncytial virus F protein variants are immunogenic and protective in rodent models of RSV infection. npj Vaccines. 2020 Feb 14;5(1):16.

325. Cunningham CK, Karron R, Muresan P, McFarland EJ, Perlowski C, Libous J, et al. Live-attenuated respiratory syncytial virus vaccine with deletion of RNA synthesis regulatory protein M2-2 and cold passage mutations is overattenuated. Open Forum Infect Dis. 2019 May 6;6(6).

326. Karron RA, Luongo C, Mateo JS, Wanionek K, Collins PL, Buchholz UJ. Safety and Immunogenicity of the Respiratory Syncytial Virus Vaccine RSV/ΔNS2/Δ1313/I1314L in RSV-Seronegative Children. J Infect Dis. 2020 Oct 12;222(1):82–91.

327. Luongo C, Winter CC, Collins PL, Buchholz UJ. Respiratory Syncytial Virus Modified by Deletions of the NS2 Gene and Amino Acid S1313 of the L Polymerase Protein Is a Temperature-Sensitive, Live-Attenuated Vaccine Candidate That Is Phenotypically Stable at Physiological Temperature. J Virol. 2012/12/12. 2013 Feb;87(4):1985–96.

328. Rey-Jurado E, Soto J, Gálvez N, Kalergis AM. A safe and efficient BCG vectored vaccine to prevent the disease caused by the human Respiratory Syncytial Virus. Hum Vaccines Immunother. 2017 Sep;13(9):2092–7.

329. Domachowske JB, Khan AA, Esser MT, Jensen K, Takas T, Villafana T, et al. Safety, Tolerability and Pharmacokinetics of MEDI8897, an Extended Half-life Single-dose Respiratory Syncytial Virus Prefusion F-targeting Monoclonal Antibody Administered as a Single Dose to Healthy Preterm Infants. Pediatr Infect Dis J. 2018 Sep;37(9):886–92.

330. Pamela Griffin M, Yuan Y, Takas T, DeVincenzo J, Domachowske JB, Simoes EA, et al. 901. MEDI8897 Prevents Serious RSV Disease in Healthy Preterm Infants. Open Forum Infect Dis. 2019 Oct 23;6(Supplement_2):S27–S27.

331. Maas B, Aliprantis A, Wolford D, Fayad G, Vora K, Geng D, et al. 1384. RSV Monoclonal Antibody (MK-1654) Phase 1 Pharmacokinetics (PK) in Healthy Adults and Population PK Modeling to Support Pediatric Development. Open Forum Infect Dis. 2018 Nov 26;5(suppl_1):S424–5.

332. Herberg JA, Kaforou M, Wright VJ, Shailes H, Eleftherohorinou H, Hoggart CJ, et al. Diagnostic test accuracy of a 2-transcript host RNA signature for discriminating bacterial vs viral infection in febrile children. JAMA - J Am Med Assoc. 2016 Aug 23;316(8):835–45.

333. Finotello F, Di Camillo B. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. Brief Funct Genomics. 2014 Sep 18;14(2):130–42.

334. Tarca AL, Romero R, Draghici S. Analysis of microarray experiments of gene expression profiling. Am J Obstet Gynecol. 2006 Aug;195(2):373–88.

335. Govindarajan R, Duraiyan J, Kaliyappan K, Palanisamy M. Microarray and its applications. J Pharm Bioallied Sci. 2012 Aug;4(Suppl 2):S310–2.

336. Routledge R. Fisher's Exact Test. Encyclopedia of Biostatistics. 2005. (Major Reference

Works).

337. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics - A bioconductor package for quality assessment of microarray data. Bioinformatics. 2009 Feb;25(3):415–6.

338. Klaus B, Reisenauer S. An end to end workflow for differential gene expression using Affymetrix microarrays. F1000Research. 2016 Jun 15;5:1384.

339. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 2002 Feb 15;30(4):e15–e15.

340. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003 Apr 1;4(2):249–64.

341. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010 Oct;26(19):2363–7.

342. Du P, Kibbe WA, Lin SM. lumi: A pipeline for processing Illumina microarray. Bioinformatics. 2008 Jul;24(13):1547–8.

343. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr;43(7):e47.

344. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011/05/05. 2011 Jun 15;27(12):1739–40.

345. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545–50.

346. Jolliffe IT. Principal Component Analysis and Factor Analysis. In: Jolliffe IT, editor. New York, NY: Springer New York; 1986. p. 115–28.

347. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol. 2017 Nov 3;13(11):e1005752.

348. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773–82.

349. Valero-Mora PM. ggplot2: Elegant Graphics for Data Analysis . 2nd ed. Vol. 35, Journal of Statistical Software. Springer Publishing Company, Incorporated; 2010.

350. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2006 Apr;8(1):118–27.

351. P B. pvca: Principal Variance Component Analysis (PVCA). 2020.

352. Li S. Regulation of Ribosomal Proteins on Viral Infection. Cells. 2019 May 27;8(5):508.

353. Lacy P. Mechanisms of Degranulation in Neutrophils. Allergy, Asthma Clin Immunol. 2006/09/15. 2006 Sep 15;2(3):98.

354. Semerad CL, Liu F, Gregory AD, Stumpf K, Link DC. G-CSF is an essential regulator of neutrophil trafficking from the bone marrow to the blood. Immunity. 2002 Oct;17(4):413–23.

355. Rosebeck S, Leaman DW. Mitochondrial localization and pro-apoptotic effects of the interferon-inducible protein ISG12a. Apoptosis. 2008 Apr;13(4):562–72.

356. Lin M, Jackson P, Tester AM, Diaconu E, Overall CM, Blalock JE, et al. Matrix metalloproteinase-8 facilitates neutrophil migration through the corneal stromal matrix by collagen degradation and production of the chemotactic peptide Pro-Gly-Pro. Am J Pathol. 2008/06/13. 2008 Jul;173(1):144–53.

357. Welin A, Amirbeagi F, Christenson K, Björkman L, Björnsdottir H, Forsman H, et al. The human neutrophil subsets defined by the presence or absence of OLFM4 both transmigrate into tissue in vivo and give rise to distinct NETs in vitro. PLoS One. 2013 Jul 29;8(7):e69575–e69575.

358. Tang BM, Shojaei M, Parnell GP, Huang S, Nalos M, Teoh S, et al. A novel immune biomarker IFI27 discriminates between influenza and bacteria in patients with suspected respiratory infection. Eur Respir J. 2017;49(6).

359. Tang BM, Shojaei M, Teoh S, Meyers A, Ho J, Ball TB, et al. Neutrophils-related host factors associated with severe disease and fatality in patients with influenza infection. Nat Commun. 2019;10(1):3422.

360. Craig A, Mai J, Cai S, Jeyaseelan S. Neutrophil recruitment to the lungs during bacterial pneumonia. Infect Immun. 2009;77(2):568–75.

361. Hufford MM, Richardson G, Zhou H, Manicassamy B, García-Sastre A, Enelow RI, et al. Influenza-Infected Neutrophils within the Infected Lungs Act as Antigen Presenting Cells for Anti-Viral CD8+ T Cells. PLoS One. 2012 Oct 8;7(10):e46581.

362. Heidema J, Lukens M V., van Maren WWC, van Dijk MEA, Otten HG, van Vught AJ, et al. CD8 + T Cell Responses in Bronchoalveolar Lavage Fluid and Peripheral Blood Mononuclear Cells of Infants with Severe Primary Respiratory Syncytial Virus Infections . J Immunol. 2007 Dec;179(12):8410–7.

363. Gagro A, Tominac M, Krsulović-Hresić V, Baće A, Matić M, Drazenović V, et al. Increased Toll-like receptor 4 expression in infants with respiratory syncytial virus bronchiolitis. Clin Exp Immunol. 2004 Feb;135(2):267–72.

364. Merekoulias G, Alexopoulos EC, Belezos T, Panagiotopoulou E, Jelastopulu DME. Lymphocyte to monocyte ratio as a screening tool for influenza. PLoS Curr. 2010 Mar 29;2:RRN1154–RRN1154.

365. St. John AL, Rathore APS. Adaptive immune responses to primary and secondary dengue virus infections. Nat Rev Immunol. 2019;19(4):218–30.

366. Richards KA, Chaves FA, Sant AJ. The memory phase of the CD4 T-cell response to influenza virus infection maintains its diverse antigen specificity. Immunology. 2011/03/29. 2011 Jun;133(2):246–56.

367. Lee FE, Falsey AR, Halliley JL, Sanz I, Walsh EE. Circulating Antibody-Secreting Cells during Acute Respiratory Syncytial Virus Infection in Adults. J Infect Dis. 2010 Dec;202(11):1659–66.

368. Blanco E, Pérez-Andrés M, Arriba-Méndez S, Contreras-Sanfeliciano T, Criado I, Pelak O, et al. Age-associated distribution of normal B-cell and plasma cell subsets in peripheral blood. J Allergy Clin Immunol. 2018;141(6):2208-2219.e16.

369. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995 Jul 25;57(1):289–300.

370. Plant VU centre for plant systems biology. Ghent university venn diagram tool [Internet]. Available from: http://bioinformatics.psb.ugent.be/webtools/Venn/

371. Vinzi VE, Chin WW, Henseler J, Wang H. Handbook of Partial Least Squares: Concepts, Methods and Applications. 1st ed. Springer Publishing Company, Incorporated; 2010.

372. Lê Cao K-A, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics. 2011;12(1):253.

373. Sharma-Chawla N, Sender V, Kershaw O, Gruber AD, Volckmar J, Henriques-Normark B, et al. Influenza A Virus Infection Predisposes Hosts to Secondary Infection with Different Streptococcus pneumoniae Serotypes with Similar Outcome but Serotype-Specific Manifestation. Pirofski L, editor. Infect Immun. 2016 Dec 1;84(12):3445 LP – 3457.

374. Hosakote YM, Jantzi PD, Esham DL, Spratt H, Kurosky A, Casola A, et al. Viral-mediated inhibition of antioxidant enzymes contributes to the pathogenesis of severe respiratory syncytial virus bronchiolitis. Am J Respir Crit Care Med. 2011 Jun;183(11):1550–60.

375. Björnsdottir H, Welin A, Michaëlsson E, Osla V, Berg S, Christenson K, et al. Neutrophil NET formation is regulated from the inside by myeloperoxidase-processed reactive oxygen species. Free Radic Biol Med. 2015 Dec;89:1024–35.

376. Hendaus MA, Jomha FA, Alhammadi AH. Virus-induced secondary bacterial infection: a concise review. Ther Clin Risk Manag. 2015 Aug 24;11:1265–71.

377. Brereton RG, Lloyd GR. Partial least squares discriminant analysis: taking the magic away. J Chemom. 2014 Apr 1;28(4):213–25.

378. Rodríguez-Pérez R, Fernández L, Marco S. Overoptimism in cross-validation when using partial least squares-discriminant analysis for omics data: a systematic study. Anal Bioanal Chem. 2018 Sep;410(23):5981–92.

379. Guvenel A, Jozwik A, Ascough S, Ung SK, Paterson S, Kalyan M, et al. Epitope-specific airway-resident CD4+ T cell dynamics during experimental human RSV infection. J Clin Invest. 2020 Jan 2;130(1):523–38.

380. Raiden S, Sananez I, Remes-Lenicov F, Pandolfi J, Romero C, De Lillo L, et al. Respiratory syncytial virus (RSV) infects CD4+ T cells: Frequency of circulating CD4+ RSV+ T cells as a marker of disease severity in young children. J Infect Dis. 2017;215(7):1049–58.

381. Mayne M, Cheadle C, Soldan SS, Cermelli C, Yamano Y, Akhyani N, et al. Gene Expression Profile of Herpesvirus-Infected T Cells Obtained Using Immunomicroarrays: Induction of Proinflammatory Mechanisms. J Virol. 2001 Dec 1;75(23):11641 LP – 11650.

382. Tate MD, Ioannidis LJ, Croker B, Brown LE, Brooks AG, Reading PC. The role of neutrophils during mild and severe influenza virus infections of mice. PLoS One. 2011 Mar 14;6(3):e17618–e17618.

383. Tecle T, White MR, Gantz D, Crouch EC, Hartshorn KL. Human neutrophil defensins increase neutrophil uptake of influenza A virus and bacteria and modify virus-induced respiratory burst responses. J Immunol. 2007 Jun;178(12):8046–52.

384. Pierantoni A, Esposito ML, Ammendola V, Napolitano F, Grazioli F, Abbate A, et al. Mucosal delivery of a vectored RSV vaccine is safe and elicits protective immunity in rodents and nonhuman primates. Mol Ther - Methods Clin Dev. 2015;2:15018.

385. Administration U. F and D. Respiratory syncytial virus infection: developing antiviral drugs for prophylaxis and treatment guidance for industry. Rockville, MD; 2017.

386. Leclercq M, Vittrant B, Martin-Magniette ML, Scott Boyer MP, Perin O, Bergeron A, et al. Large-Scale Automatic Feature Selection for Biomarker Discovery in High-Dimensional OMICs Data . Vol. 10, Frontiers in Genetics . 2019. p. 452.

387. Tomic A, Tomic I, Rosenberg-Hasson Y, Dekker CL, Maecker HT, Davis MM. SIMON, an Automated Machine Learning System, Reveals Immune Signatures of Influenza Vaccine Responses. J Immunol. 2019 Aug 1;203(3):749 LP – 759.

388. Davis J, Goadrich M. The Relationship between Precision-Recall and ROC Curves. In: Proceedings of the 23rd International Conference on Machine Learning. New York, NY, USA: Association for Computing Machinery; 2006. p. 233–240. (ICML '06).

389. Song F, Guo Z, Mei D. Feature Selection Using Principal Component Analysis. In: 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization. 2010. p. 27–30.

390. Webb GI. Naïve Bayes BT - Encyclopedia of Machine Learning. In: Sammut C, Webb GI, editors. Boston, MA: Springer US; 2010. p. 713–4.

391. Li SZ, Jain A, editors. LDA (Linear Discriminant Analysis) BT - Encyclopedia of Biometrics. In Boston, MA: Springer US; 2009. p. 899.

392. Breiman L. Random Forests. Mach Learn. 2001 Oct;45(1):5–32.

393. Mora-Buch R, Dotti I, Planell N, Calderón-Gómez E, Jung P, Masamunt MC, et al. Epithelial IL-1R2 acts as a homeostatic regulator during remission of ulcerative colitis. Mucosal Immunol. 2016;9(4):950–9.

394. Yan B, Chen G, Saigal K, Yang X, Jensen ST, Van Waes C, et al. Systems biology-defined NF-kappaB regulons, interacting signal pathways and networks are implicated in the malignant phenotype of head and neck cancer cell lines differing in p53 status. Genome Biol. 2008;9(3):R53.

395. Choudhary S, Boldogh S, Garofalo R, Jamaluddin M, Brasier AR. Respiratory syncytial virus influences NF-kappaB-dependent gene expression through a novel pathway involving MAP3K14/NIK expression and nuclear complex formation with NF-kappaB2. J Virol. 2005 Jul;79(14):8948–59.

396. Hoang LT, Tolfvenstam T, Ooi EE, Khor CC, Naim ANM, Ho EXP, et al. Patient-Based Transcriptome-Wide Analysis Identify Interferon and Ubiquination Pathways as Potential Predictors of Influenza A Disease Severity. PLoS One. 2014 Nov 3;9(11):e111640.

397. Schmitz N, Kurrer M, Bachmann MF, Kopf M. Interleukin-1 Is Responsible for Acute Lung Immunopathology but Increases Survival of Respiratory Influenza Virus Infection. J Virol. 2005 May 15;79(10):6441 LP – 6448.

398. Sims JE, Gayle MA, Slack JL, Alderson MR, Bird TA, Giri JG, et al. Interleukin 1 signaling occurs exclusively via the type I receptor. Proc Natl Acad Sci U S A. 1993 Jul 1;90(13):6155–9.

399. Rauschmayr T, Groves RW, Kupper TS. Keratinocyte expression of the type 2 interleukin 1 receptor mediates local and specific inhibition of interleukin 1-mediated inflammation. Proc Natl Acad Sci U S A. 1997 May 27;94(11):5814–9.

400. Liu W, Yan M, Liu Y, McLeish KR, Coleman WGJ, Rodgers GP. Olfactomedin 4 inhibits cathepsin C-mediated protease activities, thereby modulating neutrophil killing of

Staphylococcus aureus and Escherichia coli in mice. J Immunol. 2012 Sep;189(5):2460–7.

401. Brand KH, Ahout IML, de Groot R, Warris A, Ferwerda G, Hermans PWM. Use of MMP-8 and MMP-9 to assess disease severity in children with viral lower respiratory tract infections. J Med Virol. 2012 Sep;84(9):1471–80.

402. Chen J, Cui G, Lu C, Ding Y, Gao H, Zhu Y, et al. Severe Infection With Avian Influenza A Virus is Associated With Delayed Immune Recovery in Survivors. Medicine (Baltimore). 2016 Feb;95(5):e2606–e2606.

403. Gutiérrez-Fernández A, Inada M, Balbín M, Fueyo A, Pitiot AS, Astudillo A, et al. Increased inflammation delays wound healing in mice deficient in collagenase-2 (MMP-8). FASEB J. 2007/03/28. 2007 Aug;21(10):2580–91.

404. Foronjy RF, Taggart CC, Dabo AJ, Weldon S, Cummins N, Geraghty P. Type-I interferons induce lung protease responses following respiratory syncytial virus infection via RIG-I-like receptors. Mucosal Immunol. 2015;8(1):161–75.

405. Williams KE, Olsen DR. Matrix metalloproteinase-1 cleavage site recognition and binding in full-length human type III collagen. Matrix Biol. 2009;28(6):373–9.

406. Liu X, Yan X, Zhong B, Nurieva RI, Wang A, Wang X, et al. Bcl6 expression specifies the T follicular helper cell program in vivo. J Exp Med. 2012/09/17. 2012 Sep 24;209(10):1841-S24.

407. Xu F, Kang Y, Zhuang N, Lu Z, Zhang H, Xu D, et al. Bcl6 Sets a Threshold for Antiviral Signaling by Restraining IRF7 Transcriptional Program. Sci Rep. 2016;6(1):18778.

408. Ise W, Fujii K, Shiroguchi K, Ito A, Kometani K, Takeda K, et al. T Follicular Helper Cell-Germinal Center B Cell Interaction Strength Regulates Entry into Plasma Cell or Recycling Germinal Center Cell Fate. Immunity. 2018 Apr;48(4):702-715.e4.

409. Benedetti Gassen R, Fazolo T, Nascimento de Freitas D, Borges TJ, Maito F, Mendes DB, et al. RSV Downregulates IL-21/IL-21R On TFH Cells Via PD-L1 Induction In APCs Impairing Protective Humoral Responses. bioRxiv; 2017.

410. Panesso S, Best V, Cao RG, Ye F, Heinonen S, Ramilo O, et al. Reduced Numbers of T-Cell, B-Cell, and CD4+ Follicular Helper T Cell (Tfh) Populations in Infants With Acute Respiratory Syncytial Virus (RSV) Infection. Open Forum Infect Dis. 2016 Dec 1;3(suppl_1).

411. Yao S, Jiang L, Lu Q, Dent AL, Sun J. Macrophage Bcl6 expression promotes viral dissemination and host inflammation during influenza infection. J Immunol. 2016 May 1;196(1 Supplement):78.6 LP-78.6.

412. Ning S, Pagano JS, Barber GN. IRF7: activation, regulation, modification and function. Genes Immun. 2011;12(6):399–414.

413. Amet T, Son YM, Jiang L, Cheon IS, Huang S, Gupta SK, et al. BCL6 represses antiviral resistance in follicular T helper cells. J Leukoc Biol. 2017/05/26. 2017 Aug;102(2):527–36.

414. Zhu B, Zhang R, Li C, Jiang L, Xiang M, Ye Z, et al. BCL6 modulates tissue neutrophil survival and exacerbates pulmonary inflammation following influenza virus infection. Proc Natl Acad Sci. 2019 Jun 11;116(24):11888 LP – 11893.

415. Toufiq M, Roelands J, Alfaki M, Syed Ahamed Kabeer B, Saadaoui M, Lakshmanan AP, et al. Annexin A3 in sepsis: novel perspectives from an exploration of public transcriptome data. Immunology. 2020 Jul 18;n/a(n/a).

416. Le Cabec V, Maridonneau-Parini I. Annexin 3 is associated with cytoplasmic granules in

neutrophils and monocytes and translocates to the plasma membrane in activated cells. Biochem J. 1994 Oct;303 (Pt 2)(Pt 2):481–7.

417. Read CB, Kuijper JL, Hjorth SA, Heipel MD, Tang X, Fleetwood AJ, et al. Cutting Edge: identification of neutrophil PGLYRP1 as a ligand for TREM-1. J Immunol. 2015;194(4):1417–21.

418. Park SY, Jing X, Gupta D, Dziarski R. Peptidoglycan recognition protein 1 enhances experimental asthma by promoting Th2 and Th17 and limiting regulatory T cell and plasmacytoid dendritic cell responses. J Immunol. 2013/02/18. 2013 Apr 1;190(7):3480–92.

419. Zarei Ghobadi M, Mozhgani S-H, Farzanehpour M, Behzadian F. Identifying novel biomarkers of the pediatric influenza infection by weighted co-expression network analysis. Virol J. 2019;16(1):124.

420. Salanova B, Choi M, Rolle S, Wellner M, Luft FC, Kettritz R. Beta2-integrins and acquired glycoprotein IIb/IIIa (GPIIb/IIIa) receptors cooperate in NF-kappaB activation of human neutrophils. J Biol Chem. 2007 Sep;282(38):27960–9.

421. Mittal D, Lepletier A, Madore J, Aguilera AR, Stannard K, Blake SJ, et al. CD96 Is an Immune Checkpoint That Regulates CD8⁺ T-cell Antitumor Function. Cancer Immunol Res. 2019 Apr 1;7(4):559 LP – 571.

422. Guler ML, Gorham JD, Dietrich WF, Murphy TL, Steen RG, Parvin CA, et al. Tpm1, a Locus Controlling IL-12 Responsiveness, Acts by a Cell-Autonomous Mechanism. J Immunol. 1999 Feb 1;162(3):1339 LP – 1347.

423. Zöller J, Ebel J-F, Khairnar V, Schmitt V, Klopfleisch R, Meiners J, et al. CEACAM1 regulates CD8+ T cell immunity and protects from severe pathology during Citrobacter rodentium induced colitis. Gut Microbes. 2020 Nov 1;11(6):1790–805.

424. Barnes M, Freudenberg J, Thompson S, Aronow B, Pavlidis P. Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. Nucleic Acids Res. 2005 Oct 19;33(18):5914–23.

425. Kuo WP, Jenssen T-K, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics. 2002 Mar;18(3):405–12.

426. Park PJ, Cao YA, Lee SY, Kim J-W, Chang MS, Hart R, et al. Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference. J Biotechnol. 2004;112(3):225–45.

427. Tan PK, Downey TJ, Spitznagel ELJ, Xu P, Fu D, Dimitrov DS, et al. Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res. 2003 Oct;31(19):5676–84.

428. Petersen D, Chandramouli GVR, Geoghegan J, Hilburn J, Paarlberg J, Kim CH, et al. Three microarray platforms: an analysis of their concordance in profiling gene expression. BMC Genomics. 2005;6(1):63.

429. Järvinen A-K, Hautaniemi S, Edgren H, Auvinen P, Saarela J, Kallioniemi O-P, et al. Are data from different gene expression microarray platforms comparable? Genomics. 2004 Jun;83(6):1164–8.

430. Woo Y, Affourtit J, Daigle S, Viale A, Johnson K, Naggert J, et al. A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. J Biomol

Tech. 2004 Dec;15(4):276–84.

431. Sacco RE, Durbin RK, Durbin JE. Animal models of respiratory syncytial virus infection and disease. Curr Opin Virol. 2015 Aug;13:117–22.

432. Olivier A, Gallup J, De MacEdo MMMA, Varga SM, Ackermann M. Human respiratory syncytial virus A2 strain replicates and induces innate immune responses by respiratory epithelia of neonatal lambs. Int J Exp Pathol. 2009 Aug;90(4):431–8.

433. Sitthicharoenchai P, Alnajjar S, Ackermann MR. A model of respiratory syncytial virus (RSV) infection of infants in newborn lambs. Cell Tissue Res. 2020 May;380(2):313–24.

434. Taylor G. Animal models of respiratory syncytial virus infection. Vaccine. 2016/11/29. 2017;35(3):469–80.

435. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE. Respiratory syncytial virus infection in elderly and high-risk adults. N Engl J Med. 2005 Apr;352(17):1749–59.

436. Fleming DM, Taylor RJ, Lustig RL, Schuck-Paim C, Haguinet F, Webb DJ, et al. Modelling estimates of the burden of Respiratory Syncytial virus infection in adults and the elderly in the United Kingdom. BMC Infect Dis. 2015;15(1):443.

437. Tseng HF, Sy LS, Ackerson B, Solano Z, Slezak J, Luo Y, et al. Severe Morbidity and Short- and Mid- to Long-term Mortality in Older Adults Hospitalized with Respiratory Syncytial Virus Infection. J Infect Dis. 2020 Jun 27;

438. Coultas JA, Smyth R, Openshaw PJ. Respiratory syncytial virus (RSV): A scourge from infancy to old age. Thorax. 2019 Oct 1;74(10):986 LP – 993.

439. Zhang Y, Wang Y, Gilmore X, Xu K, Wyde PR, Mbawuike IN. An aged mouse model for RSV infection and diminished CD8+ CTL responses. Exp Biol Med. 2002 Feb;227(2):133–40.

440. Wong TM, Boyapalle S, Sampayo V, Nguyen HD, Bedi R, Kamath SG, et al. Respiratory Syncytial Virus (RSV) infection in elderly mice results in altered antiviral gene expression and enhanced pathology. PLoS One. 2014;9(2):e88764.

441. Kawasaki Y, Hosoya M, Katayose M, Suzuki H. Role of serum neutralizing antibody in reinfection of respiratory syncytial virus. Pediatr Int. 2004 Apr;46(2):126–9.

442. Ogilvie MM, Santhire Vathenen A, Radford M, Codd J, Key S. Maternal antibody and respiratory syncytial virus infection in infancy. J Med Virol. 1981;7(4):263–71.

443. Holberg CJ, Wright AL, Martinez FD, Ray CG, Taussing LM, Lebowitz MD. Risk factors for respiratory syncytial virus-associated lower respiratory illnesses in the first year of life. Am J Epidemiol. 1991 Jun;133(11):1135–51.

444. Lamprecht CL, Krause HE, Mufson MA. Role of maternal antibody in pneumonia and bronchiolitis due to respiratory syncytial virus. J Infect Dis. 1976 Sep;134(3):211–7.

445. Piedra PA, Jewell AM, Cron SG, Atmar RL, Paul Glezen W. Correlates of immunity to respiratory syncytial virus (RSV) associated-hospitalization: Establishment of minimum protective threshold levels of serum neutralizing antibodies. Vaccine. 2003 Jul;21(24):3479–82.

446. Stensballe LG, Ravn H, Kristensen K, Agerskov K, Meakins T, Aaby P, et al. Respiratory syncytial virus neutralizing antibodies in cord blood, respiratory syncytial virus hospitalization, and recurrent wheeze. J Allergy Clin Immunol. 2009 Feb;123(2):398–403.

447. Connor E. Reduction of respiratory syncytial virus hospitalization among premature infants and infants with bronchopulmonary dysplasia using respiratory syncytial virus immune globulin prophylaxis. Pediatrics. 1997 Jan;99(1):93–9.

448. Roca A, Abacassamo F, Loscertales MP, Quintó L, Gómez-Olivé X, Fenwick F, et al. Prevalence of respiratory syncytial virus IgG antibodies in infants living in a rural area of Mozambique. J Med Virol. 2002 Aug;67(4):616–23.

449. Openshaw PJ, Chiu C. Protective and dysregulated T cell immunity in RSV infection. Curr Opin Virol. 2013 Aug;3(4):468–74.

450. Fishaut M, Tubergen D, McIntosh K. Cellular response to respiratory viruses with particular reference to children with disorders of cell-mediated immunity. J Pediatr. 1980 Feb;96(2):179–86.

451. Hall CB, Powell KR, Macdonald NE, Gala CL, Menegus ME, Suffin SC, et al. Respiratory Syncytial Viral Infection in Children with Compromised Immune Function. N Engl J Med. 1986 Jul;315(2):77–81.

452. El Saleeby CM, Suzich J, Conley ME, DeVincenzo JP. Quantitative Effects of Palivizumab and Donor-Derived T Cells on Chronic Respiratory Syncytial Virus Infection, Lung Disease, and Fusion Glycoprotein Amino Acid Sequences in a Patient Before and After Bone Marrow Transplantation. Clin Infect Dis. 2004 Jul;39(2):e17–20.

453. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 2009 Jul;6(7):e1000097.

454. Openshaw PJ. The Mouse Model of Respiratory Syncytial Virus Disease. In Springer, Berlin, Heidelberg; 2013. p. 359–69.

455. Verheijden KAT, Henricks PAJ, Redegeld FA, Garssen J, Folkerts G. Measurement of airway function using invasive and non-invasive methods in mild and severe models for allergic airway inflammation in mice. Front Pharmacol. 2014;5 AUG:190.

456. Irvin CG, Bates JHT. Measuring the lung function in the mouse: The challenge of size. Respir Res. 2003;4(1):4.

457. Nordic Cochrane Centre TCC. Review Manager (RevMan). Version 5.3. 2014.

458. Gaeta NC, Ribeiro BLM, Alemán MAR, Yoshihara E, Marques EC, Hellmeister AN, et al. Serological investigation of antibodies against respiratory viruses in calves from Brazilian family farming and their relation to clinical signs of bovine respiratory disease. Pesqui Vet Bras. 2018 Apr;38(4):642–8.

459. Antonis AFG, De Jong MC, Van Der Poel WHM, Van Der Most RG, Stockhofe-Zurwieden N, Kimman T, et al. Age-dependent differences in the pathogenesis of bovine respiratory syncytial virus infections related to the development of natural immunocompetence. J Gen Virol. 2010/06/16. 2010;91(10):2497–506.

460. Kalina W V., Anderson ML, Gershwin LJ. Alternaria aerosol during a bovine respiratory syncytial virus infection alters the severity of subsequent re-infection and enhances IgE production. Comp Immunol Microbiol Infect Dis. 2006;29(2–3):138–56.

461. Bingham HR, Morley PS, Wittum TE, Bray TM, Ellis JA, Queen WG, et al. Effects of 3-methylindole production and immunity against bovine respiratory syncytial virus on development of respiratory tract disease and rate of gain of feedlot cattle. Am J Vet Res. 2000

Oct;61(10):1309–14.

462. Elvander M, Baule C, Persson M, Egyed L, Ballagi-Pordány A, Belák S, et al. An Experimental Study of a Concurrent Primary Infection with Bovine Respiratory Syncytial Virus (BRSV) and Bovine Viral Diarrhoea Virus (BVDV) in Calves. Acta Vet Scand. 1998;39(2):251–64.

463. Gershwin LJ, Himes SR, Dungworth DL, Giri SN, Friebertshauser KE, Camacho M. Effect of bovine respiratory syncytial virus infection on hypersensitivity to inhaled micropolyspora faeni. Int Arch Allergy Immunol. 1994 May;104(1):79–91.

464. Allen JW, Viel L, Bateman KG, Nagy E, Røsendal S, Shewen PE. Serological titers to bovine herpesvirus 1, bovine viral diarrhea virus, parainfluenza 3 virus, bovine respiratory syncytial virus and Pasteurella haemolytica in feedlot calves with respiratory disease: associations with bacteriological and pulmonary cyto. Can J Vet Res. 1992 Oct;56(4):281–8.

465. Blodörn K, Hägglund S, Gavier-Widen D, Eléouët JF, Riffault S, Pringle J, et al. A bovine respiratory syncytial virus model with high clinical expression in calves with specific passive immunity. BMC Vet Res. 2015 Mar;11(1):76.

466. Uttenthal Å, Larsen LE, Philipsen JS, Tjørnehøj K, Viuff B, Nielsen KH, et al. Antibody dynamics in BRSV-infected Danish dairy herds as determined by isotype-specific immunoglobulins. Vet Microbiol. 2000 Oct;76(4):329–41.

467. Rogers SAB, Murphy C. Four acres and a donkey. Ordinary lives. 1979 Mar;163(3):143 p.

468. Orro T, Pohjanvirta T, Rikula U, Huovilainen A, Alasuutari S, Sihvonen L, et al. Acute phase protein changes in calves during an outbreak of respiratory disease caused by bovine respiratory syncytial virus. Comp Immunol Microbiol Infect Dis. 2011;34(1):23–9.

469. Stewart RS, Gershwin LJ. Systemic and secretory antibody responses to sequential bovine respiratory syncytial virus infections in vaccinated and nonvaccinated calves. Am J Vet Res. 1990 Oct;51(10):1596–602.

470. Kalina W V., Woolums AR, Berghaus RD, Gershwin LJ. Formalin-inactivated bovine RSV vaccine enhances a Th2 mediated immune response in infected cattle. Vaccine. 2004;22(11–12):1465–74.

471. Gershwin LJ, Dungworth DL, Himes SR, Friebertshauser KE. Immunoglobulin e responses and lung pathology resulting from aerosol exposure of calves to respiratory syncytial virus and micropolysporafaeni. Int Arch Allergy Immunol. 1990;92(3):293–300.

472. Stewart RS, Gershwin LJ. Role of IgE in the pathogenesis of bovine respiratory syncytial virus in sequential infections in vaccinated and nonvaccinated calves. Am J Vet Res. 1989 Mar;50(3):349–55.

473. Stewart RS, Gershwin LJ. Detection of IgE antibodies to bovine respiratory syncytial virus. Vet Immunol Immunopathol. 1989 Mar;20(4):313–23.

474. Chang J, Choi SY, Jin HT, Sung YC, Braciale TJ. Improved Effector Activity and Memory CD8 T Cell Development by IL-2 Expression during Experimental Respiratory Syncytial Virus Infection. J Immunol. 2004;172(1):503–8.

475. Sacco RE, McGill JL, Pillatzki AE, Palmer M V., Ackermann MR. Respiratory Syncytial Virus Infection in Cattle. Vet Pathol. 2014;51(2):427–36.

476. Graham BS, Bunion LA, Wright PF, Karzon DT. Reinfection of mice with respiratory syncytial virus. J Med Virol. 1991 May;34(1):7–13.

477. Shao HY, Huang JY, Lin YW, Yu SL, Chitra E, Chang CK, et al. Depletion of regulatory T-cells leads to moderate B-cell antigenicity in respiratory syncytial virus infection. Int J Infect Dis. 2015;41:56–64.

478. Harker JA, Godlee A, Wahlsten JL, Lee DCP, Thorne LG, Sawant D, et al. Interleukin 18 Coexpression during Respiratory Syncytial Virus Infection Results in Enhanced Disease Mediated by Natural Killer Cells. J Virol. 2010 Apr;84(8):4073–82.

479. Dakhama A, Lee YM, Ohnishi H, Jing X, Balhorn A, Takeda K, et al. Virus-specific IgE enhances airway responsiveness on reinfection with respiratory syncytial virus in newborn mice. J Allergy Clin Immunol. 2009;123(1).

480. Dakhama A, Park JW, Taube C, Chayama K, Balhorn A, Joetham A, et al. The role of virus-specific immunoglobulin E in airway hyperresponsiveness. Am J Respir Crit Care Med. 2004;170(9):952–9.

481. Barends M, Boelen A, De Rond L, Kwakkel J, Bestebroer T, Dormans J, et al. Influence of respiratory syncytial virus infection on cytokine and inflammatory responses in allergic mice. Clin Exp Allergy. 2002 Mar;32(3):463–71.

482. Matsuse H, Behera AK, Kumar M, Rabb H, Lockey RF, Mohapatra SS. Recurrent Respiratory Syncytial Virus Infections in Allergen-Sensitized Mice Lead to Persistent Airway Inflammation and Hyperresponsiveness. J Immunol. 2000;164(12):6583–92.

483. Bramley AM, Khan MA, Manson HE, Hegele RG. Development of respiratory syncytial virus “bronchiolitis” in guinea pigs does not reflect an allergic predisposition in the host. Chest. 2003 Aug;124(2):671–81.

484. Peebles RS, Sheller JR, Johnson JE, Mitchell DB, Graham BS. Respiratory syncytial virus infection prolongs methacholine-induced airway hyperresponsiveness in ovalbumin-sensitized mice. J Med Virol. 1999 Feb;57(2):186–92.

485. Brown PM, Harford TJ, Agrawal V, Yen-Lieberman B, Rezaee F, Piedimonte G. Prenatal exposure to respiratory syncytial virus alters postnatal immunity and airway smooth muscle contractility during early-life reinfections. PLoS One. 2017;12(2):e0168786.

486. Dakhama A, Park J-W, Taube C, Joetham A, Balhorn A, Miyahara N, et al. The Enhancement or Prevention of Airway Hyperresponsiveness during Reinfection with Respiratory Syncytial Virus Is Critically Dependent on the Age at First Infection and IL-13 Production. J Immunol. 2005 Aug;175(3):1876–83.

487. Ostler T, Pircher H, Ehl S. “Bystander” recruitment of systemic memory T cells delays the immune response to respiratory virus infection. Eur J Immunol. 2003;33(7):1839–48.

488. Jiang X Bin, Huang M, Cui XF, Zhu Y, Yin KS, Yao K. Respiratory syncytial virus infection differentiates airway dysfunction in the central and peripheral airways in OVA-sensitized mice. Exp Lung Res. 2012 Nov;38(9–10):453–62.

489. Culley FJ, Pollott J, Openshaw PJM. Age at first viral infection determines the pattern of T cell-mediated disease during reinfection in adulthood. J Exp Med. 2002;196(10):1381–6.

490. Harker J, Bukreyev A, Collins PL, Wang B, Openshaw PJM, Tregoning JS. Virally Delivered Cytokines Alter the Immune Response to Future Lung Infections. J Virol. 2007 Dec;81(23):13105–11.

491. Zhai S, Hu L, Zhong L, Guo Y, Dong L, Jia R, et al. Respiratory syncytial virus aggravates renal injury through cytokines and direct renal injury. Front Cell Infect Microbiol. 2016;6(SEP):112.

492. Tregoning JS, Pribul PK, Pennycook AMJ, Hussell T, Wang B, Lukacs N, et al. The chemokine MIP1α/CCL3 determines pathology in primary RSV infection by regulating the balance of t cell populations in the murine lung. PLoS One. 2010;5(2):e9381–e9381.

493. Nagata DEDA, Ting HA, Cavassani KA, Schaller MA, Mukherjee S, Ptaschinski C, et al. Epigenetic control of Foxp3 by SMYD3 H3K4 histone methyltransferase controls iTreg development and regulates pathogenic T-cell responses during pulmonary viral infection. Mucosal Immunol. 2015 Sep;8(5):1131–43.

494. Wang J, Kong L, Luo Q, Li B, Wu J, Liu B, et al. Dual Effects of Respiratory Syncytial Virus Infections on Airway Inflammation by Regulation of Th17/Treg Responses in Ovalbumin-Challenged Mice. Inflammation. 2014 Dec;37(6):1984–2005.

495. Muñoz JL, McCarthy CA, Clark ME, Hall CB. Respiratory syncytial virus infection in C57BL/6 mice: clearance of virus from the lungs with virus-specific cytotoxic T cells. J Virol. 1991 Aug;65(8):4494–7.

496. Krishnamoorthy N, Khare A, Oriss TB, Raundhal M, Morse C, Yarlagadda M, et al. Early infection with respiratory syncytial virus impairs regulatory T cell function and increases susceptibility to allergic asthma. Nat Med. 2012 Oct;18(10):1525–30.

497. Zeng S, Wu J, Liu J, Qi F, Kimura Y, Cao Y, et al. Infection with respiratory syncytial virus influences FasL-mediated apoptosis of pulmonary γδ T cells in a murine model of allergen sensitization. J Asthma. 2014 May;51(4):360–5.

498. Zhang L, Liu J, Wang E, Wang B, Zeng S, Wu J, et al. Respiratory syncytial virus protects against the subsequent development of ovalbumin-induced allergic responses by inhibiting Th2-type γδ T cells. J Med Virol. 2013 Jan;85(1):149–56.

499. Lee DCP, Harker JAE, Tregoning JS, Atabani SF, Johansson C, Schwarze J, et al. CD25+ Natural Regulatory T Cells Are Critical in Limiting Innate and Adaptive Immunity and Resolving Disease following Respiratory Syncytial Virus Infection. J Virol. 2010;84(17):8790–8.

500. Rutigliano JA, Johnson TR, Hollinger TN, Fischer JE, Aung S, Graham BS. Treatment with Anti-LFA-1 Delays the CD8+ Cytotoxic-T-Lymphocyte Response and Viral Clearance in Mice with Primary Respiratory Syncytial Virus Infection. J Virol. 2004 Mar;78(6):3014–23.

501. Jiang X Bin, Wang ZD, Zhu Y, Zhang XL, Cui XF, Yao K, et al. Inhibition of CD8+ T lymphocytes attenuates respiratory syncytial virus-enhanced allergic inflammation. Respiration. 2009;77(1):76–84.

502. Sun J, Cardani A, Sharma AK, Laubach VE, Jack RS, Müller W, et al. Autocrine regulation of pulmonary inflammation by effector T-cell derived IL-10 during infection with respiratory syncytial virus. PLoS Pathog. 2011;7(8).

503. Yao S, Jiang L, Moser EK, Jewett LB, Wright J, Du J, et al. Control of pathogenic effector T-cell activities in situ by PD-L1 expression on respiratory inflammatory dendritic cells during respiratory syncytial virus infection. Mucosal Immunol. 2015 Jul;8(4):746–59.

504. Wang J, Wu J, Kong L, Nurahmat M, Chen M, Luo Q, et al. BuShenYiQi Formula strengthens Th1 response and suppresses Th2-Th17 responses in RSV-induced asthma exacerbated mice. J Ethnopharmacol. 2014 May;154(1):131–47.

505. Bukreyev A, Belyakov IM, Berzofsky JA, Murphy BR, Collins PL. Granulocyte-Macrophage Colony-Stimulating Factor Expressed by Recombinant Respiratory Syncytial Virus Attenuates Viral Replication and Increases the Level of Pulmonary Antigen-Presenting Cells. J Virol. 2001 Dec;75(24):12128–40.

506. Cortjens B, De Boer OJ, De Jong R, Antonis AFG, Sabogal Piñeros YS, Lutter R, et al. Neutrophil extracellular traps cause airway obstruction during respiratory syncytial virus disease. J Pathol. 2016 Feb;238(3):401–11.

507. Cormier SA, You D, Honnegowda S. The use of a neonatal mouse model to study respiratory syncytial virus infections. Expert Rev Anti Infect Ther. 2010 Dec;8(12):1371–80.

508. Chapoval S, Dasgupta P, Dorsey NJ, Keegan AD. Regulation of the T helper cell type 2 (Th2)/T regulatory cell (Treg) balance by IL-4 and STAT6. J Leukoc Biol. 2010/03/24. 2010 Jun;87(6):1011–8.

509. Capella C, Chaiwatpongsakorn S, Gorrell E, Risch ZA, Ye F, Mertz SE, et al. Prefusion F, postfusion F, G antibodies, and disease severity in infants and young children with acute respiratory syncytial virus infection. J Infect Dis. 2017;216(11):1398–406.

510. Gershwin LJ, Gunther RA, Anderson ML, Woolums AR, McArthur-Vaughan K, Randel KE, et al. Bovine respiratory syncytial virus-specific IgE is associated with interleukin-2 and -4, and interferon-γ expression in pulmonary lymph of experimentally infected calves. Am J Vet Res. 2000;61(3):291–8.

511. Jensen IP, Thisted E, Glikmann G, Obel N, Kofoed PE, Sambo M, et al. Secretory IgM and IgA antibodies to respiratory syncytial virus in nasopharyngeal aspirates: A diagnostic supplement to antigen detection. Clin Diagn Virol. 1997;8(3):219–26.

512. Stokes KL, Currier MG, Sakamoto K, Lee S, Collins PL, Plemper RK, et al. The Respiratory Syncytial Virus Fusion Protein and Neutrophils Mediate the Airway Mucin Response to Pathogenic Respiratory Syncytial Virus Infection. J Virol. 2013 Sep 15;87(18):10070–82.

513. Nakayama T, Sonoda S, Urano T, Sasaki K, Maehara N, Makino S. Detection of alpha-interferon in nasopharyngeal secretions and sera in children infected with respiratory syncytial virus. Pediatr Infect Dis J. 1993 Nov;12(11):925–9.

514. Becker S, Quay J, Soukup J. Cytokine (tumor necrosis factor, IL-6, and IL-8) production by respiratory syncytial virus-infected human alveolar macrophages. J Immunol. 1991 Dec;147(12):4307–12.

515. Jaovisidha P, Peeples ME, Brees AA, Carpenter LR, Moy JN. Respiratory syncytial virus stimulates neutrophil degranulation and chemokine release. J Immunol. 1999 Sep;163(5):2816–20.

516. Nguyen DT, Louwen R, Elberse K, van Amerongen G, Yüksel S, Luijendijk A, et al. Streptococcus pneumoniae Enhances Human Respiratory Syncytial Virus Infection In Vitro and In Vivo. PLoS One. 2015 May 13;10(5):e0127098–e0127098.

517. Roe MFE, Bloxham DM, White DK, Ross-Russell RI, Tasker RTC, O'Donnell DR. Lymphocyte apoptosis in acute respiratory syncytial virus bronchiolitis. Clin Exp Immunol. 2004 Jul;137(1):139–45.

518. Emboriadou M, Hatzistilianou M, Magnisali C, Sakelaropoulou A, Exintari M, Conti P, et al. Human neutrophil elastase in RSV bronchiolitis. Ann Clin Lab Sci. 2007;37(1):79–84.

519. Zhu L, Liu L, Zhang Y, Pu L, Liu J, Li X, et al. High Level of Neutrophil Extracellular Traps Correlates With Poor Prognosis of Severe Influenza A Infection. J Infect Dis. 2018 Jan 17;217(3):428–37.

520. Ljungman P, Ward KN, Crooks BN, Parker A, Martino R, Shaw PJ, et al. Respiratory virus infections after stem cell transplantation: a prospective study from the Infectious Diseases

Working Party of the European Group for Blood and Marrow Transplantation. Bone Marrow Transplant. 2001 Sep;28(5):479–84.

521. Lalueza A, Folgueira D, Díaz-Pedroche C, Hernández-Jiménez P, Ayuso B, Castillo C, et al. Severe lymphopenia in hospitalized patients with influenza virus infection as a marker of a poor outcome. Infect Dis (Auckl). 2019 Jul 3;51(7):543–6.

522. Huang I, Pranata R. Lymphopenia in severe coronavirus disease-2019 (COVID-19): systematic review and meta-analysis. J Intensive Care. 2020;8(1):36.

523. Choi YS, Dieter JA, Rothaeusler K, Luo Z, Baumgarth N. B-1 cells in the bone marrow are a significant source of natural IgM. Eur J Immunol. 2012 Jan;42(1):120–9.

524. Savage HP, Yenson VM, Sawhney SS, Mousseau BJ, Lund FE, Baumgarth N. Blimp-1-dependent and -independent natural antibody production by B-1 and B-1-derived plasma cells. J Exp Med. 2017/07/11. 2017 Sep 4;214(9):2777–94.

525. Choi YS, Baumgarth N. Dual role for B-1a cells in immunity to influenza virus infection. J Exp Med. 2008 Dec;205(13):3053–64.

526. Penaloza-MacMaster P, Barber DL, Wherry EJ, Provine NM, Teigler JE, Parenteau L, et al. Vaccine-elicited CD4 T cells induce immunopathology after chronic LCMV infection. Science. 2015 Jan;347(6219):278–82.

527. Peters VA, Joesting JJ, Freund GG. IL-1 receptor 2 (IL-1R2) and its role in immune regulation. Brain Behav Immun. 2012/11/27. 2013 Aug;32:1–8.

528. Ji Z, He L, Regev A, Struhl K. Inflammatory regulatory network mediated by the joint action of NF-kB, STAT3, and AP-1 factors is involved in many human cancers. Proc Natl Acad Sci U S A. 2019 May;116(19):9453–62.

529. Coates BM, Staricha KL, Wiese KM, Ridge KM. Influenza A Virus Infection, Innate Immunity, and Childhood. JAMA Pediatr. 2015 Oct;169(10):956–63.

530. Hirohata S. Human Th1 responses driven by IL-12 are associated with enhanced expression of CD40 ligand. Clin Exp Immunol. 1999 Jan;115(1):78–85.

531. Lam LKM, Watson AM, Ryman KD, Klimstra WB. Gamma-interferon exerts a critical early restriction on replication and dissemination of yellow fever virus vaccine strain 17D-204. npj Vaccines. 2018;3(1):5.