

UNIVERSITY OF OXFORD

Characterisation of small RNAs in response to Human Vaccination and Infection

DPhil Thesis

(Michaelmas Term 2022)

Dr Ruth Elizabeth Drury (1125792)

Reading for Doctor of Philosophy in Paediatrics

Kellogg College

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

United Kingdom

Word count: 49,448

Dedication

Dad

Thank you for instilling an enthusiasm for science in me

Ruth

x

"C'est la Vie"

Edele, Keavy, Lindsay, and Sinéad (Blue Peter, 1998)

Acknowledgements

Firstly, thanks go to Andrew Pollard. I remember our first meeting. I'm not sure I made a good impression, but you sent me off with a to-do list and an open door on return, and here I am, 8 years later, finally handing in my DPhil. You have provided me with many opportunities, access to incredible studies, and the space to develop my projects as I saw fit. Despite your heavy workload, you always provided me with support when called upon.

Thanks to Daniel O'Connor. This thesis is what it is because of you. You have been a supervisor and a friend. I may not have always agreed with your advice, but I have always respected it. You are an incredible academic, and I hope we continue collaborating in the future. Our chats on philosophy and politics are a highlight of my time at OVG. Thanks for all your help and support—I honestly can't thank you enough.

To Sagida Bibi—a friend and colleague who provided me with a ready source of conversation, a listening ear and an encyclopaedic knowledge of where everything was in the lab. I could not have functioned in that place without you.

To Marta Valente Pinto, you were there at the start. You encouraged me and were a dear friend throughout our time as clinical fellows and PhD students. Those early years would not have been the same without you. I fully expect to see you pioneering research in Portugal. I look forward to seeing where academia takes you.

Big thanks to the many people who helped me with this work - Lizzy for holding my hand when doing Luminex and Lisa for doing the same. Thanks to Dylan Sheerin and Chris for the mouse samples to support the EUCLIDS work. Thanks to Amy Beveridge for her photographic knowledge of where my samples were across the many OVG freezers and generally being really nice. Thanks to Flo for your questions and enthusiastic engagement whenever I had something to present—also, thanks for being just generally awesome.

Massive thank you to Tanya for working with me on the EUCLIDS PCRs and then getting stuck in with the *in-vitro* work too. I enjoy our chats about research, my findings, and the trials and tribulations of academia. Your drive, enthusiasm and resilience will stand you in good stead in your pursuit of a PhD. Best of luck!

Thank you, Natalie Maalouf, for running the EUCLIDS RT-PCRs. It was a pleasure to work with you. I promise there will be a paper with your name on very soon!

To Debbie and Mari, aka "Team Gene", I had the BEST time with you both. Working with you on DNA extractions was the best thing I ever did in my DPhil. Those memories are epic. I cannot thank you both enough for your comradeship as we faced down the challenges of our DPhils.

Thanks to Dan Drew for listening and giving me a new perspective.

Thank you, staff, at OVG - it's been a wonderful place to work and study. Your hard work underpins all of the work in this thesis.

Thank you to the participants (and their parents) of the OVG trials covered in this thesis—this work is based on your kindness.

Thank you to my friends and family for your love and support. Katherine Handel, you are my rock. Thank you, Franzi Warren, for always taking an interest, being up for coffee and a play date with the kids, and being there whenever a curve ball was thrown my way. My life in Oxford is all the better for having you both in it.

Dad and Lyn—you always encouraged me without pushing. Thanks for feeding the children and me every Tuesday and looking after them so I could scrape a few more hours of writing. You are always there to support Catherine and me, whatever the problem, and you do it gladly, without complaint. I am very lucky.

Thank you, mum, for getting me. For always being there, saying the right thing at the right time. You are my best friend.

Finally, to Catherine, Emmeline and Bryn. The lights of my life. Catherine, you made many sacrifices to enable me to do this Ph.D. You have taken on so much of the household and the childcare so that I could just get this thesis finished. I really am grateful for everything. I know it’s been hard. I hope I have the opportunity to repay your support in the future. I am looking forward to returning to normal family life now because none of this matters without the three of you.

Declaration of Contributions

All RNA extraction, RT-PCR, *in-vitro* work, data processing, normalisation and analyses were performed by the author unless otherwise stated below.

Research questions were conceived by the author with support and guidance from Professor Andrew Pollard, Dr Daniel O'Connor, and Dr Irina Mohorianu

sRNA-seq was run by the Oxford Genomics Centre, Welcome Trust Centre for human genetics, University of Oxford, overseen by Simon Engledow and Theodosios Kyriakou.

All clinical samples were obtained and processed for storage by Oxford Vaccine Group staff.

In addition:

Work pertaining to Chapters 2, 3 and 4:

- Samples used in Chapters 2, 3 and 4 were generated in the EUCLIDS study (EudraCT number 2014-000126-38). This study was overseen by Harri Hughes, Hannah Robinson, Marta Valente Pinto, and Professor Andrew Pollard.

- RT-PCR data presented in Chapter 2 were generated by the author, and Natalie Maalouf and Tanya Dinesh, under supervision by the author

- Mouse models of vaccination used in Chapter 2 were conducted by Dr Dylan Sheerin with oversight from Dr Christina Dold, who presented the author with unprocessed mouse plasma and saline-perfused liver specimens

- IButton® temperature data used in Chapters 2, 3 and 4 were processed by Dr Valente Martina Pinto.

- Whole blood RNA-sequencing data in Chapter 4 and Chapter 6 were processed and analysed by Dr Daniel O'Connor unless otherwise stated.

- Luminex data in Chapter 3 was generated by the author with supervision from Elizabeth Jones and Dr Lisa Stockdale

- Proteomics Mass spectrometry data in Chapter 3 were generated and aligned to the human proteome by Svenja Hester and Georgina Berridge, overseen by Professor Roman Fischer. Dr Sagida Bibi, Svenja Hester and Georgina Berridge did the plasma protein extraction and digestion and ran the LC-M/S.

Work pertaining to Chapter 6:

- Samples used in Chapter 6 were obtained from the Vaccines Against Typhoid Study (VAST), EudraCT Number: 2014-002978-36. This study was overseen by Elizabeth Jones, Maria Moore and Jonathan Gardner, Dr Celina Jin, and Professor Andrew Pollard.

- Cellular, Vi-antibody titres and Fc-mediated functional measurements used for correlation analysis in Chapter 6 were generated by Dr Celina Jin and Dr Jennifer Hill at the Ragon Institute with oversight from Professor Galit Alter. Elizabeth Jones and Dr Celina Jin generated data on Vi-specific ex vivo ELISpots used for correlation analysis in Chapter 6. CyTof data referred to in Chapter 6 were generated by Dr Deborah Cross and Dr Marije Verheul.

Work pertaining to Chapter 7:

- Samples used in Chapter 7 were obtained from the COV001 and COV002 studies, ClinicalTrials.gov Identifiers: NCT04324606 and NCT04400838, respectively. These studies were overseen by Hannah Robinson, Dr Sagida Bibi, Dr Pedro Folegatti, Dr Kate Emery and Professor Andrew Pollard.
- Cytokine data used in data integration in the COVID multi-omics study in Chapter 7 were generated and analysed by Dr Susana Camara.

The Impact of Covid-19 on this work

The work in this thesis was affected by the Covid-19 pandemic. The prioritisation of lab space for conducting the ChAdOx1 nCoV-19 clinical trials meant limited space for wet lab exploration of findings within this thesis. In addition, a multi-omics study was set up to explore the safety of ChAdOx1 nCoV-19 on break-through infection in COVID-19. I was asked to run the sRNA-sequencing aspect of that study, which impacted the time available for the original projects in this thesis. My work within the COVID-19 multi-omics project has been incorporated into results Chapter 7.

Papers

Publications of work presented in this thesis:

O'Connor, Pinto, M, Sheerin D, Tomic A, **Drury R.** *et al.* Gene expression profiling reveals insights into infant immunological and febrile responses to group B meningococcal vaccine. *Mol. Syst. Biol.* **16**, 9888 (2020*) [Publication of the cytokine data presented in Chapter 4]*

Associated Publications undertaken during this DPhil

Drury, RE., et al. "The Clinical Application of MicroRNAs in Infectious Disease." *Frontiers in Immunology*, vol. 8, no. SEP, 2017, doi:10.3389/fimmu.2017.01182 *[review of the roles of miRNAs in infectious disease]*

Drury, RE, et al. "The Effect of H1N1 Vaccination on Serum MiRNA Expression in Children: A Tale of Caution for MicroRNA Microarray Studies." *PLoS ONE*, vol. 14, no. 8, Public Library of Science, 2019, doi: 10.1371/journal.pone.0221143. *[Primary research study assessing the association between vaccination and serum miRNA expression using miRNA microarray profiling]*

Cross, D., **Drury, R.,** Hill, J. & Pollard, A. J. Epigenetics in Sepsis: Understanding Its Role in Endothelial Dysfunction, Immunosuppression, and Potential Therapeutics. *Front. Immunol.* **10**, 1363 (2019). *[Review of epigenetics in sepsis]*

Drury, R. E. & O'Connor, D. Symptom study app provides real-world data on COVID-19 vaccines. *Lancet. Infect. Dis.* **21**, 890–891 (2021). *[Comment on the*

results of the ZOE symptom study app, which reported common side effects of COVID-19 ChAdOx1 nCoV-19 based on big data generated in a community science project]

Manuscripts in preparation:

The effect of vaccination on circulating plasma miRNAs in infants: *Based on the findings from Chapters 3,4, and 5.*

The effect of exposure and infection with *S*. Typhi on sRNA expression in peripheral mononuclear cells—insights from a human challenge study. *Based on the findings from Chapter* 6

Systems vaccinology brings insights into the influence of ChAdOx1 nCoV-19 vaccination on the immune response to break through infection with SARS-CoV-2: *It will include findings from Chapter 7.*

An exploration of non-miRNA sequences identified across three sRNA-sequencing studies. *This paper will add to the literature on non-miRNAs identified in sRNA-seq experiments and how this relates to RNA expression profiles and full-length RNA expression in paired mRNA sequencing data. Very little data is out there on this.*

Quantifying haemolysis using an sRNA signature in plasma sRNA-seq experiments: rescuing signals in haemolysed samples. *To help researchers*

account for haemolysis in plasma sRNA-seq experiments when it was not quantified.

Table of Contents

Dedication i
Acknowledgements iii
Declaration of Contributions vi
Papers ix
Table of Contents xii
List of Tables xxv
List of Figures xxvii
Abbreviations xxxvii
Glossary of Terms xxxix
Abstract xl
1 Introduction 43
1.1 Using vaccination and challenge studies to probe the role of miRNAs in the immune response 44
1.2 Literature review 45
1.2.1 Vaccines 45
1.2.1 Omics and multi-omics data—a tool for understanding the immune response 66
1.2.2 Small non-coding RNAs 72
1.2.3 Technical aspects of miRNA and sRNA studies 96
1.2.4 Critical knowledge gaps in the literature 103

1.3 Using vaccine and infection studies to interrogate the role of sRNAs in vaccination and infection. 103

1.3.1 The EUCLIDS study: exploring how vaccination influences sRNAs and proteins in plasma, and the links to vaccine reactogenicity 104

1.3.2 Vaccines Against Typhoid (VAST): measuring sRNAs in PBMCS to dissect the immune response to vaccination, identifying correlates of protection, and exploring the effect of *S* Typhi exposure and infection..... 104

1.3.3 The COVID-19 Vaccine and Infection multi-omics study (CoVI-multi omics study): studying the influence of COVID-19 infection on whole blood sRNA expression in the context of SARS-CoV-2 vaccination. 105

1.4 Research Questions 105

1.5 Thesis Aims and Objectives..... 106

2 General Methods 107

2.1.1 Sample acquisition 107

2.2 SRNA-seq..... 107

2.2.1 RNA quantification..... 107

2.2.2 sRNA Library Preparation..... 107

2.2.3 sRNA-seq 108

2.2.4 Data processing and QC 108

2.2.5 Statistical analysis..... 109

2.3 Frequently used statistics 110

2.4 Public Data Availability..... 110

2.5 Development of an RNA sequencing alignment pipeline 110

2.5.1 Considerations for an sRNA alignment and assignment pipeline: 112

2.5.2 Final pipeline 116
2.5.3 Assignment 119
2.5.4 Assignment option used in this thesis 121
2.6 MiRNA-mRNA target interactions 121
2.7 miRNA enrichment analyses 121
2.7.1 miRNA vs mRNAs 123
2.7.2 miRNA pathway analysis 123
3 The influence of vaccination on sRNA expression in plasma in infants 125
3.1 Introduction 125
3.1.1 4CMenB: a model for understanding vaccine reactogenicity 125
3.1.2 The EUCLIDS study 126
3.2 Chapter aims 128
3.3 Chapter Methods 128
3.3.1 EUCLIDS study population 128
3.3.2 EUCLIDS study intervention 128
3.3.3 Temperature measurement 129
3.3.4 Measurement of sRNAs in infant plasma 129
3.3.5 Mouse vaccination study 135
3.4 Results 136
3.4.1 Sample demographics 136
3.4.2 sRNA-seq of plasma prior to and 24-hours after vaccination in 4-month-old infants. 137

3.4.3 RT-qPCR data validation of miRNA results from the sRNA-seq data 156

3.4.4 sRNA-seq of plasma before and 24-hours after vaccination in 12-month-old infants. 159

3.4.5 Identifying the cellular and exRNA source of differentially expressed miRNAs 168

3.4.6 The function and targets of differentially expressed miRNAs 173

3.4.7 A mouse vaccination model reveals that a vaccine-induced reduction of plasma miR-122-5p is a cross-species phenomenon linked to vaccine-related temperature rises. 174

3.5 Discussion 177

3.5.1 Summary of key results 177

3.5.2 miRNAs comprise a small proportion of sRNA features in plasma 178

3.5.3 Vaccination influences miRNA expression in plasma and is associated with vaccine reactogenicity. 178

3.5.4 Where, why, and how of differentially expressed miRNAs 179

3.5.5 Role of differentially expressed miRNAs within the immune system 184

3.5.6 Targets of differentially expressed miRNAs 185

3.5.7 Vaccine-induced changes in sRNA-dfs 186

3.6 Conclusion 187

4 The influence of Vaccination on the Infant Plasma Proteome 189

4.1 Introduction 189

4.2 Chapter aims 189

4.3 Chapter Methods 190
4.3.1 Timepoints studied 190
4.3.2 Sample selection 190
4.3.3 Mass spectrometry 190
4.3.4 Luminex 194
4.3.5 Clinical CRP measurement 195
4.3.6 Pathway enrichment 196
4.4 Results 196
4.4.1 Sample demographics 196
4.4.2 Mass spectrometry measurement of plasma proteins prior to and 4 and 24 hours after vaccination 196
4.4.3 Luminex assay of plasma proteins prior to and 4 and 24 hours after vaccination 206
4.4.4 Functions of differentially expressed proteins after vaccination .. 214
4.5 Discussion 219
4.5.1 Vaccination stimulates an acute-phase response 219
4.5.2 Changes in plasma cytokines and chemokines post-vaccination . 224
4.5.3 Kinetics of protein expression 226
4.5.4 Validation of results across assays 226
4.5.5 Comparing findings with the broader literature 226
4.5.6 Changes in plasma protein levels are greater when 4CMenB is co-administered 228

4.6 Conclusion 229

5 Multi-omics Integration of plasma Protein, plasma sRNA and whole-blood RNA sequencing data from an Infant Vaccine Study 230

5.1 Introduction 230

5.2 Chapter aims 230

5.3 Chapter methods 231

5.3.1 Samples and datasets 231

5.3.2 Assessing the overlap between differentially expressed features 231

5.3.3 Pathway enrichment and overlap 231

5.3.4 Integration of omics datasets based on feature expression profiles 232

5.3.5 In-vitro model 236

5.4 Results 237

5.4.1 Samples 237

5.4.2 The overlap between detected and differentially expressed sRNAs, mRNA and protein data sets at the feature level 238

5.4.3 Overlap between miRNA and mRNA and protein data sets at baseline vs 24-hour post-vaccination at the pathway level 243

5.4.4 Integration of omics datasets based on feature expression profiles 246

5.4.5 Multi-omics integration provides potential insights into the regulation of the acute phase response by the liver 250

5.4.6 IL6 inhibits miR-122-5p release from hepatocytes 253

5.5 Chapter Discussion 254

5.5.1 General comparability of the omics datasets 254

5.5.2 A network of post vaccination immune perturbation 256

5.5.1 Finding a PLS signature that captures reactogencity could facilitate vaccine design..... 257

5.5.2 The role of the liver in vaccine responses..... 258

5.5.3 Limitation of integration 258

5.5.1 Why does IL-6 lead to a drop in miR-122-5p release?..... 260

5.6 Overall discussion of the EUCLIDS multi-omics study 261

5.6.1 Insights into the biology of the innate immune response in infants generated from the EUCLIDS study. 261

5.6.2 Age-related effects 262

5.6.3 Drivers of vaccine-related fever..... 263

5.6.4 Clinical application of study results 265

5.6.5 Limitations of the EUCLIDS multi-omics study 266

5.7 Conclusion..... 267

6. sRNA insights into typhoid fever and typhoid vaccines in the context of a human challenge model 269

6.1 Introduction 269

6.1.1 Typhoid..... 269

6.1.2 Typhoid pathogenesis 270

6.1.3 Preventing typhoid 270

6.1.4 Challenges in typhoid..... 271

6.1.5 Vaccines Against *Salmonella* Typhi (VAST) study: an ideal model to dissect vaccine-mediated immunity and host response to *S.* Typhi..... 272

6.1.6 Key findings from VAST..... 274

6.1.7 VAST: a unique opportunity to describe vaccine-mediated immunity and host response to *S.* Typhi at the sRNA level..... 275

6.1.1 VAST sRNA study set-up 276

6.2 Chapter aims 278

6.3 Methods 279

6.3.1 VAST study details..... 279

6.3.2 sRNA sequencing..... 280

6.3.3 Data processing 280

6.4 Results 281

6.4.1 A note on terminology..... 281

6.4.2 Demographics..... 281

6.4.3 sRNA composition of PBMC samples 283

6.4.4 Theme 1: Does typhoid vaccination influence PBMC sRNA expression?..... 285

6.4.5 Theme 2: Does *S.* Typhi exposure and infection influence sRNA expression in PBMCs? 299

6.5 Discussion 321

6.5.1 Theme 1: Does vaccination influence sRNA expression in PBMCs? 321

6.5.2 Theme 2: Does exposure to *S.* Typhi influence sRNA expression in PBMCs? 323

6.5.3 Study strengths 329

6.5.4 Study limitations 329

6.6 Conclusion 331

7 The influence of COVID-19 on sRNAs in whole blood in the context of a ChAdOx1 nCoV-19 trial 333

7.1 Chapter preface 333

7.2 Introduction 333

7.2.1 SARS-CoV-2 and COVID-19 333

7.2.2 ChAdOx-1/nCoV-19 335

7.2.3 The ChAdOx1 The ChAdOx-1/nCoV-19 trials 335

7.2.4 A multi-omics study to assess the influence of ChAdOx-1/nCoV-19 on acute COVID-19 336

7.2.5 An opportunity to understand the role of sRNAs in the immune response to viruses and vaccine-mediated responses during infection. 337

7.3 Aims 338

7.4 Methods 338

7.4.1 Samples and study set up 338

7.4.2 sRNA specific methods 341

7.4.3 mRNA sequencing data 341

7.4.4 Cytokine and chemokine data 342

7.4.5 MiRNA pathway enrichment analysis 342

7.5 Results 342

7.5.1 COVI multi-omics Stage I sRNA results 342

7.5.2 Stage II: A larger confirmatory study comparing health with acute COVID-19 and the influence of ChAdOx1 nCoV-19 vaccination on COVID-19. 350

7.6 Discussion 376

7.6.1 COVID-19 influences whole blood sRNA expression compared with health 376

7.6.2 Studying the evolution of COVID-19 over the first week of illness 377

7.6.3 Comparing COVID-19 with COVID-like illness 380

7.6.4 The effect of ChAdOx1 nCoV-19 vaccination on COVID-19 382

7.6.5 Study Strengths 382

7.6.6 Study limitations and considerations 383

7.7 Conclusion 385

8 Discussion 386

8.1 Key study findings 386

8.1.1 EUCLIDS 386

8.1.2 VAST 388

8.1.3 COVID-19 vaccination and infection study 390

8.1.4 Comparing distribution of RNA classes across studies 391

8.2 Answering the key research questions of this thesis with a view across the three studies 392

8.2.1 Do vaccination and infection alter miRNA expression? 392

8.2.2 Can miRNAs or sRNA-dfs predict or shape response to vaccination and/or infection? 394

8.2.3 Do miRNAs play a functional role the immune response to infection/vaccination? 395

8.3 Potential translation of study findings to the clinic 397

8.4 Thesis limitations 399

8.5 Future work 403

8.6 Conclusion 404

9 Appendix 405

9.1 Appendix for Chapter 2 405

9.2 Appendix for Chapter 3 406

9.2.1 Associations between miRNA expression and degree of sample haemolysis 407

9.2.2 Correlations between duplicate samples across the two EUCLIDS sRNA runs 415

9.2.3 Comparing sRNA $\log_2$ fold changes at 4 months and 12-months 416

9.2.4 Euclids 4-month cohort Library size statistics 424

9.2.5 EUCLIDS 4-month-old cohort—miRNA reads 425

9.2.6 Selection of miRNA for RT-qPCR normalisation 431

9.2.7 EUCLIDS library size statistics 432

9.2.8 EUCLIDS 12-month-old cohort—miRNA reads 433

9.2.9 EUCLIDS 12-month-old cohort—sRNA reads 437

9.3 Appendix for Chapter 4 440

9.3.1 Mass spectrometry Data QC 440

9.4 Appendix for Chapter 5 462

9.5 Appendix for Chapter 6 467

9.5.1 Additional background information on *S.* Typhi and typhoid fever 467

9.5.2 Key findings in other datasets from VAST 473

9.5.3 QC and normalisation of sRNA data 486

9.5.1 Differentially expressed sRNA classes after *Vaccination* 490

9.5.2 Post-vaccination miRNA foldchanges in a publicly available dataset do not validate specific differential expression results but do correlate with post-Vi-vaccine foldchanges. 494

9.5.3 The challenge of identifying correlates of protection in VAST because of participants who are inherently resistant to typhoid. 498

9.5.1 sPLS-DA to identify predictors of vaccine outcome 505

9.5.2 Establishing whether pathways are positively enriched for miRNA targeting by at 10 days after vaccination in diagnosed people or negatively enriched for targeting in people who did not develop typhoid. 506

9.5.3 Differentially expressed sRNA classes after *S*. Typhi challenge ... 510

9.5.4 Pathway enriched amongst sRNA-dfs differentially expressed at 12 and 24-hour post challenge 514

9.5.5 Differences between control vaccinated individuals and Vi-vaccinated individuals at D0 (28 days after vaccination, pre-challenge) ... 522

9.5.6 sPLS-DA to differentiate healthy samples from acute typhoid samples. 524

9.6 Appendix for Chapter 7 529

9.6.1 The COV studies: the first clinical trials of ChAdOx1 nCoV-19 529

9.6.2 sRNA-seq QC 531

9.6.3 differentially expressed sRNAs arise from genes involved in metabolism and RNA silencing, but there is little relationship between mRNA and sRNA expression 7

10 References 13

List of Tables

Table 1 Vaccine platforms in clinical use. This table does not contain platforms that are in development, e.g., DNA vaccines 47

Table 2 Vaccine and host factors, which influence vaccine reactogenicity..... 59

Table 3 Vaccine-related miRNA studies–all studies identified in the literature are presented..... 87

Table 4 Source, biogenesis, and function of sRNA-derived fragments..... 92

Table 5 microRNA target site prediction software/data bases and the algorithms underlying them..... 100

Table 6 MiRNA:mRNA target validation methods 101

Table 7 Options for dealing with multi-mapping reads 113

Table 8 Options that can be implemented in assignment..... 120

Table 9 Demographics of participants with samples submitted for sRNA-seq .. 137

Table 10 Number of features included in the EUCLIDS differential expression analysis 138

Table 11 association testing between % feature type abundance and vaccine (both groups) status and haemolysis..... 142

Table 12 DE miRNAs in plasma pre versus 24 hours post-vaccination 145

Table 13 DE sRNA-dfs, pre vs post-vaccination..... 153

Table 14 association testing between % feature type abundance and vaccine status and haemolysis..... 160

Table 15 Cellular distribution of differentially expressed miRNAs according to miRNA expression atlases..... 171

Table 16 The key processes that DE proteins play a role in..... 215

Table 17 number of samples with paired data collected across all omics datasets. 238

Table 18 Sampled timepoints broken down by vaccine arm in stage I and stage 2 of the sRNA-seq study..... 277

Table 19 The numbers of participant samples subject to sRNA-seq broken down by vaccine group and challenge outcome.. 282

Table 20 DE miRNAs between time points..... 287

Table 21 Stage 1 sample numbers. 343

Table 22 Stage II sample numbers. 351

List of Figures

Figure 1 miRNA biogenesis and mechanism of action [140]. 76

Figure 2 Examples of how miRNAs regulate genes in the immune response and how pathogens can exploit host miRNAs 84

Figure 3 Common steps in the investigation of miRNAs. 97

Figure 4 The overlap between different classes of RNAs at the sRNA-df level. Many sequences multimap to multiple RNAs. 111

Figure 5 Overview of alignment and assignment pipeline. 117

Figure 6 Example of a .sum file: 118

Figure 7 Example Gene expression profiles. 120

Figure 8 mRNA and miRNA enrichment analyses. 122

Figure 9 MIEAA2 GSEA style pathway enrichment analysis 124

Figure 10 EUCLIDS study Overview (adapted from EUCLIDS study protocol) . 127

Figure 11 Overview of study samples and assays studied in this thesis. 130

Figure 12 Global changes in the expression of sRNAs after vaccination in 4-month-old infants 140

Figure 13 The influence of haemolysis on the global expression of sRNAs in 4-month-old infants 141

Figure 14 Influence of vaccination on miRNA expression 144

Figure 15 Dot plot showing the expression of miRNAs DE 24 hours after vaccination compared with baseline when both vaccine groups were combined. 146

Figure 16 Log_2 fold changes for miRNAs showing significant differential expression 24 hours after vaccination in the 4CMenB or Routine group. 147

Figure 17 Distribution of p-values from pre- versus post-vaccine differential expression analysis shown in two ways. 148

Figure 18 Correlations between post-vaccination miRNA expression and maximal post-vaccination temperature. 149

Figure 19 Correlations between post-vaccination miRNA expression and neutrophil abundance 24 hours after vaccination. 150

Figure 20 Influence of vaccination on sRNA-dfs in infant plasma 152

Figure 21 RT-qPCR of miRNAs in infant plasma samples before and after vaccination 157

Figure 22 Correlation between post-vaccination miRNA expression and markers of reactogenicity 158

Figure 23 Global changes in the expression of sRNAs after vaccination in 12-month-old infants.. 161

Figure 24 Influence of vaccination on miRNA expression in 12-month-old infants at baseline versus 24 hours 163

Figure 25 Differential expression results of miRNAs in 12-month-old participants representing a genuinely independent validation cohort. 164

Figure 26 Correlation between post-vaccination expression of miRNAs and neutrophil abundance in the 12-month-old cohort. 165

Figure 27 Influence of vaccination on sRNA-df expression in 12-month-old infants at baseline versus 24 hours 167

Figure 28 Cross plot comparing log_2 fold changes 24-hours post-vaccination at 4 months versus 12 months. 168

Figure 29 XDEC exRNA carrier deconvolution results. 170

Figure 30 Gene set enrichment results for the 4-month-old cohort using all miRNAs. 174

Figure 31 Experimental design to determine whether vaccination influences miRNA expression in mouse plasma and liver. 175

Figure 32 Expression of miR-122-5p in mice plasma 24 hours after vaccination 176

Figure 33 Normalised expression of miR-122-5p in mouse liver 24 hours after vaccination in each vaccine group. 177

Figure 34 Influence of vaccination on protein abundances 198

Figure 35 Dot plot showing the expression of DE proteins 24 hours after vaccination compared with baseline when both vaccine groups were combined. Raw p-values are shown. 199

Figure 36 Influence of vaccination on protein abundance in 12-month-old infants 200

Figure 37 Dot plots of proteins validated as DE expressed after vaccination in the 12-month cohort. 201

Figure 38 heatmap of proteins validated as DE 24 hours after vaccination. 202

Figure 39 Volcano plots showing results of the 4 month cohort differential abundance analysis when both groups are analysed separately..... 203

Figure 40 Proteins showed differential abundance between vaccine groups at 4 hours and 24 hours after multiple testing adjustments.. 204

Figure 41 Volcano plots showing results of differential abundance analysis when both groups are analysed separately in the 12 month cohort. 204

Figure 42 Proteins DE between vaccine group in the 12-month-old cohort. 205

Figure 43 Relationship between plasma protein abundance 24 hours post-vaccination and markers of vaccine reactogenicity. 206

Figure 44 Influence of vaccination on cytokine, chemokine and acute-phase proteins as measured by Luminex..... 208

Figure 45 Dot plots of DE proteins in 4-month-old infants. 209

Figure 46 Influence of vaccination on cytokine, chemokine and acute-phase proteins as measured by Luminex in 12-month-old infants. 210

Figure 47 Dot plots of DE proteins in the 12-month-old infants as measured by Luminex. Numbers represent raw p-values. 211

Figure 48 Dot plots of DE proteins between vaccine groups in 4-month-old infants. = ... 212

Figure 49 12 Dot plots of lab-measured plasma CRP levels between vaccine groups in 12-month-old infants. 213

Figure 50 Correlation between plasma protein levels 24 hours post-vaccination and neutrophil abundance in 4-month-old infants. 214

Figure 51 Protein expression enrichment results 218

Figure 52 Algorithm for building the interaction network. 236

Figure 53 Overlaps between features detected in the mRNA, sRNA and protein datasets. 238

Figure 54 Chord plot showing the relationship between features (miRNAs, sRNA-dfs, proteins and mRNAs) DE at baseline versus 24 hours post-vaccination.

Figure 55 GSEA using miRNA targets as gene sets in the mRNA data analysing baseline versus 24 hours post-vaccination. 241

Figure 56 GSEA of the mRNA data (baseline versus 24 hours post-vaccination) using targets of DE liver-enriched miRNAs 242

Figure 57 Correlation between DE sRNA-dfs and mRNA counterparts. Correlation tested with Spearman correlation.. 243

Figure 58 Venn diagram of overlapping terms derived in gene set analyses of baseline vs 24 hours for the mRNA, miRNA and protein datasets in the 4-month cohort. 244

Figure 59 Consolidation of miRNA, mRNA and protein gene ontology networks. 245

Figure 60 Multi-omics CCA of 4 and 12-month cohort pre and post-vaccine samples 247

Figure 61 DIABLO output..... 249

Figure 62 Gene set enrichment on features selected by the DIABLO model built to differentiate the vaccine groups..... 250

Figure 63 Interaction network centred on the hepatic response..... 252

Figure 64 In-vitro model..... 254

Figure 65 VAST study schema. 274

Figure 66 How study time points can be exploited to answer questions about vaccination and the host response to S. Typhi. 278

Figure 67 Starburst plot showing the number of samples broken down by vaccine group and outcome..... 282

Figure 68 Relative abundance (by number of reads) of each class of sRNA identified in the sRNA-seq data. 284

Figure 69 Overview of the number of DE sRNAs between time points..... 286

Figure 70 Immunoglobulin gene sRNA-dfs upregulated at 7 days after vaccination. 290

Figure 71 Bar plot showing the number of DE sRNAs between Vi-vaccine groups at each time point. 292

Figure 72 Identifying correlates of protection 295

Figure 73 Agreement plot comparing log_2 foldchanges between outcome groups in Vi-PS vaccinees (x-axis) and Vi-TCV vaccinees (y-axis). 296

Figure 74 Agreement plot comparing baseline to 10 days post-vaccination log_2 foldchanges between outcome groups.. 297

Figure 75 Lollipop plot summarising MIEAA2 miRNA pathway enrichment results comparing sRNA expression 10 days after vaccination in those that did and did not develop typhoid after challenge. 298

Figure 76 Bar plot showing the number of DE sRNAs between Vi-vaccine groups after challenge.. 300

Figure 77 PCA of samples taken pre-challenge (day of challenge) and 12 hours and 1-day post-challenge. 302

Figure 78 Changes in sRNA feature expression in the first 24 hours after challenge. 305

Figure 79 Lollipop plot summarising miRNA pathway enrichment results comparing sRNA expression at D0 versus D1. 306

Figure 80 Genes involved in the Salmonella infection KEGG pathway. 308

Figure 81 PCA of samples taken at D0 and D1 coloured by vaccine group..... 309

Figure 82 A: Agreement plot of log_2 foldchanges Vi- and control vaccinees at D0 (x-axis) and D1 (y-axis). 310

Figure 83 comparing effect sizes at D0 vs D1 between diagnosed and non-diagnosed groups.. 312

Figure 84 PCA of samples taken at D0, 12 hours, D1, D7 and at diagnosis. .. 314

Figure 85 Changes in sRNA feature expression at D7 in those that remained well and at diagnosis in those that developed typhoid.l. 316

Figure 86 A: Heat map of D0, diagnosis and D7 samples, generated using DE sRNAs.. 317

Figure 87 Comparing Vi-PS and Vi-TCV at D7 in those that did not develop typhoid.. 320

Figure 88 A summary of answers to the research questions posed at the outset of this Chapter. 331

Figure 89 Visit schema.. 340

Figure 90 Schema of the two phases of this study amended from the study plan.. 341

Figure 91 Distribution of sRNA-seq reads mapping to each class of RNAs. There were no significant differences between the timepoints. 345

Figure 92 PCA analysis of all samples analysed in stage I. 347

Figure 93. Summary of the number of DE sRNAs in different comparisons, broken down by sRNA class type. 347

Figure 94 miRNA pathway enrichment analysis shows enrichment for targetting at symptom onset in participants with COVID-like illness compared with baseline and COVID-19.. 349

Figure 95 Stage I Differential expression results Panels A+B: PCA analysis of all samples analysed in stage II. Samples are coloured by..... 353

Figure 96 Volcano plots showing sRNAs that were DE at COVID-19 onset symptoms compared with baseline. 355

Figure 97 Class-wise changes in sRNA expression over time relative to baseline. Y-axis represents the mean log_2 foldchange of sRNAs within each class of sRNA with 95% confidence intervals shown in the shaded area.. 357

Figure 98 Overlap between miRNAs DE between baseline and S_0 and/or S_7, and miRNAs DE in publicly available datasets.. 358

Figure 99 Enrichment results presented for significantly enriched pathways comparing symptom onset with baseline. 360

Figure 100 Enrichment results presented for significantly enriched pathways comparing 7 days into COVID-19 with symptom onset.. 362

Figure 101 miRNA set enrichment analysis: comparing changes in miRNA expression with their targets in the mRNA data. 365

Figure 102 Baseline vs S_0 GO pathway enrichment comparisons. 367

Figure 103 S_0 vs S_7 GO pathway enrichment comparisons..... 369

Figure 104 Global differences between ChAdOx1 nCoV-19 and control vaccine groups at COVID-19 symptom onset in sRNA average log_2 fold change. 371

Figure 105 COVID-19 DISEASE KEGG pathway with DE features highlighted.. 373

Figure 106 Chord plot showing the cross-over between DE proteins and DE mRNAs and miRNAs that encode and target them at S_0 vs baseline in the MenACWY group..... 375

Figure 107 Agreement plot comparing log_2 foldchanges between D0 and D1 (x-axis) and D0 and D7 (y-axis) in Vi-vaccinees. 524

Figure 108 Over-representation analysis showing, which pathways were overrepresented in the sRNA data.....9

Figure 109 There is a limited relationship between sRNA and expression and the full-length mRNAs they derive from. 11

Abbreviations

4CMenB: 4 component Neisseria meningitidis sero group B vaccine

BCR: B-cell receptor

ceRNAs: competitive endogenous RNAs

CMV: Cytomegalovirus

Cpm: counts per million

DTaP-IPV-Hib-hep B: Diphtheria, Tetanus, Pertussis, Inactivate polio, haemophilus influenzae type B, hepatitis B vaccine

fDCs: follicular dendritic cells

GSEA: Gene Set enrichment analysis

GO:BP: Gene ontology biological process

HAI: heamagglutination inhibition

Hib: (haemophilus influenza type B)

HPV: Human Papilloma Virus

LAIV: Live attenuated flu vaccine

Lcpm: log counts per million

MenB: Neisseria meningitidis sero group B

miRNAs: microRNAs

miRs: microRNAs

MMR: Measles, mumps, rubella vaccine

MREs: MicroRNA recognition site

mTB: mycobacteria Tuberculosis

NK cells: Natural Killer cells

ORA: Over representation analysis

OMV : Outer membrane vesicles

PAMPs: Pathogen associated molecular patterns

PCV: Pneumococcal conjuagte vaccine

PLS-DA: Patial least squares discriminant analysis

PRRs: Pathogen recognition receptors

RISC: Rna induced silencing complex

RNAi: RNA interference

RNPS: Ribonucleic acid proteins

RSV: Respiratory synsitial virus

RT-PCR: Real time quantitative PCR

***S.* Typhi**: *Salmonella enterica* serovar Typhi

SBA: Serum bacteriocidal assays

siRNA: silencing RNAs

SNPs: single nucleotide polymorphisms

sRNA-dfs: small RNA derived fragments

sRNAs: small RNAs

sRNA-seq: small RNA sequencing

TCR: T-cell receptor

TfH: T-follicular helper cells

TIV: Trivalent flu vaccine

TLRs: Toll-like receptors

Treg: T-regulatory cells

Vi-**PS**: Vi-plain polysacheride vaccine

Vi-**TCV**: Vi-tetanus conjugate vaccine

Glossary of Terms

Adverse event	Any untoward negative event that occurs after vaccination. It may or may not be related to vaccination (especially in clinical trials when all negative health occurrences are collected). Vaccine-related adverse events can vary from mild reactogenicity-related signs and symptoms to rare serious events or death. In clinical trials, they are usually graded 0 (did not occur)—5 (death) via the Common Terminology Criteria for Adverse Events (CTCAE). Local and systemic reactogenicity are examples of vaccine-related adverse events.
Antibody titres	Antibody levels measured in serum or plasma
Correlate of protection	Something which can be measured and is associated with protection from infection. Common examples include antibody titers.
Local reactogenicity	Signs and symptoms occurring after vaccination at the vaccination site.
Noise	Unwanted variation in data generated by random effects - can arise from biological or technical sources.
Plasma cells	Terminal differentiation step of mature B lymphocyte. These cells secrete high levels of antibody on activation.
Redundancy	Different features that represent the same thing. Multiple unique reads may all come from an isoform. Several isoforms come from a single gene.
Serobacteriocidal assays	The ability of a sample of serum to kill bacteria in vitro. This killing capacity stems from antibodies in the serum.
Small RNAs	Short RNA molecules typically 20 to 30 nucleotides long generated by systematic processing of precursor RNA by RNAse enzymes
sRNA derived fragments	short RNA molecules identified in sRNA-sequencing experiments that are fragments derived from RNA molecules, which are not a miRNA or piRNA.
Systemic reactogenicity	Systemic signs and symptoms occurring after vaccination, e.g., fever, malaise, GI upset, irritability.

Abstract

Background

Heterogeneity in host responses plays a role in vaccine failure, vaccine-related adverse events and infection outcomes and undermines efforts to produce universally effective vaccines and treatments. Gene expression analyses have provided insights into the immune response, but a deeper understanding requires a broader view encompassing gene regulation.

MicroRNAs (miRNAs) are small RNAs (sRNAs) that post-transcriptionally regulate protein-coding genes. Small RNA sequencing (sRNA-seq) has revealed the presence of other small RNA molecules ~20 to 30 nucleotides long, which arise from RNAs classes, including tRNAs, small nucleolar RNAs and yRNAs. The action of these so-called sRNA-derived fragments (sRNA-dfs) is poorly understood; they may play a role in gene regulation. Their role in the immune system is not known.

Research questions:

1. Does vaccination and infection alter miRNA expression?
2. Can miRNA expression predict or shape the response to vaccination and/or infection?
3. What is the role of miRNAs in the immune response to infection/vaccination?

Methods

sRNA-seq was run on samples from three studies: an infant vaccine study involving the reactogenic 4CMenB vaccine, a randomised controlled trial of two typhoid vaccines set within a typhoid vaccine-challenge study, and a ChAdOx1 nCoV-19 vaccine study that followed participants through COVID-19 infection.

Results

This thesis found vaccination and infection strongly influence miRNAs and sRNA-df expression in plasma, whole blood and PBMCs. In total, 1846 sRNAs were DE after vaccination or infection in at least one study. These features comprised miRNAs and sRNA-dfs derived from tRNAs, snoRNAs, lncRNAs and protein-coding genes.

Associations were found between aspects of the immune response and sRNA expression. The infant vaccine study revealed an association between vaccine fever and post-vaccination expression of hsa-miR-122-5p at 4 months of age. Integration of miRNA, mRNA expression, and proteomics revealed a possible link between liver-derived miRNAs and the acute phase response to vaccination via the action of IL6 on hepatocytes. An in-vitro model supported this.

The typhoid vaccine-challenge study revealed a divergent pattern of miRNA expression 10 days after vaccination is associated with vaccine-mediated protection from typhoid, and that a hypo-responsiveness to *S*. Typhi challenge is associated with susceptibility to typhoid fever.

Finally, the COVID-19 study showed downregulation of miRNAs at COVID-19 onset was associated with wide-scale upregulation of pro-inflammatory genes and pathways. Amelioration of COVID-19 disease severity after ChAdOx1 nCoV-

19 vaccination was reflected at the sRNA level, with less perturbation occurring compared with COVID-19 infection in unvaccinated individuals.

Conclusion

This thesis reports for the first time that vaccination and infection alter the expression of sRNA fragments derived from snoRNA, lncRNAs and protein-coding genes. This thesis opens up exciting avenues for investigating the role of miRNAs and sRNA-dfs as purveyors of the acute phase response and correlates of vaccine-mediated protection, whose dysregulation may contribute to disease pathogenesis, aspects hitherto overlooked.

1 Introduction

Infections cause significant morbidity and mortality worldwide. Between 10% and 20% of deaths are attributable to infections, with significant variations across regions and age ranges[1,2]. Rates are higher amongst children under five, with 49% of deaths attributed to infections[3]. Improved sanitation and nutrition reduce rates, but even in the UK, 1 in 5 childhood deaths are from infections[4].

Preventing death and morbidity from infectious diseases requires new vaccines, diagnostics, and therapeutics—this requires a better understanding of immune responses. Understanding how the immune system successfully or unsuccessfully controls an infection could facilitate development of treatments that augment host immune responses and remedy immune dysregulation, as seen in sepsis.

Vaccination saves around ~2.5 million lives every year[5]. Despite this, many life-threatening infectious diseases have no vaccine, and some existing vaccines require improvement to enhance efficacy or reduce reactogenicity. Efforts to design new vaccines are frequently undermined by disappointing clinical trial results. Failure to translate promising pre-clinical results into licensed vaccines partly reflects inadequate understanding of how vaccines induce an immune response and limited knowledge of how antigen properties, adjuvants, and formulations influence vaccine immunogenicity and reactogenicity. Heterogeneity in vaccine response prompts calls for population and individualised vaccine strategies, but this is impossible given an incomplete understanding of host-based factors influencing vaccine response.

These knowledge gaps are being addressed via "omics" approaches—studying gene and protein expression changes after vaccination, bacterial exposure, and infection[6–13]. However, interrogation of gene regulation that maintains homeostasis and instigates and terminates immune responses is equally important.

A key layer of gene regulation is via the RNA interference (RNAi) pathway based around small non-coding RNAs (sRNAs)—single-stranded RNA molecules typically 20 to 50 nucleotides long. sRNAs transcriptionally and post-transcriptionally control gene expression by binding to promoter loci and mRNAs, respectively[14–19]. microRNAs (miRNAs) are a class of sRNA that regulate around ~60% of protein-coding genes by binding to mRNA 3'UTRs[20–22,23,24]. miRNAs critically control cellular differentiation and functions, including those integral to innate and adaptive immunity[25–35]. Few studies have evaluated miRNAs in vaccine responses, and none have used next-generation sequencing to investigate vaccine effects on the "micronome".

1.1 Using vaccination and challenge studies to probe the role of miRNAs in the immune response.

The controlled nature of vaccination and challenge studies provide a unique opportunity to investigate how antigenic stimulation alters cellular and extracellular sRNA expression. Pursuing sources and targets of differentially expressed miRNAs and integrating this information with paired RNA

transcriptomic and proteomic data could reveal novel insights into the host response to vaccination and infection. Insights gained from this approach could be used to rationalise vaccine and treatment design or even enable personalised vaccine strategies and treatment options.

This thesis investigates how vaccination and infection influence sRNA expression in three clinical studies. The first study (Chapters 3-5) investigates how vaccination influences plasma sRNA and protein expression in infants and explores the link between this and vaccine reactogenicity. The second study (Chapter 6) uses a human typhoid vaccine-challenge model to explore the influence of vaccination, exposure, and infection on sRNA expression in peripheral blood mononuclear cells (PBMCs). The final study (Chapter 7) explores the influence of vaccination and infection on sRNA expression in whole blood in a viral setting using samples from a COVID-19 vaccine study. The following literature review provides information on vaccine immunology, sRNAs, multi-omics approaches, and a brief background on vaccines and diseases studied in this thesis.

1.2 Literature review

1.2.1 Vaccines

Vaccines are a key public health measure[36]. In addition to preventing death and morbidity, vaccination supports health systems by reducing the burden of infectious diseases. Vaccination brings social and economic benefits to the state

and individuals by preventing family disruption caused by illness and reducing loss of economic activity due to death, ill health, and caring responsibilities[37].

Though many regard Edward Jenner's introduction of the cowpox-based smallpox vaccine as the dawn of vaccination, the practice has earlier roots in China and India[38]. In the mid-1500s, variolation was widespread in China. Variolation involved inhaling powdered smallpox scabs or scratching them into the skin, which rendered the person immune, providing it did not cause fatal smallpox first. Today there are licenced vaccines against at least 36 infectious diseases using various vaccine platforms (Table 1)[39].

Table 1 Vaccine platforms in clinical use. This table does not contain platforms that are in development, e.g., DNA vaccines.

Vaccine Class	**Vaccine platform**	**Description**	**Examples (not an exhaustive list)**	**notes**
Whole pathogen also referred to as whole cell	Live attenuated	Live virus or bacteria, capable of infecting cells but with reduced virulence. Weakened viruses are often created through successive rounds of virus passage in non-human cell lines, which removes the selective pressure to replicate in human tissues.	Nasal flu vaccine (LAIV), BCG, Measles, mumps, and rubella (MMR), yellow fever, Typhoid vaccine (MHZ09), Varicella vaccine, Shingles vaccine, Live poliovirus, rotavirus vaccine, cholera vaccine, smallpox, Japanese encephalitis	Not suitable for pregnant women or immunosuppressed/immunodeficient populations. Often have a long duration of high levels of protection.
	Killed/inactivated whole cell	Pathogen has been killed or inactivated (e.g., formalin-inactivated) to render it not live.	Inactivated polio vaccine Whole-cell pertussis vaccine Rabies vaccine Q-fever vaccine Hepatitis A vaccine	Can be very reactogenic
	Split-virion vaccines	Whole virus is disrupted in detergent and inactivated in formaldehyde. This disrupts the structure of the virus whilst retaining important antigens.	Quadrivalent flu vaccine	

Table 1 (continued)

subunit vaccines—comprise of a single/few highly purified antigens	Plain polysaccharide vaccines	Contain just the polysaccharides from the pathogen's outer capsule	PCV (pneumococcal vaccine) Vi-polysaccharide vaccine (typhoid) MenACWY vaccine	Do not induce T-cell responses. Not immunogenic in under 2-year-olds. Generally poor at producing immunological memory responses. May require an adjuvant to enhance immunogenicity because the purity of the antigen means there is a lack of PAMPS that activate the innate immune system reducing the signals, which encourage robust B-cell responses.
	Recombinant protein vaccines	Protein antigen manufactured in a bacterial system through the expression of a transgene encoding the protein antigen of interest. May not be adjuvanted	4CMenB (Neisseria meningitidis sero group B) HPV (human papiloma virus) Hepatitis B vaccine	May require an adjuvant to enhance immunogenicity because the purity of the antigen means there is a lack of PAMPS that activate the innate immune system reducing the signals, which encourage robust B-cell responses.
	Conjugate vaccines	Polysaccharide vaccines that have been covalently bound to a protein, e.g., Tetanus toxoid or diphtheria toxoid—results in a chimeric antigen.	Conjugate flu vaccine, MenC vaccine Conjugate PCV MenACWY vaccine Vi-TCV (typhoid–Vi conjugated to tetanus toxoid) Hib (Haemophilus influenza type B)	T-cells respond to the protein part of the chimeric antigen. This means germinal centre responses can occur with affinity maturation of B-cells leading to high-affinity antibodies, antibody class switching and the creation of immunological memory.

Table 1 (continued)

	Toxoid vaccines	Toxins from pathogens that have been inactivated by heat or formaldehyde	Tetanus toxoid vaccine Diphtheria toxoid vaccine Pertussis toxoid vaccine	Protects recipients from the harmful effects of the bacterial toxin and the bacteria itself by attacking an important component.
	Virus-like particles	Viral proteins are expressed in vitro in bacteria/yeast/cell lines. These proteins self-assemble into viral-like structures containing viral antigens and are secreted from cells and purified.	HPV (human papilloma virus) Hepatitis B vaccine	Similar structure to a virus, thus can present antigens in a repetitive manner, enabling B-cell receptor crosslinking, which strengthens B-cell responses.
	Outer membrane vesicles (OMV)	Purified outer membrane vesicles, which are commonly produced by Gram-negative bacteria. Can be detergent extracted to reduce their reactogenic LPS content or can be harvested from pathogens that have been genetically modified to produce penta-acylated LpxL1 LPS, which is less reactogenic	4CmenB (Meningococcal serogroup B vaccine)	Can be associated with high reactogenicity unless genetically modified strains are used.
viral vectored vaccines		Antigens from the pathogen of interest are encoded within an attenuated or replication-deficient virus.	COVID-19 (ChAdOx SARS-CoV-2–uses replication-deficient chimp adenovirus) Ebola vaccine (rVSV-ZEBOV uses a vesicular stomatitis virus as the vector)	Produce strong TH-1 responses, which are particularly desirable for respiratory pathogens, given previous safety issues in cases where experimental vaccines have produced TH-2 responses. Cheap to produce. If replication-deficient viruses are used, then they are suitable for immunosuppressed people.

Table 1 (continued)

RNA vaccines	RNA encoding the antigen of interest is packaged into liposomes and transfects vaccine recipients' cells, which then translate the mRNA into protein	Pfizer-biontec BNT162b2 (COVID-19) Moderna COVID-19 vaccine	Very flexible
Adjuvanted vaccines	Vaccines, which are formulated with an adjuvant–usually subunit vaccines	6-in-1 vaccine (DTaP-IPV-Hib-hep B) PCV 4CMenB HPV Trivalent and quadrivalent flu vaccines	Increase the immunogenicity of vaccines, skew the immune response towards a TH1 or TH2 response but can also increase reactogenicity. Common examples: aluminium and MF59 (squalene oil).

1.2.1.1 Understanding heterogeneity in vaccine responses.

Heterogeneity in vaccine responses is poorly understood outside of clinical immunodeficiencies. Vaccine failure can be surprisingly high; for example, conjugate typhoid and influenza vaccines have efficacies of only ~55% and ~40-60%,respectively[40]. Age and sex can influence vaccine responses, but inter-individual differences persist after accounting for these factors[41–45]. Cytomegalovirus (CMV) infection is linked to poor vaccine responses through immunosenescence in T-cells and NK cells, though not all studies agree[46–50]. The microbiome may also affect vaccine responses[51]. Genome-wide association studies (GWAS) link genetic polymorphisms to vaccine immunogenicity[52–55]. Identifying host factors that predict vaccine responses could inform personalised vaccine strategies—for example, those at most risk of adverse events or least likely to mount an immunogenic response could receive modified doses or a different vaccination platform where possible.

1.2.1.2 Vaccine immunology

Vaccination exposes the immune system to pathogen antigens that induce antigen-specific antibodies and T-cells, which protect the individual upon pathogen exposure. Discovery of new cellular subsets, redundancy and pleiotropy amongst the cells and pathways of the immune response, and limitations of models in which to dissect vaccine responses create uncertainty, controversy, and contention regarding the generation and persistence of vaccine induced-immunity. This section provides a framework for understanding the immune response to vaccination but is caveated by the above statement.

1.2.1.2.1 Engagement of innate immune responses

Local antigen-presenting cells (APCs), e.g., macrophages and dendritic cells, take up vaccine-instilled antigen. Antigen pathogen-associated molecular patterns (PAMPs) activate pattern recognition receptors (PRRs), such as toll-like receptors (TLRs), on local immune cells and resident stromal cells. Macrophages and mast cells respond within minutes of vaccination, releasing pro-inflammatory factors that cause vasodilation and recruit leukocytes. Neutrophils and monocytes appear at the vaccine site within 3 hours. Later, dendritic cells, NK cells, eosinophils and lymphocytes appear[56]. Stimulation of innate immune cells activates NF-KB, inducing transcription of pro-inflammatory cytokines (e.g., IL1β, IL6, IL10, tumour necrosis factor-alpha [TNF-α], and prostaglandin-E2) and chemokines. This generates local inflammation, e.g. pain and swelling. Cytokines enter the bloodstream, inducing the hepatic acute-phase response and acting on the hypothalamus to increase basal temperature, resulting in fever (section 1.2.1.3). The importance of a robust innate response is underlined by the poor immunogenicity of subunit vaccines, which lack the PAMPS of whole-cell/viral-vector vaccines. Many subunit vaccines must be adjuvanted to be sufficiently immunogenic. Adjuvants, e.g., aluminium salts and oil-in-water emulsion (MF59), activate innate leukocytes, e.g., macrophages, increasing cytokine secretion, which promotes leukocyte recruitment, enhances antigen uptake, and encourages migration to draining lymph nodes for presentation to lymphocytes[57].

1.2.1.2.2 Induction of adaptive immunity

Antigen-bearing APCs at vaccine sites migrate to local lymph nodes, presenting antigen to T and B lymphocytes. Engagement of the T-cell receptor (TCR) or B-cell receptor (BCR) by antigen and their co-receptors (by co-stimulatory molecules on APCs) activates the lymphocytes. Most vaccines protect via antibody generation, as demonstrated by the protective effects of passive serum transfer from vaccinated to unvaccinated animals. The role of T-cells in protection is less clear—particularly against extracellular pathogens, e.g. many bacteria.

1.2.1.2.3 B-cell responses

BCRs recognise soluble ("free antigen") and native antigen presented by APCs. BCR activation induces B-lymphocyte clonal expansion into plasma cells, which initially secrete IgM antibody. Antigen binding to a BCR is insufficient to activate a B-cell. Therefore, a second signal must be received, which may or may not come from a T-cell (T-cell-dependent and T-cell-independent responses, respectively).

1.2.1.2.3.1 T-cell dependent responses

Germinal centres are the hallmark of T-cell-dependent responses. They are foci of macrophages, dendritic cells, B-cells, and T-follicular helper cells within lymph nodes that temporarily form as leukocytes interact. Within germinal centres, B-cells internalise antigen, load it onto MHC II and present it to T-follicular helper cell (TfH) TCRs. The TfH cell then provides survival and co-stimulatory signals, which cause B-cells to differentiate into extra-follicular, short-lived IgM-secreting

plasma cells, B-memory cells, or centroblasts. Centroblasts proliferate and mutate the variable region of their immunoglobulin locus (somatic hypermutation), resulting in a progeny of B-cells called centrocytes. These centrocytes migrate towards native antigen presented by follicular dendritic cells (fDCs). The centrocytes compete to bind and internalise antigen, which they then process and present to TfHs. Engaged TfHs provide a survival signal to the centrocyte, which differentiates back into a centroblast and undergoes further proliferation and somatic hypermutation. Centrocytes with high-affinity BCRs outcompete centrocytes with lower affinity BCRs for fDC presented antigen. Centrocytes with low-affinity BCRs fail to capture antigen, so they cannot present to TfHs and therefore die. Centrocyte competition for antigen creates "evolutionary pressure" favouring survival and proliferation of B-cells with high-affinity BCRs. Recurrent rounds of this selection process (affinity maturation) result in B-cells with high-avidity BCRs.

Another consequence of T-cell help is immunoglobulin class switching, resulting in IgA, IgG, IgD and IgE antibodies that have different functions[58]. T-cell help also induces some activated B-cells to become memory B-cells and long-lived plasma cells that provide long-term protection. Memory B-cells rapidly differentiate into plasma cells if their BCR is engaged by pathogen re-exposure. Long-lived plasma cells migrate to bone marrow, where they secrete antibodies for years[59,60].

1.2.1.2.3.2 T-cell independent responses

TCRs only recognise antigen bound to MHC molecules. Generally, only peptides are loaded into MHCs; thus, TCRs usually only recognise protein-based antigens (rare exceptions being zwitterionic polysaccharides and glycolipids in the case of iNKT-cell)[61]. Polysaccharides are therefore viewed as inherently T-cell independent antigens, i.e., antigens that stimulate B-cell responses without T-cell help. T-cell independent responses can be divided into T-cell independent 1 (TI-1) and T-cell independent 2 (TI-2). TI-1 antigens, e.g. LPS, activate B-cells by concurrently binding to one of their BCRs and TLRs[62]. TI-2 antigens contain repetitive epitopes—e.g., polysaccharide, which allows a single antigen molecule to simultaneously bind (cross-link) multiple BCRs on the B-cell membrane, activating the B-cell[62]. Only B-cells in the spleen's marginal zone, "marginal zone B-cells", can be activated in this way. Children under two lack mature marginal zone B-cells, so they cannot mount TI-2 responses[63]. This partly explains why young children are particularly vulnerable to encapsulated organisms and why vaccines based on TI-2 antigens, e.g. plain polysaccharide vaccines, are poorly immunogenic in young children.

B-cells activated in a T-cell-independent response undergo clonal proliferation into plasma cells that secrete IgM, IgA and IgG2, but affinity maturation and B-cell memory do not occur. This means T-cell independent antigens do not induce high-avidity or long-lived humoral responses. T-cell independent antigens can be converted into T-cell dependent antigens through conjugation to a carrier protein, e.g., tetanus toxoid. B-cell BCRs bind the polysaccharide part of the polysaccharide-conjugate protein complex, internalise the whole complex, and

then incorporate the protein part into MHCII and present it to T-follicular helper cells, providing B-cell survival signals. This strategy enables immunogenicity in young children and the generation of high-avidity antibodies and memory responses.

1.2.1.2.3.3 B-cell kinetics

The antigen-specific plasma cell peak in blood occurs around ~10 days after vaccination. Antibody levels are often highest around ~28 days post-vaccination and then wane, but the kinetics of these measures vary between vaccines and are influenced by vaccine platform, e.g., live vs non-live[64].

1.2.1.2.4 T-cell responses

Activation of TCRs results in clonal expansion, leading to a sharp increase in antigen-specific T-cells, and instigation of T-cell effector functions. MHC restriction of TCRs means vaccines based on non-protein antigens are not expected to induce adaptive T-cell responses. The role of T-cells in vaccine-induced immunity is less understood and may be more important for protection against severe disease rather than infection *per se*. Strong T-cell responses are thought to be most relevant against intracellular pathogens such as viruses and malaria[65].

TCRs on cytotoxic CD8+ve T-cells bind antigen incorporated in MHC class I molecules on all cells, then secrete perforin, granzymes, and granulysin to kill infected cells. Vaccine platforms encouraging antigen presentation in nonprofessional APCs, e.g., live viral, mRNA and viral-vector vaccines, induce strong CD8+ responses.

TCRs on CD4+ T-helper cells recognise antigen presented by MHCII molecules, which are restricted to professional APCs. After engagement of their TCR, naïve CD4+ T-cells become activated and differentiate into one of several CD4+ lineages, classically defined as Th1, Th2, TfH, Th17 and Treg (T-regulatory cells). Local soluble cytokine expression determines differentiation and lineage commitment by inducing one of several transcription factors.

CD4+ve T-helper cells secrete cytokines that skew the immune response towards either a cytotoxic T-helper 1 (TH1) response or a more antibody driven TH-2 response (TH2). Th1 and Th2 categorisations are probably overly simplistic but are a useful paradigm for understanding T-cell responses and have important consequences in vaccination. In the 1960s, a trial of a formalin-inactivated, alum-adjuvanted, whole-cell respiratory syncytial virus (RSV) vaccine in infants was tragically associated with enhanced RSV disease. One theory is that formalin-inactivation of the virus led to a Th2-skewed immune response resulting in non-neutralising antibodies and reduced CD8+ve cytotoxic T-cell responses[66,67]. This led to poor control of viral replication, poor clearance of infected cells, pulmonary eosinophilic infiltrates, and increased inflammation, resulting in worse outcomes for RSV-vaccinated children than controls upon natural RSV infection[66,67]. This catastrophic outcome set back RSV vaccine development by decades and prompted the development of vaccine platforms that induce Th-1 skewed responses, e.g., viral-vector vaccines such as ChAdOx1 (see results Chapter 8).

1.2.1.3 *Vaccine reactogenicity*

Vaccine reactogenicity is the signs and symptoms after vaccination locally (e.g., pain, swelling, redness) and systemically (e.g. fever, headaches, chills, malaise, gastrointestinal upset, appetite loss). Vaccine reactogenicity is the physical manifestation of the inflammatory response arising from innate immune system activation. After the instillation of antigen, the local release of vasoactive factors such as bradykinin and vasoactive amines increases blood flow to the area and plasma extravasation, leading to warmth, redness and swelling. Locally released cytokines, histamine, prostaglandin E_2 and ATP activate and sensitise peripheral nociceptors leading to pain[56,68]. Fever is modulated by locally generated pyrogenic factors (e.g. IL1B and IL6, and TNFα) leaching into the systemic circulation and interacting with the organum vasculosum of the lamina terminalis (OVLT) in the anterior hypothalamus. This upregulates $PGE_{2,}$ which slows the firing rate of warm-sensitive neurons in the pre-optic nucleus of the hypothalamus, altering central perception of body temperature. This leads to shivering, increased metabolism and warmth-seeking behaviour. The binding of blood-borne PAMPS to TLRs on the vessels of the OVLT may induce fever[69]. A rapid onset of fever may also be instigated via the peripheral nervous system through the vagus nerve and from neurones transmitting messages to the CNS directly from the site of inflammation. These theories are based on the attenuation of fever through vagotomy and the administration of local anaesthetic at inflamed sites[70,71].

Vaccine factors influencing reactogenicity are described in Table 2. Whether reactogenicity is important for immunogenicity is not clear. The two seem

positively correlated—for example, higher doses and inclusion of adjuvants are linked to greater immunogenicity and reactogenicity, whilst paracetamol has been shown to reduce reactogenicity but also immunogenicity for some vaccines[72]. This led to the "no pain, no gain" concept, with fears that strategies ameliorating reactogenicity will negatively impact efficacy. Correlation does not mean causation, however. If the pathways that influence reactogenicity and immunogenicity are caused by a common factor but are distinct, selectively targeting pathways responsible for reactogenicity may not jeopardise immunogenicity.

Table 2 Vaccine and host factors, which influence vaccine reactogenicity

Factor type	Feature	Influence on reactogenicity
Vaccine	Antigen	Inactivated/killed whole cell vaccines and live attenuated vaccines tend to be highly reactogenic due to the plethora of PAMPS they contain, and, in the case of live attenuated vaccines, their ability to cause mild infection. Subunit vaccines tend to be less reactogenic unless they are adjuvanted.
	Adjuvants	Including adjuvants in subunit vaccines to enhance immunogenicity often leads to greater reactogenicity due to activation of innate leukocytes and subsequent release of pro-inflammatory cytokines[56]. The type of adjuvant chosen also influences reactogenicity[73].
	Contaminating pyrogens	Vaccine subunits produced through bacterial cell culture can contain trace amounts of bacterial pyrogens such as endotoxin, lipoprotein, peptidoglycan, lipoproteins, and bacterial DNA[74]. The presence of LPS in the outer membrane vesicle component of 4CMenB is thought to be partially responsible for its reactogenicity[75,76].
	Dose	Higher vaccine doses are usually associated with greater reactogenicity, meaning a trade-off must often be made between reactogenicity and immunogenicity when deciding upon dose.
	Concomitant vaccine administration	In some cases, the concomitant administration of vaccines is associated with greater reactogenicity–For example, the concomitant administration of 4CMenB with routine vaccinations increases the incidence of fever than when vaccines are given separately[77].

Table 2 (continued)

	Vaccine schedule	The influence of repeated vaccine doses on reactogenicity varies between vaccines—two recently introduced COVID-19 vaccines are good examples; compared with the first dose, reactogenicity is usually lower after the second dose of the Oxford AstraZeneca adenoviral based ChAdOx1 nCoV-19 vaccine, whereas the reverse is true for the Pfizer-BioNTech mRNA based BNT162b2 vaccine[45,78–82]. A similar phenomenon has also been observed in children vaccinated against flu with TIV[83]. The reasons for this discrepancy are unclear, but could be related to tolerance or sensitization to vaccine excipients and pre-existing adaptive immune response to the antigen.
	Route	Intradermal and subcutaneous administration route is often associated with greater reports of local reactogenicity than intramuscular injection. This is hypothesised to be due to the paucity of nociceptors and antigen-presenting cells in muscle compared with skin and subcutaneous tissues[56].
Technical administration of vaccine		Needle length and needle angle (longer needles inserted at 90° are associated with less local reactogenicity), location of injection site, e.g., buttock associated with less local reactogenicity than thigh, and administration technique/administrator have been associated with injection site reactions[83,84].
Host factors	Age	Some evidence that younger age is associated with higher rates of fever but lower rates of local reactions[45,56].
	sex	Injection site and systemic reactions are often greater in women compared with men[45,85–87].
	Pre-existing immunity	Real-world data indicated that previous infection with covid-19 is linked to greater vaccine reactogenicity upon vaccination with ChAdOx1 nCoV-19 vaccine and BNT162b2[45]
Post vaccine intervention	Paracetamol administration	Paracetamol can reduce fever after vaccination. It is not universally advised as it has been linked to reduced immunogenicity (though the impact of this on vaccine effectiveness is unclear)[72]. Medicines regulators have advised prophylactic paracetamol for very reactogenic vaccines (e.g. 4CMenB) where it does not negatively influence immunogenicity[88].

1.2.1.4 The need for rationalised vaccine development

Vaccine development is long and costly. Around ~94% of vaccines in the preclinical phase fail to reach licensure. Failure at phase III trial stage is not

uncommon[89–91,92]. These failures are expensive—it costs US$31–68 million to take a vaccine from pre-clinical to phase II trials[93]. The costs of bringing vaccines to market include the cost of all the failures along the way—thus, it is estimated to take at least 10 years and ~$319–469 million (range: $137 million–$1·1 billion) to bring a successful vaccine to market[92,93]. These costs are passed onto health care providers (potentially limiting access for citizens, especially those in low and middle-income countries) and hinder development of vaccines unlikely to draw in sufficient profits—e.g. serious but rare infections. Vaccine development is undermined by incomplete knowledge of the factors governing vaccine efficacy and reactogenicity. Furthermore, clinical trial costs limits cohort sizes meaning decisions about antigens, dose, formulation, schedule, route, etc., must be based on limited data. Failure to fully explore each of these factors increases the risk of disappointing trial results or unanticipated safety concerns[94].

Vaccine reactogenicity can be problematic. Firstly, excessive reactogenicity prevents licensure. Secondly, reactogenicity can undermine public confidence deterring individuals/guardians from vaccinating themselves/their children/wards. Concerns about reactogenicity can fuel 'scares', negatively impacting vaccine uptake. For example, the highly reactogenic whole-cell pertussis vaccine contributed to a 1970s scare that the vaccine caused neurological side effects. Vaccine coverage dropped from 79% to 31%, resulting in national pertussis epidemics[95]. Concern regarding one vaccine potentially stokes concerns around others, undermining vaccine programmes as a whole. This reduces vaccination rates, leading to a resurgence of vaccine-preventable

diseases and loss of herd immunity that protects the immunologically vulnerable[96–98]. Vaccine reactogenicity can have implications for healthcare resources too. For example, the introduction of the 4CMenB vaccine led to an increase in infants admitted to hospital for full septic screens as clinicians struggled to differentiate post-4CMenB fever from serious bacterial infection[99].

A rationalised approach to vaccine design is required to reduce failure in clinical vaccine trials and minimise vaccine reactogenicity. Identifying surrogates and correlates of protection and reactogenicity are key to this.

1.2.1.4.1 Correlates of protection

This thesis uses Qin and Gilbert's definition of correlates of protection defined as an immune marker, which correlates with protection and may or may not be mechanistically linked to that protection[100]. They subdivide correlates into mechanistic and non-mechanistic (i.e., statistical correlates). They deprecate the term "surrogate of protection", which others use to describe mechanistic correlates of protection that have a threshold above which protection is guaranteed[100,101]. Common correlates of protection include antibody titres or measuring the ability of antibodies in plasma or serum to stick to and neutralise the pathogen via functional assays, e.g., complement-mediated serum bactericidal assays (SBAs), neutralisation/pseudo neutralisation assays, or haemagglutination inhibition assays (used for influenza).

Ideally, a threshold for protection for the correlate can be established, which predicts the proportion of a population protected from a disease after vaccination

via seroepidemiological studies. These thresholds sometimes infer vaccine effectiveness negating the need for large field trials to demonstrate efficacy. For example, studies of army recruits in the 1960s showed an SBA titre of ≥4 conferred protection against serogroup C meningococcal infection[95]. Based on this work, WHO guidelines decree meningococcal vaccines can be licenced from an efficacy point of view if they induce a ≥4-fold rise in SBA titres in ≥90% of subjects[102]. Such correlates reduce the costs and time to completion of clinical trials because a relatively small, short-term trial is sufficiently powered to estimate vaccine efficacy precisely. It also means early phase clinical trials can provide evidence of whether a vaccine is likely to work before instigating costly phase 3 trials.

Not all vaccines or diseases have robust correlates of protection, e.g., typhoid (see Chapter 6), greatly hindering research. New correlates based on a systems serology approach (e.g., antibody-dependent neutrophil phagocytosis) are being investigated. Even if a threshold for protection cannot be established, any correlate associated with protection could help identify the best vaccine antigen/formulation/dose in pre-clinical or early-phase clinical studies, facilitating vaccine development.

1.2.1.4.2 Correlates of reactogenicity

Quality control (QC) operations use models to infer the pyrogenicity of vaccine batches before release. The outdated rabbit pyrogenicity test involved injecting rabbits with vaccine intravenously and measuring their temperature. The monocyte activation test (MAT) involves *in vitro* stimulation of monocytes with

vaccine and subsequent measurement of pyrogenic cytokine release—IL1β, IL6 or TNFα[74]. The MAT assay is used to QC 4CMenB lots before release to prevent unacceptable fever rates. No other standardised correlates of reactogenicity exist for any vaccine.

Novel correlates of reactogenicity could aid the selection of vaccine antigen, dose, adjuvant, etc., preclinically or in early-phase clinical trials. Systemic and local side effects are the clinically relevant endpoints but are subjective measures open to intraoperative variability. For example, rates of fever, are underestimated in infants using through intermittent axillary temperature measurement post-vaccination when compared with technologies like the iButton® (a probe fixed to the skin that captures continuous temperature data over 24 hours)[103]. Variability in reactogenicity measurements creates noise, reducing power, meaning more participants must be recruited to assess reactogenicity. A measure strongly associated with reactogenicity that is objectively, accurately, and easily measured would be useful.

1.2.1.5 Using systems vaccinology to identify biomarkers of vaccine reactogenicity.

The Biomarkers for Enhanced Vaccine Immunosafety (BioVacSafe) consortium was set up as part of the Innovative Medicines Initiative, a public-private venture funded by the European Union and the European pharmaceutical industry. The project used a "systems vaccinology" approach in 14 randomised controlled trials (RCTs), mainly relating to viral vaccines, to identify novel correlates of reactogenicity. The approach involved measuring serum cytokines

and whole blood gene expression profiles to describe the early immune response to vaccination. Despite differential expression of several cytokines and genes after vaccination, none correlated with reactogenicity. A study of adjuvanted influenza-H1N1 vaccine found an association between the composite score of an individual's adverse events (i.e., summed scores for injection site pain, fatigue, anorexia etc.) and serum expression of IP-10 (CXCL10)[104]. Plasma levels of this cytokine as well as IFN-γ, IL6, IL10, MCP-1, and MIG were also associated with adverse event symptom severity scores (fever, fatigue, headache chills, systemic upset) after the second dose of Pfizer/BioNTech BNT162b2 mRNA vaccine[105]. A study of two adjuvanted hepatitis B vaccines, HBsAg-AS01B and HBsAg-Alum, found CRP, IL-6, IP-10, IFN-γ, MIP-1β and MCP-2 were associated with reactogenicity. However, instead of reporting individual associations between cytokines and reactogenicity scores, the authors performed a principal component analysis (PCA) of cytokine expression in post-vaccination samples[73]. The aforementioned cytokines/acute phase proteins contributed the most to PC1, and the PC1 score was associated with the composite score of an individual's adverse events. This suggests the general profile, rather than the expression of individual cytokines, best associates with vaccine reactogenicity.

A limitation of published reactogenicity biomarker studies is that they are not conducted in children who are the greatest vaccine recipients. Furthermore, they measure a minority of acute phase proteins, e.g., CRP, whilst the wider acute phase is not studied. In addition, published studies mainly focus on viral vaccines, which may induce a different response to bacterial vaccines due to

different PAMPS. The kinetics of the immune response to live viral versus bacterial vaccines are also different. These knowledge gaps must be addressed.

In sum, a better understanding of the immune response after vaccination is needed. A revolutionary approach to this is the use of omics technologies, introduced in the following section.

1.2.1 Omics and multi-omics data—a tool for understanding the immune response

The immune system is complex, comprising many inputs, outputs, and interactions, all carefully controlled through intricate feedforward and feedback mechanisms. The traditional deductive approach to dissecting immune responses takes hypotheses built on what is already known and designs experiments accordingly. Iterative cycles of this process lead to new insights but can miss the unexpected. For this reason, many exciting discoveries had to arise serendipitously through experimental error and the pursuit of spurious results. Deductive reasoning approaches can also struggle to decode processes within multifaceted systems containing redundancy and co-dependent interactions. Big data offer an alternative approach. Big data involves generating hundreds to billions of data points, allowing simultaneous measurement of many molecules/pathways in a complex system. Patterns and associations emerging from big data can generate new hypotheses quicker than classical deductive approaches. Machine learning can augment big-data analyses, uncovering patterns and associations, which the human brain and traditional statistics cannot discern easily. In its purity, data-driven approaches are free from the

constraints of prior knowledge, potentially leading to insights not possible through deductive approaches. The pursuit of novel biomarkers is particularly amenable to data-driven approaches—the biological reasons behind a biomarker do not need to be known; it just needs to predict the outcome of interest. In reality, big data and deductive reasoning are complementary methods. A machine learning algorithm or a large association study can produce a result; however, biological knowledge may be needed to make sense of that result, update current hypotheses, and translate it into a useful output.

"Omics" technologies generate big data in biology, simultaneously measuring the "entire" population of biological units, e.g., DNA sequences (genome), RNA molecules (transcriptome), microRNAs (miRome/micronome), methylation sites (methylome) and histone modifications (collectively the epigenome), proteins (proteome), metabolites (metabolome) and cells. Technical, economic, and time limitations mean omics technologies do not usually assay the entire population of features; nevertheless, a substantial proportion is captured, certainly more than traditional methods (e.g., PCR, western blots, chromatography, ELISAs). An advantage of omics technologies is that they are agnostic, i.e., genes/proteins/metabolites/cells of interest do not need prespecifying. This is ideal when the features of importance are unknown. Several omics technologies may be used within a study ("multi-omics"), and attempts made to integrate their outputs (a "systems biology" approach). Omics studies are bringing insights into the immune system. For example, RNA-sequencing of whole blood revealed the role of tryptophan metabolism during acute typhoid[106].

Multi-omics studies generate complementary layers of information, producing a holistic overview of the immune response not captured by studying single omics datasets in isolation. For example, explaining potential disconnect between mRNA and protein expression via post-transcriptional regulation. Proof that systems biology approaches can untangle the complexities of immune responses comes from a seasonal flu vaccine study, which integrated antibody titres, cytokines levels, flow cytometry, and transcriptomic analyses of PBMC subsets[107]. The study revealed correlations between gene expression and vaccine immunogenicity and identified the role of *CAMKIV* in regulating vaccine responses[107]. Other examples include a study, which described the evolution of the neonatal immune system in the first week of life by integrating transcriptomic, proteomic, metabolomic, cytokine/chemokine, and single-cell immune phenotyping[108]; and a combined whole blood single-cell sequencing and proteomics study that described COVID-19 disease severity at the molecular level[109].

One clinical application of omics is biomarker discovery in infectious disease, e.g. by identifying host signatures that diagnose infections caused by difficult to detect pathogens. This is essentially what clinicians do when evaluating a patient's signs and symptoms, but "omics" methods take this approach further, using gene, protein or metabolomic signature differences to differentiate clinically similar presentations. For example, transcriptomics and a machine learning approach identified a 2-gene host signature that can differentiate viral infections from serious bacterial infections in children[110,111]. This signature has successfully been transferred onto point-of-care technology[112].

1.2.1.1 Challenges of omics data

Omics technologies are expensive; thus, large sample sizes can be prohibitively costly, meaning studies are often underpowered to detect changes in highly variable features and features with small effect sizes. This is compounded by a multiple testing problem (see below). Bioinformatic challenges also exist. Omics data is usually computationally intense to process, QC and analyse. Factors for consideration include identifying and dealing with outlier samples, unwanted noise, confounding factors (e.g., batch effects) and redundancy. The lack of universally standardised methods and pipelines for these steps makes it challenging to select the optimal way to analyse data.

The ability to measure many features leads to vast numbers of data points that cannot be easily comprehended. Dimension reduction methods such as PCA and clustering methods that summarise differences and similarities between samples and identify sources of variation in the data can help with this. Network analysis makes it easier to see how features relate to each other and how information flows through systems. It may help prioritise particular features for follow-up studies, e.g., by identifying highly connected hubs, which might not be the most differentially expressed features but are highly influential within systems. Gene set enrichment analyses (GSEA) can summarise the output of differential expression analyses into the cellular processes in which differentially expressed genes participate.

Multiple testing is an issue in omics data because using the standard p-value cut-off of 0.05, means observing one or more significant results by chance is

highly likely. A multiple testing correction can somewhat control for this issue but increases the chance of type II errors and reduces the power to detect significant differences by increasing the significance threshold. These risks are compounded by relatively small sample sizes common in omics studies, which limits power. In addition, effect sizes of the most differentially expressed features ("top hits") are often over-estimated in differential expression analyses involving many features ("winners curse") and may not represent the true top hits if the whole population had been studied (essentially regression to the mean). Biomarker studies, particularly those involving supervised machine learning, are at risk of overfitting if many features are used, despite methods designed to reduce this possibility (e.g., cross-validation). To negate many of these concerns, high-quality, reproducible science in omics studies requires replicating results in an independent validation cohort.

Another issue with omics technologies is technical bias. For example, adapter-dependent sequence ligation in RNA-seq library preparation and non-specific binding by microarrays[113–115,116]. Data processing and analysis decisions can also introduce biases, particularly if assumptions central to big-data analysis are violated (e.g., RNA/protein/cell content between two conditions is identical, and most features are not differentially expressed). It is, therefore, prudent to validate results by another, usually low-throughput, method, e.g., PCR. Furthermore, technical validation using a cheaper, simpler technology can economically enable assay of interesting hits in validation cohorts and *in vitro* and facilitate translation into the clinical setting.

1.2.1.2 Integrating multi-omics data

Multi-omics datasets need to be integrated to unleash maximum potential, producing insights not notable by measuring each omics in isolation. Integration remains a significant challenge, partly because each omics set is subject to different sources of noise and bias. Methods, such as neural networks, partial least squares and tensor-based approaches, can uncover patterns of variation within datasets arising from groups of connected "cross-omics" features and achieve multi-omics integration. However, they are still in relative infancy, of variable quality and sophistication, and often difficult to understand, implement, and interpret by non-bioinformaticians[117].

1.2.1.3 Considerations for omics data analyses

An important consideration when using a "systems biology" approach is "what is the question?" as this informs the experimental design. Most questions can be divided into two themes: 1) what is the mechanism? 2) can a biomarker be identified? In the first instance, a paired design is helpful as this removes unwanted biological variation that can obscure the signal of interest. A paired design may not be helpful in biomarker identification if it will be used in the clinic when baseline samples are unavailable; in this situation, the biomarker signal must outshine biological variation. The same is true for the output of supervised machine learning algorithms. Some model outputs are predictive but obscure, so they are useful in biomarker questions but less useful for mechanistic insights. Furthermore, depending on the question being asked, feature selection might be necessary–for example, in biomarker studies, clinical feasibility dictates that biomarker signatures entail a small number of easy-to-assay features.

To summarise, omics studies are bringing insights into host response to infection and vaccination. A truly holistic insight requires studying not just genes and proteins but also the mechanisms that regulate their expression, e.g., microRNAs and other non-coding RNAs.

1.2.2 Small non-coding RNAs

1.2.2.1 Non-coding RNAs are crucial to the complexity of life

In 2007, Gerstein et al. updated the definition of a gene: "A gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products"[118]. This was in response to the findings of large consortium projects such as the Encyclopedia of DNA Elements (ENCODE), which found the genome is "pervasively transcribed" but not translated. The updated definition shifted focus from genes primarily encoding proteins to a broader definition encompassing non-coding RNAs—genes whose functional product is RNA. The importance of non-coding RNAs in cellular processes is increasingly recognised, underpinned by several observations:

1) Developmental complexity correlates with the proportion of non-coding DNA in an organism's genome, not the number of protein-coding genes[119].

2) Around 76% of the genome is transcribed, yet only 3% is protein-coding[120,121].

3) Most single nucleotide polymorphisms (SNPs) associated with diseases lie in non-coding DNA regions, including non-coding RNA genes[122].

4) Ribonucleic proteins—complexes of non-coding RNA and proteins—play essential roles in transcribing, modifying, and translating coding genes.

Acknowledgement of the above has increased interest in the role of non-coding RNAs in host response to infection and vaccination.

1.2.2.2 Small non-coding RNA

Non-coding RNAs are a heterogenous group of RNA species broadly categorised based on size and/or function. The term "small RNA" (sRNA) is ambiguous, sometimes meaning RNA molecules shorter than 200 nucleotides long and sometimes meaning RNAs whose precursors have been processed by endonucleases into very short (~22 to 34 nucleotide) single-stranded molecules. This thesis uses the latter definition. A well-recognised species of sRNAs are microRNAs. Other sRNAs are cleaved products of full-length non-coding species (e.g. tRNAs and snoRNAs) and are usually referred to as fragments or halves; this thesis will refer to them as sRNA-derived fragments (sRNA-dfs) (Table 4).

1.2.2.3 microRNAs

MicroRNAs (miRNAs, miRs) are ~22 nucleotides long and primarily post-transcriptionally regulate protein-coding genes. They were first discovered in nematodes in 1993 when a short RNA molecule called lin-4 was found to regulate lin-14 expression[123]. miRNAs are evolutionarily conserved from viruses to eukaryotes[124–132]. Humans have ~2,564 miRNAs according to the miRNA registry, miRbase[20–22]. The curators of MirGeneDB claim only ~550 miRNAs exist as some miRNAs identified by bioinformatic prediction tools have not yet been validated by RNA-seq[133]. Box 1 contains miRNA nomenclature.

miRNAs are transcribed as long primary transcripts (pri-miRs). Pri-miRs are processed into mature miRNA molecules, which are loaded onto the RNA-induced silencing complex (RISC)—a protein of the RNA interference pathway. A RISC-bound miRNA binds to complementary sequences on an mRNA 3'UTR, inhibiting translation (Figure 1). Only nucleotides 2 to 7 ("seed region") of a miRNA must match perfectly for mRNA binding to occur[134]. Consequently, one miRNA can target many mRNAs, and one mRNA can be targeted by many miRNAs, with bound miRNAs cooperatively controlling mRNA translation[135,136,25]. The many-to-many interactions between miRNAs and mRNAs create complex regulatory networks allowing cells to respond quickly to stimuli without excessive perturbation of cell function and homeostasis. Because multiple miRNAs interact with multiple mRNAs, a single function cannot be ascribed to any miRNA. Even though specific miRNAs may be frequently associated with a process, e.g., miR-155 often promotes pro-inflammatory responses, each miRNA is pleiotropic, with their effects depending upon the context, e.g., cell type, cell activation status, and concomitant expression of mRNAs and other miRNAs[137].

Collectively, miRNAs regulate ~60% of protein-coding genes[20,24]. They fine-tune protein production by:

- Acting as buffers so small fluctuations in mRNA transcription do not change protein output[25,141,142].
- Involvement in negative feedback and feedforward loops[23].
- Creating a threshold that means translation cannot occur until the mRNA:miRNA ratio tips in favour of the mRNA. This threshold level can be adjusted by varying expression of either the mRNA or miRNA[143].

- miRNAs may curate a transcription factor signal by targeting one or several genes induced by that transcription factor.

Box 1 miRNA terminology—adapted from[35]

Naming convention. The term microRNA is commonly abbreviated to miRNA or miR. Initial miRNAs discovered in *Caenorhabditis elegans* were prefixed with "lin" and "let", e.g. lin-4 and let-7. Subsequently, miRNAs have been numbered in the order in, which they were discovered, e.g. miR-1. Their prefix is usually shortened to miR and often pre-fixed by their species, e.g. hsa-miR-21 (homosapien-microRNA-21).

-5p and 3-p suffixes: Denotes whether the mature miRNA is derived from the 5'arm or the 3'arm of the stem-loop precursor: e.g. hsa-miR-21-3p, hsa-miR-21-5p.

*** suffix**: A * suffix denotes the miRNA strand from the miRNA duplex that is less abundantly expressed as determined by experiment, e.g. miR-56 (the predominant product) and miR-56* (the opposite miRNA strand). The -5p/-3p suffix supersedes this annotation.

MiRNA family: Closely related miRNAs whose mature sequences differ by 1 nucleotide, e.g. the let-7 family, comprising of let-7a, let-7b, let-7c etc.

MiRNA cluster: Some miRNA genes are found grouped together in polycistronic units with a shared promoter, meaning they are transcribed together as a long primary transcript and are, therefore, co-expressed. E.g. miR-17-19 cluster.

Micronome: The entire miRNA expression profile of a cell or sample.

Seed sequence: Nucleotides 2-7 from the 5' end of the miRNA molecule. Canonical targeting of mRNAs requires that the miRNA seed sequence be perfectly complementary to an mRNA site for binding to occur.

MicroRNA Regulatory Element (MRE): A sequence that is complementary to a miRNA's seed site, usually meaning the miRNA binding site on an RNA molecule

AntagomiR/ antimiR: A miRNA inhibitor- an artificially produced single-stranded RNA molecule, which is perfectly complementary to (and therefore binds to) a target miRNA molecule preventing the miRNA from functioning.

MiRNA mimics: artificially produced analogues of miRNAs. Used to upregulate the levels of specific miRNAs artificially.

RNA interference (RNAi), small interfering RNAs (siRNA) and miRNA. Small interfering RNAs (siRNAs) are double-stranded RNA molecules, which bind to targets with perfectly homologous sequences. MicroRNA and siRNA have analogous but not identical biogenesis. siRNAs are produced from DICER cleavage of endogenously (e.g. mRNA transcript) or exogenously (e.g. viral transcript) derived long double-stranded transcripts. MicroRNA and siRNA pathways conjoin at the point where the guide strand is loaded into RISC complexes[138]. Unlike miRNAs, siRNA bind targets with perfect complementarity resulting in mRNA cleavage. In line with several authors, this review uses RNAi as an umbrella term for the process of miRNA and siRNA molecules interfering with gene expression. It should be noted that some authors ascribe RNAi solely to the process of siRNA mRNA degradation.

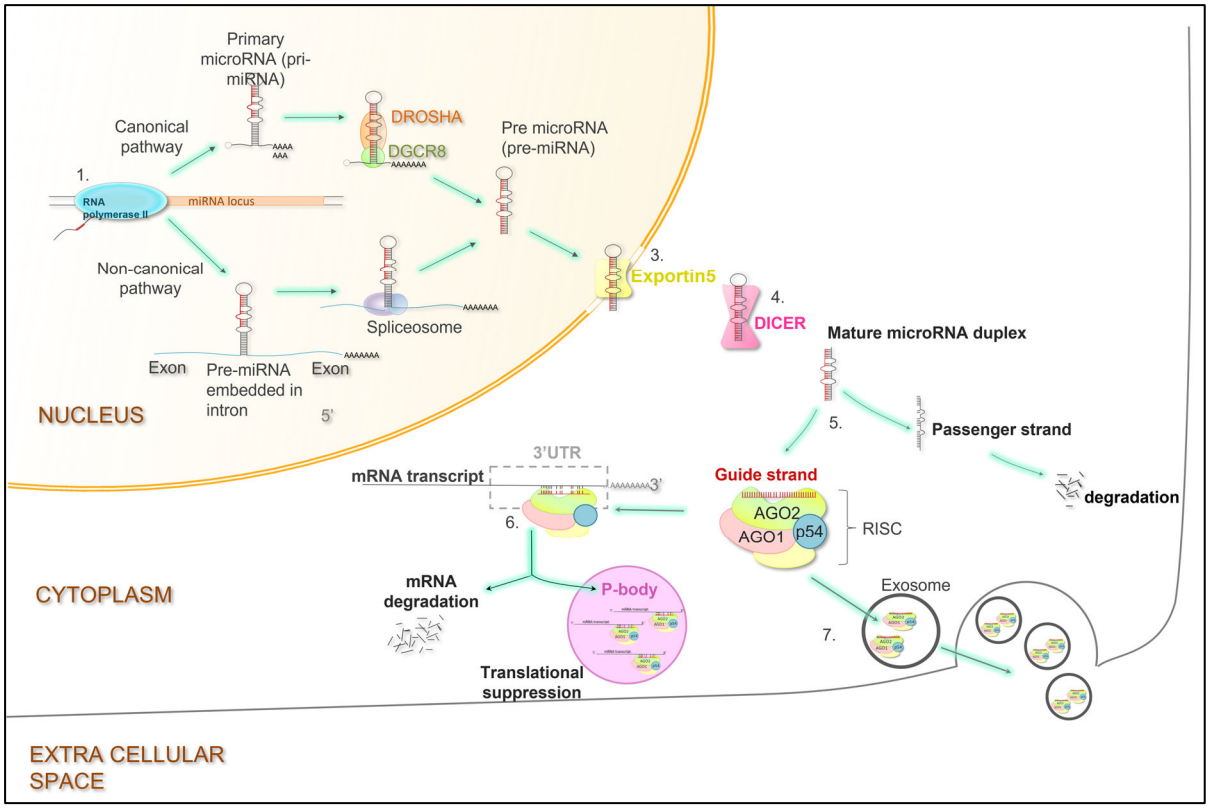


Figure 1 miRNA biogenesis and mechanism of action (composed from references[15–19]) and adapted from[35].

***1**. Within the genome, microRNA genes can be found as independent transcriptional units or embedded within introns and occasionally exons of other genes. In the canonical pathway, microRNAs are transcribed as long primary miRNA transcripts (pri-miRNAs) by RNA polymerase II (and occasionally RNA polymerase III). Pri-miRNAs can be several kilobases long and can contain stem-loops of several mature miRNAs (giving rise to miRNA clusters-see Box 1). In the non-canonical pathway, miRNA precursors lie in mRNA introns ("miRtrons")*

***2.** Pri-miRNAs are processed by the nuclear protein DGCR8 (DiGeorge syndrome critical 8 region) and the enzyme DROSHA into hair pin-shaped structures called pre-miRNA transcripts. In the non-canonical pathway, miRNA precursors in mRNA introns are spliced out, bypassing DGCR8/DROSHA.*

***3**. Pre-miRNAs are exported to the cytoplasm by exportin-5.*

***4**. The enzyme DICER cleaves the pre-miRNA hairpin loop to produce a mature miRNA duplex.*

***5**. One strand of the miRNA duplex (the guide strand) associates with Argonaut (AGO) protein in the RNA-induced silencing complex (RISC), creating the miRISC complex. The remaining strand (termed the passenger strand) is degraded. In most cases, there is a preference for, which strand is incorporated due to factors like thermodynamic stability.*

***6**. The mature single-stranded miRNA in the miRISC complex binds to complementary sequences in the 3' untranslated region of the mRNA molecule, preventing translation. Bound mRNA may be degraded or stored for translation later[139]. Bound mRNA may be sequestered into processing bodies (p-bodies), possibly for later release[140]. After targeting an mRNA, the miRISC can be recycled, and so one miRNA molecule can*

(Figure 1 caption continued)
sequentially target multiple mRNAs
7. *miR: AGO and miRISC complexes can be exported from the cell in exosomes and microvesicles or bound to high-density lipoproteins. Recipient cells may take up these exported miRNAs, wherein they post-transcriptionally regulate gene expression.*

1.2.2.4 Regulating miRNAs

Pri-miRNAs are transcribed by RNA polymerase II and are subject to transcription factors and DNA regulatory elements like mRNAs. miRNAs are one of the fastest-produced transcripts, with up to 10^5 copies per cell generated per day—far higher than mRNA transcription rates.

miRNA turnover is not well understood. Compared with mRNAs, which have half-lives of ~10 hours, most miRNAs are very stable, with half-lives lasting days[144]. miRNA stability is attributed to AGO binding, which protects miRNAs from RNases. There are exceptions to miRNA stability—miRNA half-lives can be as short as 1-2 hours, depending on the temporal and cellular context[145]. Fast miRNA turnover has been attributed to target-mediated miRNA decay. This occurs when a high degree of miRNA-mRNA complementarity causes a conformational change in the AGO-miRNA complex. As a result, the 3' end of the miRNA, usually buried inside AGO, becomes exposed to nucleases and terminal nucleotidyltransferases that perform 3' trimming or nucleotide extensions ("non-templated addition"), respectively, which influence stability[146,147]. Conformational change of AGO can also cause its ubiquitination, leading to proteasomal degradation, releasing the miRNA, which RNAses then degrade[148].

Competitive endogenous RNAs (ceRNAs) control miRNA activity and turnover by binding miRNAs via miRNA response elements (MREs) (Box 1)[149]. ceRNAs are often long non-coding RNAs, including circular RNAs. ceRNAs containing many MREs are sometimes called "miRNA sponges". ceRNA miRNA sequestration results in upregulation of miRNA targets. mRNAs can act as ceRNAs, meaning an mRNA can influence the expression of another mRNA through competition for miRNA binding[150].

1.2.2.5 Extracellular miRNAs

miRNAs are surprisingly stable extracellularly (ex-miRNAs) in serum/plasma, saliva, urine, CSF, and breast milk[151–153]. Their stability is attributed to protection from circulating RNAases by their incorporation into exRNA carriers: extracellular vesicles (exosomes, microvesicles and apoptotic bodies), high-density lipoprotein (HDL) and AGO[151,154–156]. sRNA-dfs are also found extracellularly. sRNA profiles of each exRNA carrier contain distinct sRNA profiles[157,158].

sRNAs may appear in biofluids as a by-product of cell death, passive diffusion across the membrane and/or active/selective release from cells. Evidence for active sorting comes from studies showing the composition of miRNAs released into cell media does not reflect intracellular miRNA composition[159160151,160–162]. One reason for active miRNA release could be "waste disposal", i.e., using miRNA efflux to control intracellular miRNA levels[160]. Another possible function is intercellular regulation of gene expression as miRNAs taken up from the extracellular environment influence gene expression[163–169]. For example, dendritic cells horizontally transfer miR-155 and miR-146a to each other via exosomes, modulating the response to endotoxin[170]. The mechanism behind

selective release is not fully understood. YBXI, Caveolin-1, nSMase2, SRSF1, Sumoylated hnRNPA2B and SYNCRIP are implicated in sorting miRNAs into extracellular vesicles[171–176]. miRNA sorting by hnRNPA2B and SYNCRIP depends on an EXO motif (GGCU) in the miRNA.

1.2.2.6 The role of miRNAs in the immune response

miRNAs mediate processes integral to innate and adaptive immunity, such as cell differentiation, activation and effector functions, pathway regulation and, occasionally, direct targeting of pathogens (Figure 2)[25–35]. Studies have shown miRNAs play a role in over 50 infections[35]. This is comprehensively discussed in my review of the clinical application of microRNAs in infectious disease[35]. A precis of this review is presented below.

1.2.2.6.1 miRNAs influence leukocyte development and activation

Evidence for the integral role of miRNAs in the immune system was first revealed in mice studies, which found miRNA expression in haematopoietic stem cells influences lymphocyte production[26,27-28]. Later studies revealed characteristic, stage-specific miRNA expression profiles during T-cell development with miRNAs governing progression down a CD4+ or CD8+ lineage[177-178]. MiR-125b, miR-21, miR-155, miR-146, miR-181, miR-150, and the miR 17-19 cluster are particularly important in regulating T-lymphocyte development[179,179]. A similar set of miRNAs, miR-150, miR-34, miR-21, miR-155, miR-125, miR-146, mir-181 are important in B-cell development and function[180]. These miRNAs are frequently identified infection biomarkers, reflecting their essential role in immune responses generally. miRNAs set a transcriptional threshold for translation, stopping naïve T-cells from differentiating into effector

cells before TCR activation. For example, naïve T-cells express miR-125b, which targets factors promoting differentiation (including IFN-γ, IL2-subunit receptor beta, IL10-subunit receptor alpha and BLIMP1). MiR-125 b is downregulated upon T-cell activation enabling differentiation into an effector T-cell[181]. MiRNAs are also involved in developing and activating innate immune cells, including NK cells, macrophages, and dendritic cells[32,138,182–186].

1.2.2.6.2 miRNAs allow leukocytes to respond rapidly

miRNAs allow cells to respond quickly to their surroundings through stand-by translation. For example, resting NK cells express miR-30e, which targets perforin mRNA, enabling perforin to be transcribed but not translated during cellular quiescence. This creates a poised state so that upon NK cell activation, rapid miR-30e downregulation enables prompt perforin production, enabling cytolysis of infected cells[187].

MicroRNAs promote inflammatory responses by inhibiting translation of genes that suppress inflammation. For example, several TLRs induce miR-155, which targets anti-inflammatory factors, including SH2-containing inositol 5'-phosphatase 1 (SHIP-1) and suppression of cytokine signalling 1 (SOCS1)[188–190]. MiRNAs also negatively regulate inflammation, e.g., miR-146a negatively regulates TLR signalling by targeting *IRAK1* and *TRAF6,* preventing transcription of pro-inflammatory mediators by NFκβ[184,185]. MicroRNAs also regulate innate immune responses of non-leukocytes, e.g., hepatocytes and epithelial cells[191–195].

MicroRNAs promote balanced immune responses that can eliminate pathogens with minimal collateral host damage. For example, TLR signalling induces miRNAs that repress elements within TLR signalling pathways, including TLRs themselves[29,30,188,196–199]. This creates feedforward and feedback loops that fine-tune TLR signalling so pathogens are eradicated in a controlled manner[188]. For example, macrophage TLR4 signalling induces both the proinflammatory miR-155 and IL-10, a cytokine, which inhibits miR-155 transcription. This system caps miR-155 production, preventing excessive inflammation[200].

MicroRNAs can curate immune responses, adapting responses to combat viral, bacterial, fungal, or protozoal infections, e.g., by controlling macrophage polarisation and skewing the TH1/TH2 balance. MiR-29, for example, promotes a TH2 response by targeting T-bet, eomesodermin and IFN-γ that promote TH1 differentiation[33].

MicroRNAs are integral to adaptive immune responses such as antibody affinity maturation, e.g., germinal centre formation is miRNA dependent as shown by global knockout of miRNAs in B lymphocytes (via conditional DICER knockdown)[201].

1.2.2.6.3 Polymorphisms in miRNAs are associated with susceptibility to infection

SNPs in miRNA loci or microRNA binding sites are associated with susceptibility to, and outcome in, leprosy, brucellosis, HIV, EBV, HCV, HPV, hCMV and hepatitis B infection[202–209]. Although confirming the functional nature of such

associations is difficult, these findings are in keeping with miRNA importance in immunity.

1.2.2.6.4 miRNAs may target pathogens directly.

miRNAs are potential independent mediators of immunity through direct targeting of pathogen genomes and transcripts—a well-documented phenomenon in plants and insects. Studies of human cells show miRNAs target human T-cell leukaemia virus 1, vesicular stomatitis virus, enterovirus-71, human papillomavirus and malaria parasites, inhibiting pathogen replication, enhancing pathogen clearance and/or inducing latency[210–2155]. Recent bioinformatics and *in vitro* studies predict several miRNAs target the SARS-CoV-2 genome and suggest SARS-CoV-2 may act as a miRNA sponge, but this has yet to be shown *in vivo*[216,217]. Interestingly, even if miRNA targeting of SARS-CoV-2 occurs *in vivo*, the effect is not entirely predictable as host miRNA binding sometimes facilitates pathogen survival[218]. For example, miR-122-5p binding to the 5'UTR of the hepatitis C virus (HCV) genome creates a 3' overhang, shielding the virus from nuclear degradation[218,219,220]. This partly explains the hepatocyte tropism of HCV, as miR-122-5p is generally viewed as liver-specific.

Evidence miRNAs target viral genomes/transcripts *in vivo* comes from successful attenuation of live viral vaccines by integrating MREs into the viral genome in clinical trials[221222–226]. The relative importance of miRNAs as direct mediators of immunity is disputed. Many argue the interferon response in mammals supersedes antiviral RNA interference, making it obsolete. Others emphasize highly mutating viruses escape miRNA targeting through mutation as recently demonstrated in clinical trials of miRNA attenuated vaccines[214,226].

1.2.2.6.5 Pathogens exploit cellular miRNAs.

Pathogens can subvert and exploit miRNAs to create an immune-tolerant environment necessary for pathogen survival and latency (Figure 2). HIV, dengue, vaccinia, EBV and HBV target proteins in the RNAi pathway (e.g., DICER) to globally suppress cellular miRNA production, circumnavigating miRNA-mediated innate immunity[227–231]. Enterovirus, salmonella, CMV, HPV, HBV, and HCV induce and suppress specific cellular miRNAs that inhibit and promote immune responses,respectively[232–239].

1.2.2.6.6 Extracellular miRNAs in the immune response

Ex-miRNAs play a role in immune responses and may be biomarkers of infection and vaccine responses.

1.2.2.6.6.1 The functional role of extracellular miRNAs in the immune response

Functional roles of ex-miRNAs are being uncovered. Mittlebrun et al. showed T-cells, B-cells, and dendritic cells release miRNAs in exosomes, which APCs take up where they influence gene expression[169]. Yahyaei et al. found miR-233, and miR-29a were upregulated in serum from people resistant to HIV and discovered this was due to selective PBMC release leading to downregulated intracellular levels that hindered infection[160].

Extracellular miRNAs contribute to the pathophysiology of infection; for example, erythrocytes infected with *Malaria falciparum* release extracellular vesicles containing miR-451a. Endothelial cells internalise these vesicles leading to the downregulation of proteins that maintain endothelial integrity, thus contributing to endothelial dysfunction in severe malaria[240].

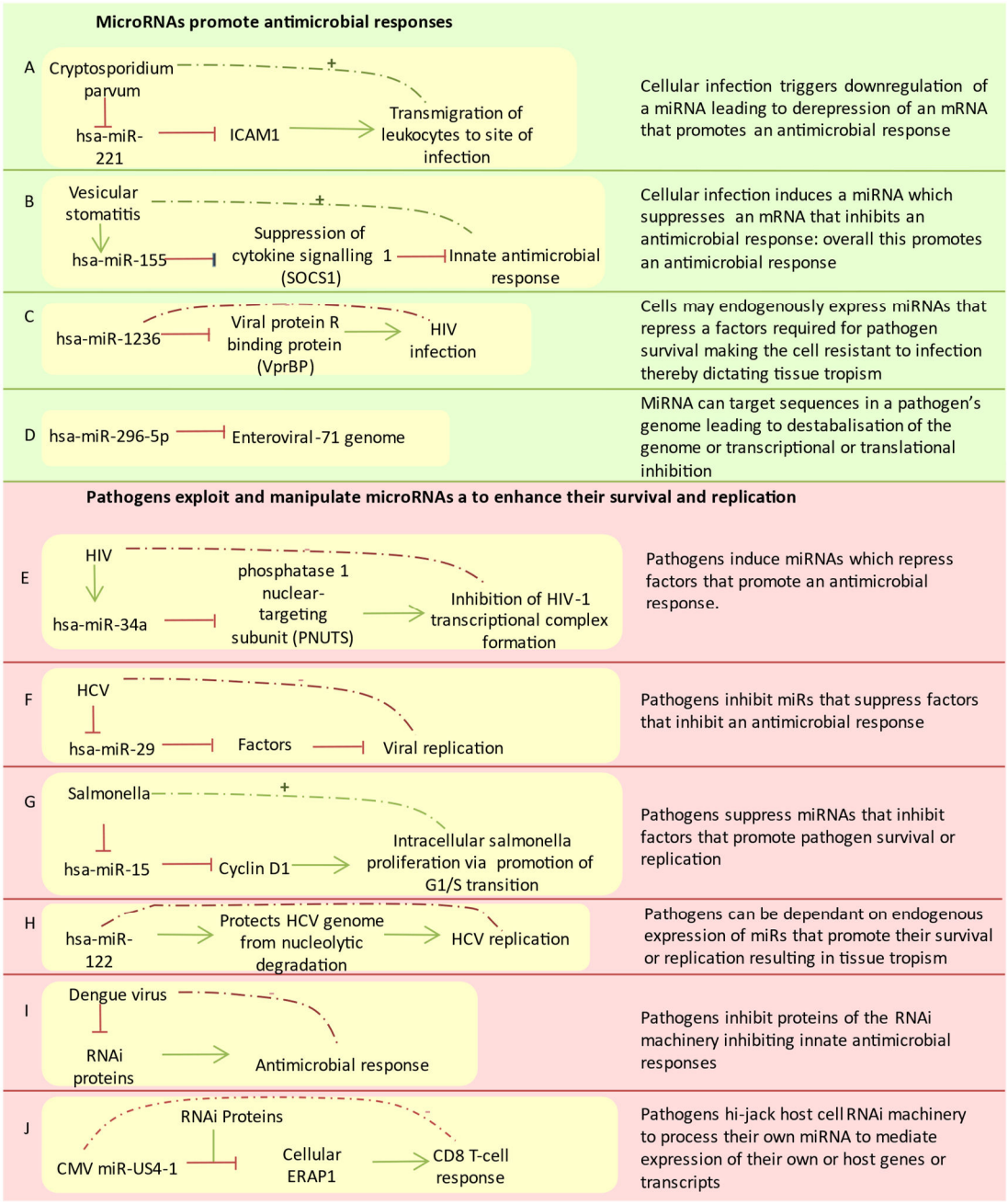


Figure 2 Examples of how miRNAs regulate genes in the immune response—green boxes and how pathogens can exploit host miRNAs—red boxes. Adapted from reference 35

An issue for determining the role of ex-miRNAs *in vivo* is establishing donor and recipient cells. Identifying surface markers of miRNA-carrying exosomes (e.g., CD44) could help, but miRNA bound to protein and HDL is difficult to track[241].

Novel methods for tracking and studying exosomes will address this issue[169,242,243].

1.2.2.6.6.2 Extracellular miRNAs as infection and vaccine biomarkers.

Ex-miRNA stability makes them attractive for use in clinical settings as biomarkers or diagnostics because sample handling can be less stringent. There is a growing literature associating ex-miRNAs with infectious disease, including sepsis outcomes[35,244,245,246,247,248]. Heterogeneity in study design, miRNA detection, data normalization and analysis, small sample sizes, and frequent absence of validation cohorts hinder reproducibility of results and cross-comparison between studies—these issues must be overcome to progress the field.

1.2.2.6.7 miRNAs and vaccine responses

Evidence that miRNAs play a role in vaccine responses comes from studies linking genetic variation in miRNA genes with hepatitis B surface antigen and diphtheria toxoid antibody titres. In both cases, hsa-mir-146a SNPs were associated with post-vaccination antibody levels[249,250].

Few studies investigate the effect of vaccination on miRNA expression (Table 3). Several studies have measured intracellular or exosomal miRNA profiles of ex-vivo stimulated bulk PBMCs or B and T lymphocytes from vaccinated participants (Table 3)[251,252]. In such studies, the participants are vaccinated, and 4 weeks later, cells are *ex vivo* stimulated by vaccine antigen, and miRNA expression assessed. Often results are compared between vaccine responders and vaccine non-responders (usually defined by antibody titres). The issue with these studies

is they are not accompanied by measurement of antigen-specific cells (ASCs). Higher levels of a particular miRNA in cells or supernatants in vaccine responders may simply reflect more ASCs. One study cultured THP-1s in serum from vaccinated participants and stimulated the cells with influenza vaccine. The study revealed an association between THP-1, IL6 and IFNB expression and levels of miR-451a and miR-7704 in serum but did not prove the association was causative[253]. Furthermore, it did not look at whether vaccination influenced miR-451a expression, nor did it link the result to a clinical output, e.g., vaccine reactogenicity or antibody titres[253].

Three studies have examined the influence of vaccination on miRNA expression in blood (Table 3). All were flu vaccine studies. One study found five miRNAs in PBMCs within two weeks of vaccination were associated with haemagglutinin antibody titres—these miRNAs were predicted to target the interferon pathway. Another study found flu vaccination increased serum miR-150 after one to two months and correlated with haemagglutinin antibody titres. Both studies lacked a validation cohort. Another study found an initial association between vaccination and miRNA expression, but the findings did not validate in an independent cohort[254]. Given these limitations, more studies are required to confirm the effects of vaccination on miRNA expression.

Table 3 Vaccine-related miRNA studies–all studies identified in the literature are presented.

Vaccine	Substrate	Method	Findings	ref
Tri-valent split virion inactivated influenza vaccine	Seventy-five adults' PBMCs were taken at baseline and 1,3,7, and 14 days after vaccination.	Microarrays measured mRNA, and the micronome was measured using HTC Edge-seq (probe-based technology).	miRNA signatures changed over time but not after multiple testing. miRNA expression correlated with haemagglutinin antibody (HAI) titres, especially expression of miR-199a, miR-340, miR-944, miR-424, mir-93, miR-96, and miR-296. Expression of miRNAs predicted to target genes promoting the interferon pathway negatively correlated with haemagglutinin antibody titres	[255]
MF59-adjuvanted monovalent 2009 pandemic influenza subunit vaccine	46 adults before and 30 days after a single dose of the vaccine 50 children before and 30 days after second dose of the vaccine	RT-qPCR measured serum miR-150 at each time point	Increase in serum miR-150 after flu vaccination, which correlated with antibody titers. This signal was isolated in the nano-vesicle rather than the microvesicle fraction of serum. Felt to be a memory response as children required 2 doses of vaccine to see the effect. Intracellular miRNAs expression did not correlate with miRNA release	[151]
MF59-adjuvanted monovalent 2009 pandemic influenza subunit vaccine	22 children. Serum was taken before and 28 days after the second vaccine dose	Microarrays measured micronome expression in participant serum before and after vaccination. RT-qPCR was used to test positive findings in an independent validation cohort.	Nineteen microRNAs were differentially expressed after vaccination after adjustment for multiple testing. However, these findings were not replicated when assessed in a validation cohort.	[254]

Table 3 (continued)

BCG	PBMCs for BCG vaccinees	PBMCs before, 2 and 6 months after vaccination. Cultured in the presence or absence of mTB antigens and mTB filtrates. Expression of cellular miR-146a, miR-326 and miR-155 measured using RT-qPCR	correlation analysis revealed a negative association between the *FOXP3* transcription factor and miR-155 and between IL-17 and miR-326	[251]
Measles	PBMCs from 23 MMR vaccine recipients aged 11–40 years who were classified as low and high antibody responders	Sorted B-cells and monocytes infected with measles virus and uninfected control cells. T-cells were incubated with measles-infected monocytes or non-infected monocytes and then purified to ensure that just T-cells were analysed. sRNA-sequencing was used to measure intracellular miRNA expression in uninfected vs infected cells.	21 miRNAs were differentially expressed in B-cells from low versus high vaccine responders	[252]
Hepatitis B virus	10 participants with chronic HepB were given a therapeutic vaccine–5 were vaccine responders, 5 were non-vaccine responders	RT-qPCR of serum and serum-derived hepatitis B surface antigen (HBsAg) particles at baseline 52 and 72 weeks post-commencement of treatment	There was an increase in miR-let7f, miR-22, miR-30a and miR-122-5p in the serum and HBsAg particles of responders	[162]

Table 3 (continued)

Inactivated whole virus influenza vaccine	Serum from participants after vaccination. Plus THP-1 macrophages were incubated with serum from 6 people for 2 days and then exposed to whole cell influenza vaccine	Microarray measurement of miRNAs in each subject's serum and THP-1 supernatant. IFNB and IL6 were measured within the THP-1 monocytes by RT-qPCR.	There was an anticorrelation between 451a and miR-7704 expression in serum and IL6 and IFN-B expression in vaccine-stimulated THP-1 cells.	[253]

1.2.2.6.8 miRNA-based therapeutics for infectious disease

The role of miRNAs in host-pathogen interactions and immunity suggests that miRNA-based therapeutics could be used as adjuncts to standard antimicrobial therapies and vaccines. Agents targeting the host response rather than the microbe may be more resistant to the genetic mutations that render pathogens resistant to microbe-targeting antimicrobial treatments.

The potential for miRNA-based therapeutics in infectious disease has been shown in mice studies of EV-71 infection. EV-71 induces cellular miRNA-146a expression, negatively regulating innate immunity. Inhibition of miR-146a in EV-71 infected mice by intraperitoneal injection of a miR-146a antagomiR improved survival from 25% to 80% by reinstating an IFN-γ response[256]. Another study showed that *S.* Enteritidis infection could be suppressed by administering anti-miR-128 to infected mice[257]. Experimental agents targeting miRNAs have shown mixed success in clinical trials. Administration of miR-122 inhibitors called

Miravirsen and RG-101 are safe and well tolerated, and in the case of RG101, reduced HCV viral load in a phase 1b double-blind randomised controlled[258,259]. However, despite the initial promise, only three out of 28 participants had a sustained antiviral response at 76 weeks. 5'UTR mutations occurring in the HCV genome may account for the viral rebound. Phase II trials investigating whether RG-101 and miravirsen can augment anti-viral HCV therapy have shown promise[260259].

miRNA-augmented vaccines are another possibility. Incorporating miR-155 into an experimental HIV DNA vaccine increased ENV expression and T-cell mediated immunity by miR-155 targeting of antiviral proteins that inhibit ENV expression [261]. Despite this promising result, no follow-up studies have been reported.

1.2.2.7 SRNAs other than miRNAs

sRNA-sequencing identifies short fragments of non-miRNAs. Sometimes referred to as "fragments" or "derived fragments" (e.g. tRNA fragments, tRf's, sRNA-dfs), "halves" (e.g., tRNA halves) or "derived sRNAs" (e.g., sd-yRNA-4). They are generated from various RNAs (

Table 4). sRNA-dfs are not degradation products and function differently from the full-length RNA molecules they derive from. Many appear to play a role in transcriptional or post-transcriptional silencing. Sometimes there is an overlap between sRNA-dfs and miRNAs biogenesis and function, e.g. snoRNA-dfs[262,263,264]. sRNA-dfs have been implicated in disease pathogenesis in the past 2 years as researchers explore non-miRNA fractions within sRNA-seq studies.

Table 4 Source, biogenesis, and function of sRNA-derived fragments

sRNA	Full-length RNA	Function of full-length RNA	Biogenesis of sRNA from precursor/full-length RNA	Function of the sRNA	ref
miRNA	-	-	See Figure 1	Translation inhibition via the RNAi pathway	[15–19]
piRNA	-piRNA precursors -transposable elements (TE) Not evolutionary conserved like miRNAs, sequences evolve quickly	-	Generated from single-stranded precursors in a DICER-independent manner Primary biogenesis: piRNA precursors are cleaved into piRNAs by either being trimmed at the 3′ end by PNLDC1 or by being sequentially cleaved into consecutive mature piRNA sequences in a 5′ to 3′ direction by PLD6. piRNA abundance can be amplified in a PING PONG mechanism as they bind to their target TE mRNAs resulting in TE cleavage by PLD6, PIWIL1 or AGO3. These TE fragments bind to piRNA precursors, causing cleavage into mature piRNAs, creating a cycle of piRNA production.	Production predominantly but not exclusively limited to gametes Bind and suppress retrotransposons at the transcriptional and post-transcriptional level–particularly important in gametogenesis when widespread DNA methylation can lead to the transcription and reintegration of retrotransposons. Regulate protein-coding genes	[264]

Table 4 (continued)

tRNA fragments tRfs tR-halves	Transfer RNAs (tRNA)	Bind amino acids to allow translation of mRNA	Shown to be cleaved by Angiogenin induced during cell stress cleaves tRNAs. Other exonucleases also identified as cleavers of tRNA, including Dicer, RNase 1, RNase L, and Schlafen13/SFLN13. - Depending on where the tRNAs are cleaved, the resultant sRNAs can be described as t-RNA halves or tRNA fragments (tRfs), which can be further described as 3′ or 5′.	- Promote cell survival through their interaction with cytochrome C - Promote stress-induced granule assembly (pro-survival granules that protect cells during stress) - inhibit translation by binding to ribosome or translation factors directly to interfere with their functions - shown to interact with piwi protein to regulate transcription epigenetically - nascent RNA silencing by binding to intronic sequences - one tRNA fragment upregulated in vitro has been shown to promote RSV replication - little evidence that they act like miRNAs	265,266267–271

Table 4 (continued)

snoRNA fragments	Small nucleolar RNAs (snoRNA)	Modify nucleotides in ribosomal RNA, tRNA and snRNA. Three main classes: - SNORA are H/ACA box snoRNA: Pseudouridylation of RNAs - SNORD are C/D box snoRNAs: cleave rRNA, methylate RNAs - scaRNA are small Cajal body-specific RNAs: localise to Cajal bodies in the nucleus, guide methylation and pseudouridylation of spliceosomal RNAs	Some SNORA snoRNA-derived fragments produced by cleavage with DICER FUS (hnRNP) also identified as a possible cleaver of snoRNAs into fragments	Translational repression via RNAi Associate with AGO proteins and may function like miRNAs May be a source of piRNAs	[262] [263]
snRNA fragments	Small nuclear RNAs (often called U RNAs)	Part of the spliceosome. Involved in mRNA spicing.	?DICER ?other endonucleases	Previously identified to be associated with AGO protein ?translational repression, ? splicing	[272,273]

Table 4 (continued)

yRNA fragments	yRNAs	Bind to Ro60 in the ribonuclear protein complex (RoRNP). This complex binds to misfolded RNAs to promote their degradation or facilitate refolding - binding of yRNAs inhibits this process. yRNA is also required for DNA replication, but its role in this is unknown.	RNase 1 Independent of DICER Y-RNA fragments derive from the 5' and 3' arms of the Y-RNA hairpin, though 5' fragments are most abundant. They are around ~21, ~30, and ~40 nucleotides long.	?Translational repression via RNAi yRNA-dfs bound to AGO2 may not post-transcriptionally repress mRNA expression, possibly suggesting they regulate gene expression through a different mechanism.	[274–277] [278]
Vault RNA fragments	Vault RNAs vt-RNAs	5% are associated with vault particles RNP complexes; the rest are free within the cell Known to play a role in drug resistance in cancer, but mechanism of action unclear	DICER NSUN2 RNA cytosine C (5)methyltransferase, RSF2	- Post-transcriptional gene silencing via association with AGO2 - nascent RNA silencing by binding to intronic sequences	[279–281]
Endo-siRNAs	mRNAs, long non-coding RNAs, antisense transcripts	Protein expression, gene regulation	DICER	Cleavage of copies of its full-length RNA via RNAi	
rRNA fragments	Ribosomal RNAs	Translate mRNAs	Can bind to AGO	Some evidence that one rRNA fragment can target mRNAs	[273,282]

1.2.2.7.1 Non-microRNA sRNAs in infection and immunity

There are tentative signs that sRNA-dfs participate in immune responses. yRNA-dfs are abundant in extracellular vesicles and are generally associated with immune activation, occasionally through TLR3 and TLR7 activation[275]. *In vitro* cellular infection with RSV, measles, mycobacterium tuberculosis, and Listeria monocytogenes induce tRNA-dfs, indicating a potential role in innate cellular immune responses[269,283,284]. *In vitro* RSV infection upregulates 5'-tRNA halves due to Angiogenin cleavage of full-length tRNAs[269,271]. 5' tRNA-Glu-CTC, tRNAs-Gly-CCC and tRNA-Lys-CTT fragments promote RSV replication [269,271,270]. 3' tRF5-Glu-CTC fragments supress APOER2 expression by binding its 3'-UTR[269,271,270].

1.2.3 Technical aspects of miRNA and sRNA studies

1.2.3.1.1 Detecting and measuring miRNAs

Figure 3 contains a schema for conducting miRNA studies. Gold standard miRNA assays are northern blots and reverse-transcription quantitative PCR, i.e., real-time PCR (RT-PCR). However, these methods are low throughput, primer/probe design can be challenging due to the short nature of miRNAs, and the high homology between miRNAs of the same family can impact specificity. Primers, including hydrolysis probe-based reporter molecules (e.g., TaqMan) or highly specific locked nucleic acids (LNA-based primers), increase specificity but may miss isomiRs. Another issue is normalization as no standard housekeeping genes exist. Suggestions from the literature include miR-103a-3p, miR-191-5p, miR-423-5p, miR-16-5p, miR-425-5p, miR-93-5p and miR-451a and RNU-6, but the tissue/cell type and conditions/disease state influences their expression making it hard to predict which sRNAs are best for a particular study[285]. Spike-in RNAs

(e.g. cel-miR-39 from *C.* elegans) offer another normalisation option, particularly for biofluids. Ideally, sRNA normalisers are selected based on miRnome data (e.g., RNA-seq data) generated from a study that closely reflects the experimental set-up, but this is often unavailable. In lieu of this, several potential housekeepers may be selected based on the literature, and then programs such as Best Keeper, geNorm or Normfinder can be used to identify the best normaliser from RT-PCR data[286–288]. Normalizing against the average expression of several putative "housekeepers" is most robust.

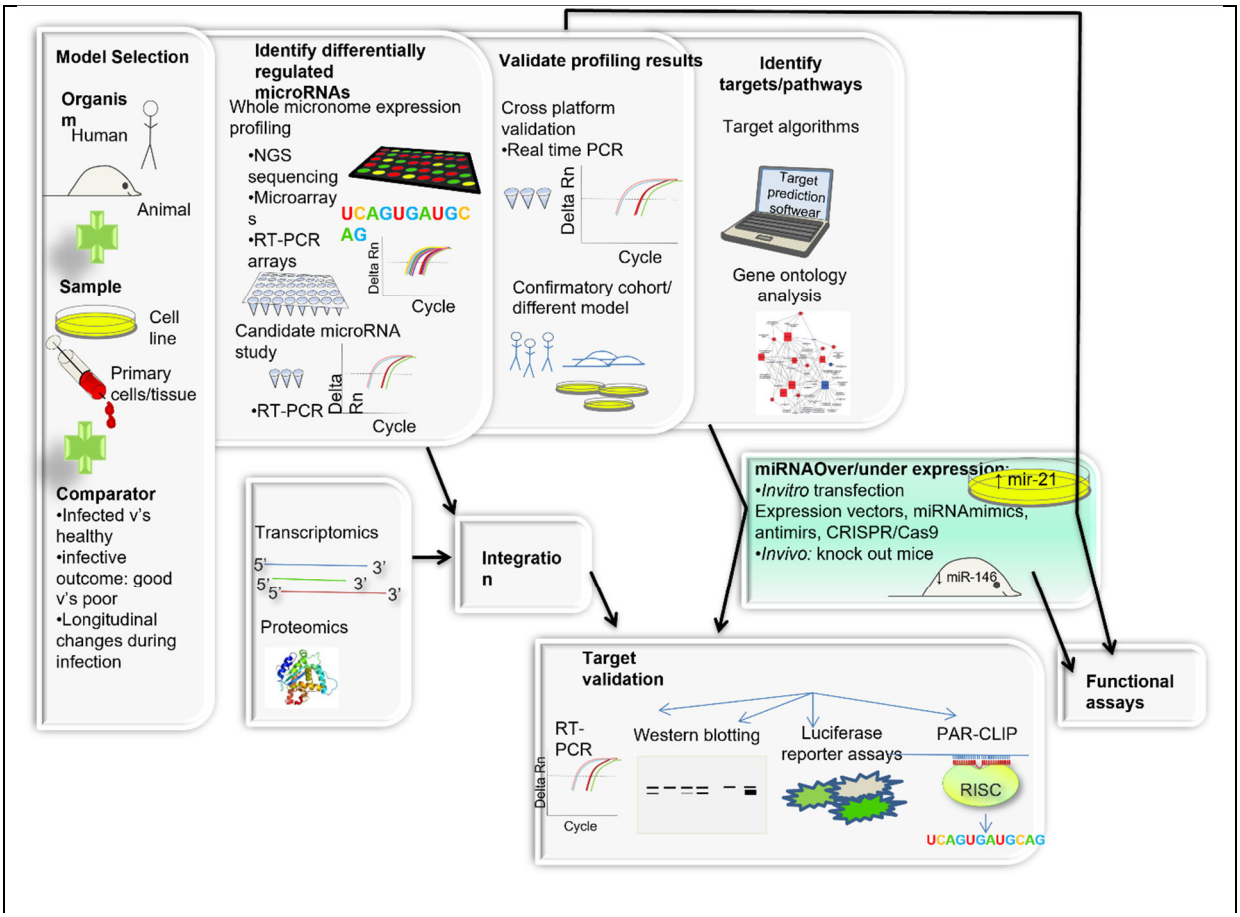


Figure 3 Common steps in the investigation of miRNAs.

High-throughput assays, e.g., hybridization-probe-based microarrays, PCR arrays, and sRNA-seq, offer greater efficiency and are more agnostic to which miRNAs are measured. A drawback to microarrays is potential non-specific binding[254]. RT-PCR arrays do not require technical validation but are restricted to the miRNAs present on the arrays. Furthermore, arrays may not capture isomiRs because the probes/primers are based on "canonical" sequences in miRbase. These sequences are not necessarily the most abundant so arrays can fail to measure the most abundant version of a miRNA[113–115].

sRNA-seq captures a broader miRNAome and other sRNA-dfs and novel miRNAs/sRNA-dfs. Its drawback includes more complicated data processing– e.g., dealing with multi-mapping reads, technical noise and data normalisation. There has been a concerted effort to standardize sRNA data processing, but there is currently no standard method, and each published pipeline has limitations (see General Methods Chapter 2)[289–291]. A technical limitation of sRNA-seq is sequence-dependent ligation and amplification, which means some sequences are overestimated, and others are underestimated or not sequenced. This issue is revealed by changes to the micronome when different adapter sequences are used. Inclusion of random nucleotides into the end of the adapters mitigates some of these issues[292]. Despite its limitations, sRNA-seq is the best approach for surveying the sRNAome.

1.2.3.1.2 miRNA target prediction

A challenge in miRNA research is identifying a miRNA's mRNA targets. One approach uses bioinformatic prediction tools such as miRanda, miRDB, PITA,

DIANA microT-CDS, and TargetScan (Table 5). These tools predict miRNA binding sites *in silico* by looking for complementary sequences between a miRNA's seed region and an mRNA's 3' UTR. To reduce false positives, information on other factors is incorporated into algorithms, including the number of MREs within a 3'UTR, the energetic stability of a miRNA:mRNA pairing (enhanced by additional complementarity between a 3'UTR and miRNA outside of the seed site and MRE accessibility given mRNA secondary structure), and evolutionary conservation of putative MREs across species (Table 5)[293]. Different target prediction algorithms produce different results; consequently, some tools take a meta-prediction approach, integrating the output of miRanda, miRDB, PITA, and TargetScan[294].

Bioinformatic prediction of miRNA:mRNA binding pairing does not mean it occurs *in vivo*. This is due to limitations of target prediction algorithms and because miRNA-mRNA binding is context specific, dependent on concomitant expression and competition between targets[303,304]. To circumvent this, databases of experimentally validated targets exist, mined from literature, e.g., miRtarbase, tarbase and miRecords[305–307]. These also have limitations with results depending on the cell types and cellular states of the experiments in the database and the quality of experimental validation (Table 6). Ultimately, there is no single approach for identifying mRNAs-miRNA interactions: bioinformatics and experimental databases are both used in the literature.

Table 5 microRNA target site prediction software/databases and the algorithms underlying them

Site (URL)	Principles of the target site prediction algorithm	ref
Targetscan (https://www.targetscan.org/vert_80/)	• Searches for conserved 8mer, 7mer, and 6mer sites that match the seed region of each miRNA and sequences with 1 mismatch to seed sequences that have compensatory complementarity to a base outside of the miRNA seed sequence. • Sites also scored using features from a biochemical model of miRNA-mediated repression based on experimental data. • Identification of MREs in open reading frames of mRNAs also provided.	295, 296
DIANA microT-CDS (http://www.microrna.gr/webServer)	• Algorithm developed using features identified in machine learning that are present in miRNA-mRNA target interaction identified using high-throughput immunoprecipitation and sequencing data • Identifies MREs in 3'UTRs and coding sequences	297
miRDB (http://www.mirdb.org/)	• Uses the MirTarget algorithm built using machine learning methods that learnt common features associated with miRNA binding and target downregulation through analysing miRNA-target interactions from RNA-sequencing experiments.	298, 299
miRanda (http://www.microrna.org/)	• seed site complementarity • Conservation • frequency of AU in the region flanking the target • target site accessibility based on secondary structure of the mRNA and the binding energy of the miRNA–target duplex • length of the UTR • Relative position of the potential MRE in the UTR ends	300, 301
PITA (http://genie.weizmann.ac.il/pubs/mir07/)	• Estimates seed sequence complementarity and G-U wobble • target site accessibility based on the energy difference between the free energy from forming miRNA-target complexes and energy cost of unpairing the target for miRNA accessibility	302

Table 6 MiRNA:mRNA target validation methods

miRNA:mRNA target validation experiment	**Method**	**Advantage**	**Disadvantage**
Measurement of mRNA/protein expression after manipulation of a miRNA	Knock up or knock down miRNA within a system and measure mRNA or protein expression by high or low-throughput methods	Simple	cannot differentiate direct and indirect targeting
Biotinylated arrays	mRNA 3'UTR probes are printed onto an array, and the miRNA binds to putative MREs	High throughput prediction	Non-specific binding between the miRNA and mRNA 3'UTR on the array, conformational shape of the 3'UTR may be altered.
Luciferase reporter	Wild type and mutant forms of a predicted mRNA target cloned downstream of a firefly luciferase reporter. This is transfected into a cell, and the putative targeting miRNA levels manipulated and changes in luciferase monitored—a reduction in luminescence indicates miRNA mediated post-transcriptional regulation.	Very specific, high-quality evidence shows functional effect	Low throughput
AGO HITS-CLIP	high-throughput technique involving sequencing of AGO-bound regions of mRNAs and miRNAs but relies on bioinformatics to "pair up" the AGO-bound mRNA sequences and miRNAs to infer mRNA:miRNA binding sites.	high-throughput technique involving sequencing of AGO-bound regions of mRNAs and miRNAs	Still relies on bioinformatics to "pair up" the AGO-bound mRNA sequences and miRNAs to infer mRNA:miRNA binding sites.
CLASH and clear-CLIP-based methods	Methods that cross-link miRNAs to mRNAs prior to sequencing	unambiguous identification of miRNA bound to mRNAs.	Recent development, so few studies exist, and a costly approach.

1.2.3.1.3 Functional studies of miRNAs

After identifying miRNAs of interest, miRNA levels are often manipulated *in vitro* or *in vivo*, and phenotypic effects studied. Upregulation can be achieved through miRNA mimics or expression vectors. miRNA mimics are synthetically produced miRNA copies passively transfected into the cell/organism, e.g., encapsulated in lipid nanoparticles. One downside of this approach is that supra-physiological levels of a miRNA can occur, producing off-target effects, and levels gradually reduce due to processing, and cell division. Stable expression can be achieved through expression vectors, e.g., lentiviral vectors. Some studies introduce pre-miRNAs, which may function more akin to wildtype miRNAs due to incorporation higher up in the RISC processing pathway. MiRNA reduction/knockdown can be achieved through anti-miRs or antagomiRs—oligos with perfect complementarity to the miRNA of interest. Binding of an anti-miR/antagomiR sequesters the miRNA from the RISC pathway, causing its degradation. Scrambled mimics and antagomiRs must be included as negative controls, as introducing foreign genetic material can induce changes within cells and organisms. Recently, CRISPR has been used to create knock-out cell lines and germline knock-out organisms[308,309]. Inducible knock-out transgenic mice help isolate the effect of a miRNA in a specific context[308,310].

1.2.3.1.4 miRNA expression atlases

The largest miRNA expression atlas is the fifth edition of the Functional Annotation of Mammalian Genome project (FANTOM5), which contains miRNA count data for 369 human cell types[311]. Other miRNA atlases are the miRNA Tissue atlas (latest version; miRNATissueAtlas2, accessible at https://ccb-

web.cs.uni-saarland.de/tissueatlas2/), the "Mammalian microRNA Expression Atlas" published by Landgraf et al., a microRNA catalogue of human peripheral blood published by Juzenas et al. (accessible at http://134.245.63.235/ikmb-tools/bloodmiRs/) and a human miRNA repertoire of different blood compounds from Leidinger et al.[312–315].

1.2.4 Critical knowledge gaps in the Literature

Thus far, evidence has been presented that biomarkers of vaccine reactogenicity and correlates of protection will help enable rational vaccine design and aid selection of vaccine candidates and parameters in preclinical and early-phase clinical trials. Furthermore, a better understanding of immune responses could facilitate the design of new diagnostics and treatments. The paucity of studies on paediatric populations who receive the bulk of vaccination must be addressed. The role of miRNA and sRNA-dfs in vaccination studies is an emerging field of interest as a more holistic understanding of the immune response is sought. This thesis will address this knowlegde gap.

1.3 Using vaccine and infection studies to interrogate the role of sRNAs in vaccination and infection.

Three clinical studies will be explored in this thesis: an infant vaccine study involving the reactogenic 4CMenB vaccine, a typhoid vaccine challenge study, and a ChAdOx1 nCoV-19 vaccine RCT. These studies cover sRNA responses across various ages, vaccines, and infections, providing a broad view of sRNAs in immune responses. All studies have associated mRNA-sequencing data and,

depending on the study, other modalities, e.g., proteomics and CyToF. This allows integration of sRNA-seq data with information on expression of miRNA targets and the transcripts that sRNA-dfs derive from. The background to each study is presented in full in the relevant results chapters, but a brief explanation is provided here for context.

1.3.1 The EUCLIDS study: exploring how vaccination influences sRNAs and proteins in plasma, and the links to vaccine reactogenicity

EUCLIDS was a randomised, descriptive, open-label study exploring the relationship between whole blood gene expression and vaccine reactogenicity and immunogenicity in 4- and 12-month-old infants. The study compared children receiving routine childhood vaccines with or without the 4CmenB vaccine. Including 4CmenB in infant vaccine schedules increases reactogenicity, but the reasons are unknown at the molecular level. The study explored the immune response early after vaccination and compared that response between vaccine regimes with different reactogenicity[76].

1.3.2 Vaccines Against Typhoid (VAST): measuring sRNAs in PBMCS to dissect the immune response to vaccination, identifying correlates of protection, and exploring the effect of *S* Typhi exposure and infection.

VAST was a Phase IIb, observer-blind, randomised controlled trial to assess the immunogenicity and protective efficacy of a Vi conjugated (Vi-TCV) and an unconjugated polysaccharide (Vi-PS) vaccine in preventing typhoid using a

challenge model of *Salmonella* Typhi infection in adults. PBMCs were available for sRNA-seq at baseline and after vaccination, challenge, and diagnosis in this study. This offered the ability to evaluate sRNAs expression within host immune responses longitudinally, from activation of the early innate immune response through to the development of adaptive immune responses, and potential vaccine-mediated immune responses to *Salmonella* Typhi exposure and infection.

1.3.3 The COVID-19 Vaccine and Infection multi-omics study (CoVI-multi omics study): studying the influence of COVID-19 infection on whole blood sRNA expression in the context of SARS-CoV-2 vaccination.

The COVID-19 Vaccine and Infection multi-omics study (CoVI-multi omics study) used samples from RCTs of the ChAdOx1 nCoV-19 vaccine. In this study, participants received either the ChAdOx1 nCoV-19 vaccine or a control vaccine (MenACWY) and were followed up to see if they developed COVID-19. sRNA-seq was done on whole-blood samples taken at baseline and during COVID-19 and non-COVID-19 infection. Inferring gene regulation by miRNAs enabled real-time assessment of innate and adaptive vaccine-induced immunity against SARS-CoV-2.

1.4 Research Questions

The above studies address three research questions

1. Does vaccination and infection alter miRNA expression?

2. Can miRNA expression predict or shape the response to vaccination and/or infection?

3. What is the role of miRNAs in the immune response to infection/vaccination?

1.5 Thesis Aims and Objectives

The aim of this thesis is to characterise the influence of vaccination and infection on sRNAs in human vaccine and challenge studies to answer the research questions posed above. The central objectives of this thesis are:

1. To use sRNA-seq to determine changes in sRNA expression before and after vaccination and/or infection across a range of sample types using blood samples from bacterial and viral vaccine and infection studies.
2. To identify potential biomarkers of vaccine reactogenicity and correlates of protection
3. To identify the pathways that differentially expressed miRNAs collectively target
4. To integrate sRNA data with mRNA and other available datasets to identify how miRNAs relate to feature expression in other datasets
5. To set up an *in-vitro* model to identify the source and role of differentially expressed miRNAs

2 General Methods

2.1.1 Sample acquisition

Clinical samples were obtained from registered clinical trials approved by research ethics committees and health regulatory authorities. Samples from the EUCLIDS study were retrieved from the Oxford Vaccine Centre biobank (IRAS ID 196396, REC reference 16/SC/0141) and analysed per OVC Biobank protocol. Samples were only used if a participant had provided written consent for storing and analysing their genetic material.

2.2 SRNA-seq

2.2.1 RNA quantification

For EUCLIDS and VAST, the concentration of RNA from cellular samples was quantified using the Qubit™ RNA HS and BR Assay Kits (Thermo Scientific, cat #:Q32851). For the COVID multi-omics study, concentration was measured using the Agilent Small RNA kit (cat#5067-1548) on the Agilent 2100 Bioanalyzer 6. RNA purity was measured on a Thermo Scientific NanoDrop Spectrophotometer at OD 260/280 and 260/230.

2.2.2 sRNA Library Preparation.

RNA was submitted to the Oxford Genomics Centre, Welcome Trust Centre for human genetics, University of Oxford, for RNA concentration (where relevant), library preparation and sequencing.

Forty microliters of RNA from EUCLIDS plasma samples were concentrated 8X (using Agencourt Ampure X beads) before library preparation.

Using the low-input protocol, sRNA libraries were prepared using the NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) kit (cat#E7330).

2.2.3 sRNA-seq

sRNA-seq was conducted on Illumina HiSeq2500 Rapid flow cells. For EUCLIDS and VAST, libraries were pooled in groups of 44 and sequenced across two flow cells at an expected depth of 5 million reads. For the COVI multi-omics study, libraries were pooled in groups of 61 and sequenced across four flow cells at an expected depth of 10 million reads. Read length was 50 nucleotides. Strand information was preserved. Random pooling of samples was stratified by participant number where possible to reduce technical noise in paired analyses.

2.2.4 Data processing and QC

Adapter-trimmed FASTQC files received from the Wellcome Trust were submitted to the sRNA alignment pipeline (section 2.5). Noise thresholds were determined empirically by plotting density log counts per million (lcpm) at different filtering thresholds to identify the lowest CPM cut-off that produced a near unimodal distribution of reads for most samples. Features with a CPM above this threshold in at least half the smallest experimental subgroup group were kept. The

exception to this was stage I of the COVI multi-omics study, where a sample size of 6 was used because there were so few samples per group. Data were normalised using quantile normalisation or the trimmed mean of M-values (TMM), depending on which method generated the best QC plots. TMM was used where possible as it accounts for library size variation without being affected by changes in the expression of highly abundant features. Potential outlying samples were identified using PCA, log ratio-mean average (MA-plots), boxplots of log-counts per million and density lcpm diagrams on normalised data.

2.2.5 Statistical analysis

All analyses were done in R unless otherwise stated[316]. Supplementary Table 1 lists utilised packages.

2.2.5.1 Differential Expression Analysis

2.2.5.1.1 EUCLIDS

Differential expression analysis of RNA-seq data was done using the EdgeR package, which implements a generalised linear model likelihood ratio test[317,318]. This was done as there were very few subgroups. Gene dispersions were estimated using the estimateGLMRobustDisp function.

2.2.5.1.2 VAST and COVI multi-omics

Differential expression analysis in the VAST and COVI multi-omics studies was done using the limma package, which fits a linear model to each gene[319–322]. Limma was used because it is easier to implement for many subgroups.

Expression values were transformed using the voom function(). Data were modelled using mixed random and fixed effects. Random effects made it possible to account for multiple samples from the same participants.

2.3 Frequently used statistics

Wilcoxon or t-tests (paired where possible) were used to compare sample medians or means based on whether data points were normally distributed. Spearman or Pearson correlation coefficients were used to estimate correlations depending on whether the relationship between the variables appeared linear (Pearson) or only monotonic (Spearman). Unless otherwise stated, results were subject to multiple correction using the Benjamini-Hochberg multiple testing correction (FDR) with a cut-off of 0.05 indicating significance[323].

2.4 Public Data Availability

Data will be deposited in the NCBI Gene Expression Omnibus (GEO) repository after completion of analyses for publication prior to manuscript submissions.

2.5 Development of an RNA sequencing alignment pipeline

There is no standard sRNA-seq alignment and assignment pipeline. Key issues with sRNA-seq data compared with mRNA-seq:

- The short nature of miRNAs and sequence homology within a miRNA family means multi-mapping sequences are common (Figure 4); this is especially true if they arise from repetitive sequences.

- miRNAs can be found within sequences of other genes; for example, mirtrons lie within introns.
- isomiRs and nucleotide editing of miRNAs mean miRNA sequences can mismatch to the canonical miRNA sequence listed in miRbase.

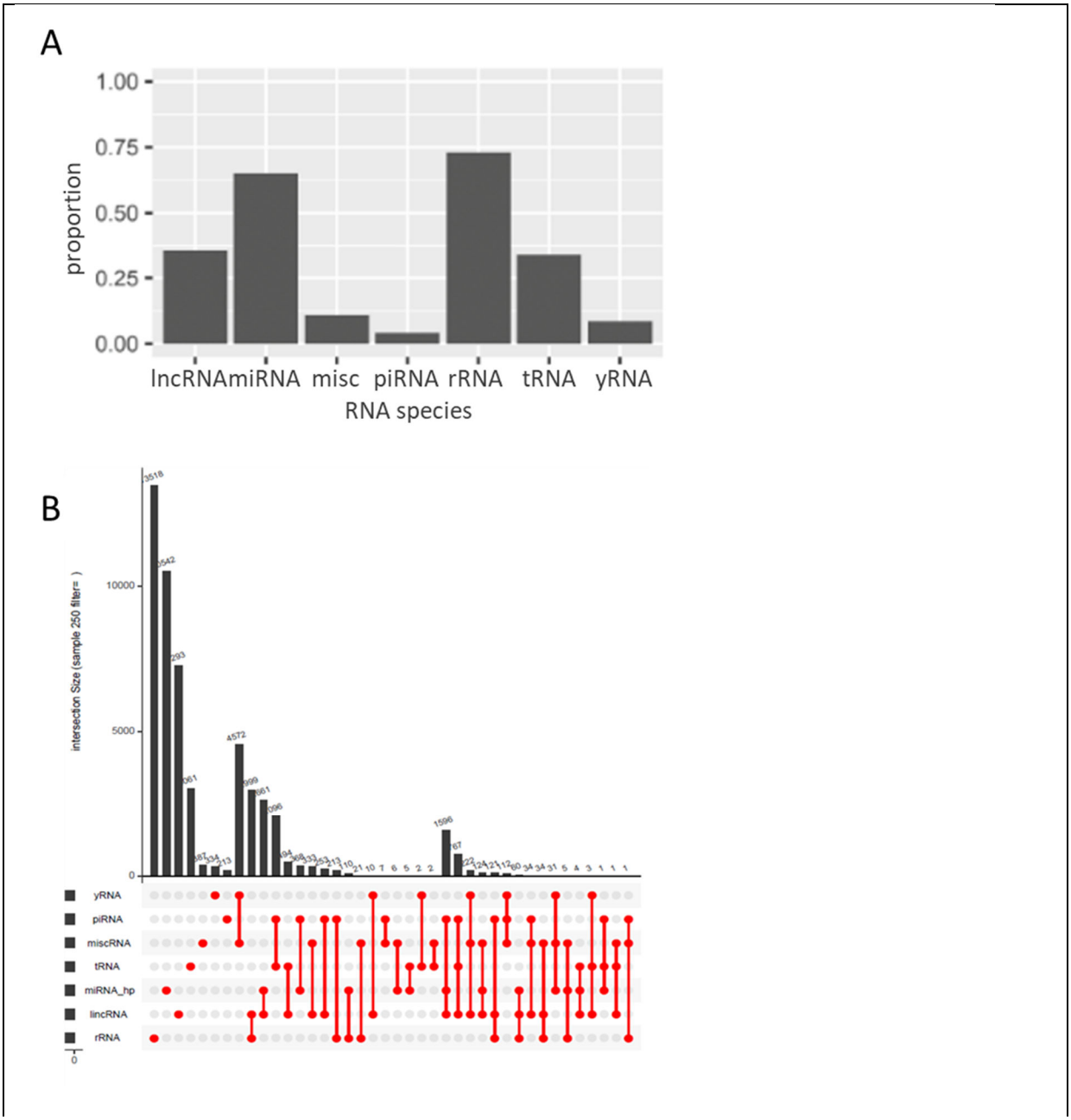


Figure 4 The overlap between different classes of RNAs at the sRNA-df level. Many sequences multimap to multiple RNAs. A. The proportion of sequences within an RNA class that do not map to another class. B. Upset plot, showing overlaps at a finer level.

2.5.1 Considerations for an sRNA alignment and assignment pipeline:

2.5.1.1 Alignment:

Two approaches can be used: align to the genome and then assign using a program such as feature counts and annotation files (e.g. GTF files), which may or may not be restricted to the sRNAome. Alternatively, align to the sRNAome directly.

2.5.1.2 Dealing with multi-mapping reads mapping equally well to several loci

General reference for this section is Deschamps-Francoeur et al. [324]. Options for dealing with multi-mapping reads, along with advantages and disadvantages, are summarised in Table 7.

Should reads mapping to different RNA classes be dealt with differently?

- Yes—miRNAs should be prioritised as they are biologically functional units with more evidence for existence compared with other putative sRNA-dfs (which could represent degradation products) and can exist within other genes: solution—use a hierarchical method to prioritise assignments based on full-length RNA species
- No—in the absence of certainty, assumptions should not be made about which full-length RNA a sequence derives from, particularly reads that map to non-miRNA loci.

Table 7 Options for dealing with multi-mapping reads

Option	**Explanation**	**Programs implementing this approach**	**Advantages**	**Disadvantages**
Ignore	Remove multi-mapping reads	Option in algorithms within star and HTSeq-count	Simple	- Not appropriate for sRNA seq data as multi-mapping sequences comprise a significant proportion of reads, and some features are inherently multi-mapping; therefore, they would be completely discarded
abundance-based approach	Assign to the most abundant read	A similar approach is implemented in Cufflinks[325]	- Removes redundancy, meaning fewer statistical tests are done, potentially improving power in differential expression analyses. - Based on Bayesian thinking—if there are genes, one is very abundant, and the other is less abundant, then an ambiguously mapping read is more likely to come from the abundant reads.	- An assumption is made about where that reads maps to - Alternative loci are not represented in the final expression matrix

Table 8 (continued)

fractional approach	Assign reads to all possible loci but divide the frequency of each read by the number of times it maps.	Commonly used method implemented in the feature counts in the subread package[326].	- This does not assume mapping and allows all possible loci to be represented in the final expression table.	- Many more features are present in the final dataset than were probably sequenced, exacerbating the multiple testing issue - There is redundant information between genes that share reads. This leads to excessive tests, worsening the multiple testing issue with potentially little gain, particularly if none of those loci is differentially expressed (as it does not matter where their sequences map). - If a gene is differentially expressed, it is not immediately clear whether it is genuinely differentially expressed or whether reads from another locus could drive differential expression. - Count data are not integer, which some differential expression packages require - odd dispersion values can arise, which some differential expression packages need to model

Table 9 (continued)

sequence level approach	Perform gene expression at the sequence level		- It removes the ambiguity around loci mapping. This can work in sRNA-seq data because the whole RNA is sequenced, meaning identical sequences can be picked up multiple times. This could not be implemented for mRNA data because random, unique, partial fragments of mRNA genes are sequenced. - It enables differential expression at the isomiR / sRNA "iso-feature" level, which could hold biological meaning.	- Excessive features included in the differential expression analysis—this causes issues with computational time and memory. It can exacerbate the multiple testing problem without bringing many benefits as there can be redundancy between isomiRs/ sequences from the same gene as their expression may be highly correlated
gene_union approach (sometimes referred to as a gene cluster approach)	A sequence, which maps to several genes, e.g., gene A & gene B, gets assigned to a new feature called "geneA_geneB"	Similar approach used in seqCluster[327] and mmQuant[328]. This approach is rarely used in RNA sequencing but common in mass spec proteomics data processing pipelines, e.g. Max Quant software	- All potential loci are represented in the gene expression table. Feature names retain information on mapping certainty, so if feature A_B is differentially expressed, it is clear that this sequence could arise from either or both genes. - It may reduce the number of features tested compared with the fractional approach. - It allows greater specificity about, which loci may be differentially expressed without resorting to differential expression at the sequence level. - Potentially a compromise between the specificity of the sequence level approach, the reduction in redundancy of the abundance-based approach, and feature retainment of the fractional approach.	

2.5.2 Final pipeline

There is no standard (or perfect) way to deal with multi-mapping reads. This can lead researchers to remap sRNA expression data using multiple approaches, which is time-consuming. A new, flexible way of mapping sRNAs was therefore created. This algorithm aimed to:

- Assign reads using the main assignment options within one programme so the user can easily compare the outputs without remapping.
- Assign miRNA sequences to mature reads rather than hairpins if the read only extended slightly into the hairpin region.

Figure 5 contains an overview of the final alignment pipeline. Adapter-trimmed FastQ files are input to the pipeline, and files from the same samples are merged. Short sequences and low-complexity sequences are discarded using fqtrim[329]. Next, reads are mapped to pre-miRNAs and mature miRNAs with 2MM allowed. In the case of mature miRNAs, soft clipping is permitted at either end, allowing reads that extend beyond the canonical mature miRNA sequence to map to the mature miRNA.

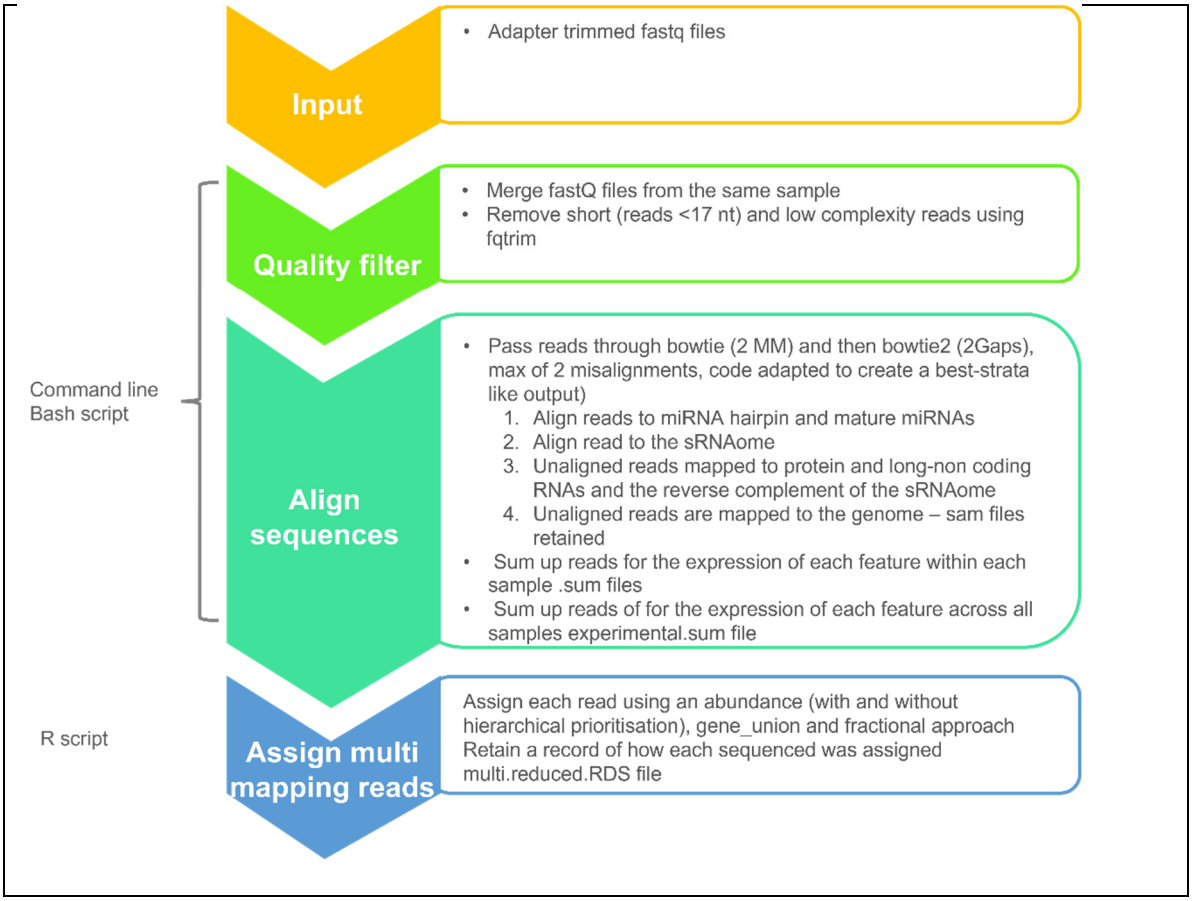


Figure 5 Overview of alignment and assignment pipeline.

Next, all reads are mapped to the sRNAome, including pre-miRs again. The sRNAome bowtie indexes were built by merging FASTA files from:

miRbase[21] (miRNAs)

SnOPY[330] (snoRNAs)

trRNA_db[331] (tRNAs)

RefSeq[332] (snRNAs, yRNAs, vault RNAs, lnc RNAs, ribosomal RNA)

These databases were chosen empirically because they assigned the most reads using the EUCLIDS data. piRNAs are excluded because no high-quality piRNA repositories exist. miscRNAs are excluded as they are a superset of more specific classes (Figure 4).

Sequences are mapped to the sRNAome using Bowtie[333]. Up to 2 mismatches are allowed, and the best-strata option is implemented. The best-strata option only retains assignment(s) with the least number of mismatches. Unaligned sequences are then mapped using Bowtie2, which allows gapped alignment[334]. Bowtie2 does not have a best-strata alignment option. Therefore, options in the Bowtie2 alignment function are tweaked, and assignments filtered using information in the CIGAR string so that the final output is analogous to Bowtie's best-strata option.

Unaligned sequences are then mapped to the protein-coding and anti-sense transcriptome (the reverse complement sequences of all genes) using the same process. Reverse complement sequences were mapped as endo-siRNAs, and piRNAs biogenesis can generate reverse complement sequences. Unmapped reads were then aligned to the genome (GRCh38) using prebuilt Bowtie and Bowtie 2 indexes[335,336]. The output of Bowtie and Bowtie2 is .sam files. These files were processed into ".sum" files containing each unique sequence's sum total (Figure 6).

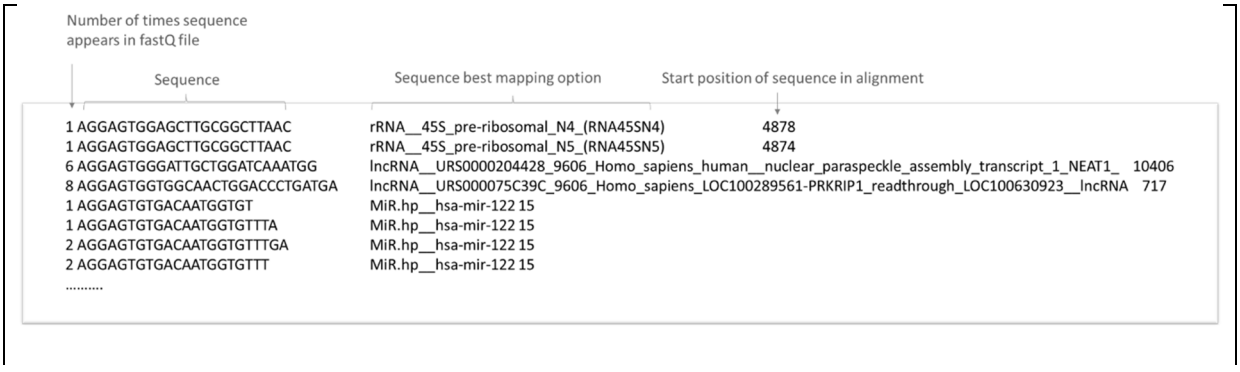


Figure 6 Example of a .sum file:

2.5.3 **Assignment**

With a hierarchical assignment approach, multi-class mapping reads are assigned to only one sRNA class using the following order: Mature miRNA>miRNA hairpin>yRNA>tRNA> snRNA>snoRNA>vault>viral miRNA hairpins>lncRNA>rRNA. Prior to assignment, sequences that map to a mature miRNA but not a miRNA's hairpin are discarded.

The final output files are gene expression matrices collapsed using the following options:

- Abundance based approach
- Abundance-based approach with hierarchical RNA class prioritisation (used for this thesis)
- Fractional counts
- Gene union counts

Each collapse option produces two expression matrices—one with ribosomal rRNAs retained and the other ribodepleted. The pipelines have been made user-friendly so that only folder names containing sample files need to be supplied to the alignment and assignment program. User options are contained in Table 10. The output is a folder containing gene expression matrices for each collapse option. A multi.reduced.RDS file is also created, which contains a record of how each sequence was assigned in each collapse option. This file can be used as a reference when reviewing differentially expressed sRNAs if the user wants to double-check where else reads may map to.

Table 10 Options that can be implemented in assignment

Option	**Method**
Discard reads that map more than a user threshold defined cutoff.	Removes highly multi-mapping reads that arise from repetitive sequences
Collapse genome mapping reads.	This maps non-sRNAome mapping reads to 100bp sliding windows along the genome
Generate gene expression profiles (also known as presence plots).	Draws sample-wise expression profiles along the gene. Allows users to see whether reads map with a random uniform distribution or at discrete loci (see Figure 7 for examples). The former suggests the feature could be a degradation product, and the latter suggests specific cleavage by an endonuclease, possibly indicating that the feature has been processed systematically and thus has a biological function.

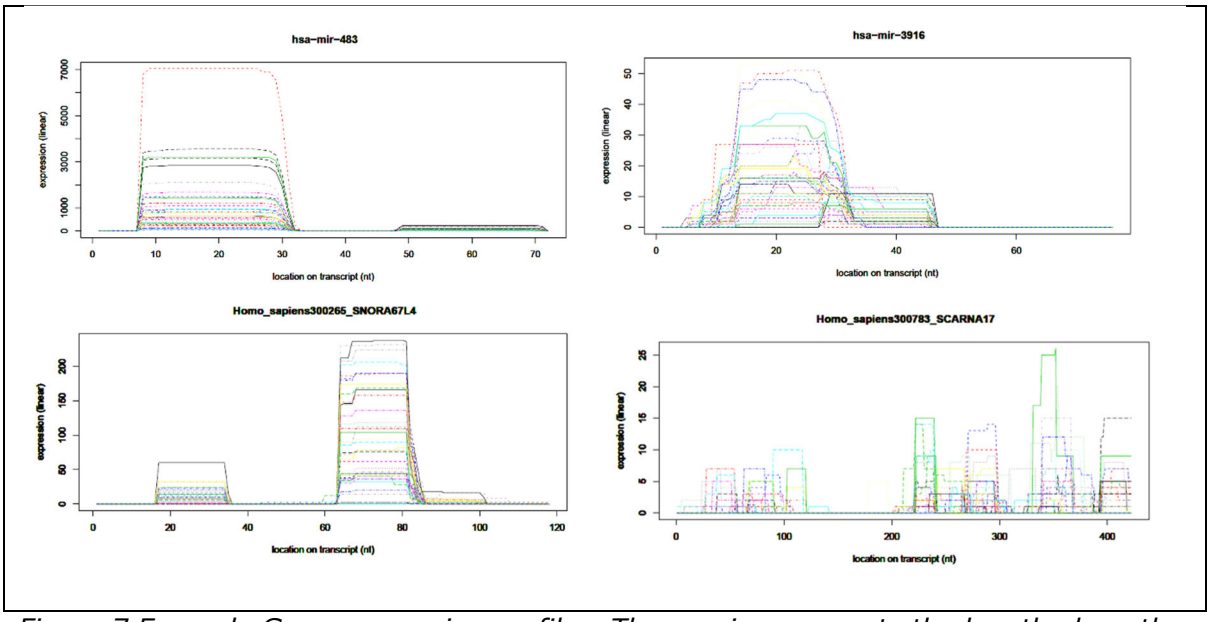


Figure 7 Example Gene expression profiles. The x-axis represents the length along the transcript. Nucleotide 1 = the first place that at least one read maps to. The y-axis represents the number of times each nucleotide in the RNA sequences was identified amongst all reads for a given sample. The expression profile for each sample is plotted in a different colour.

2.5.4 Assignment option used in this thesis

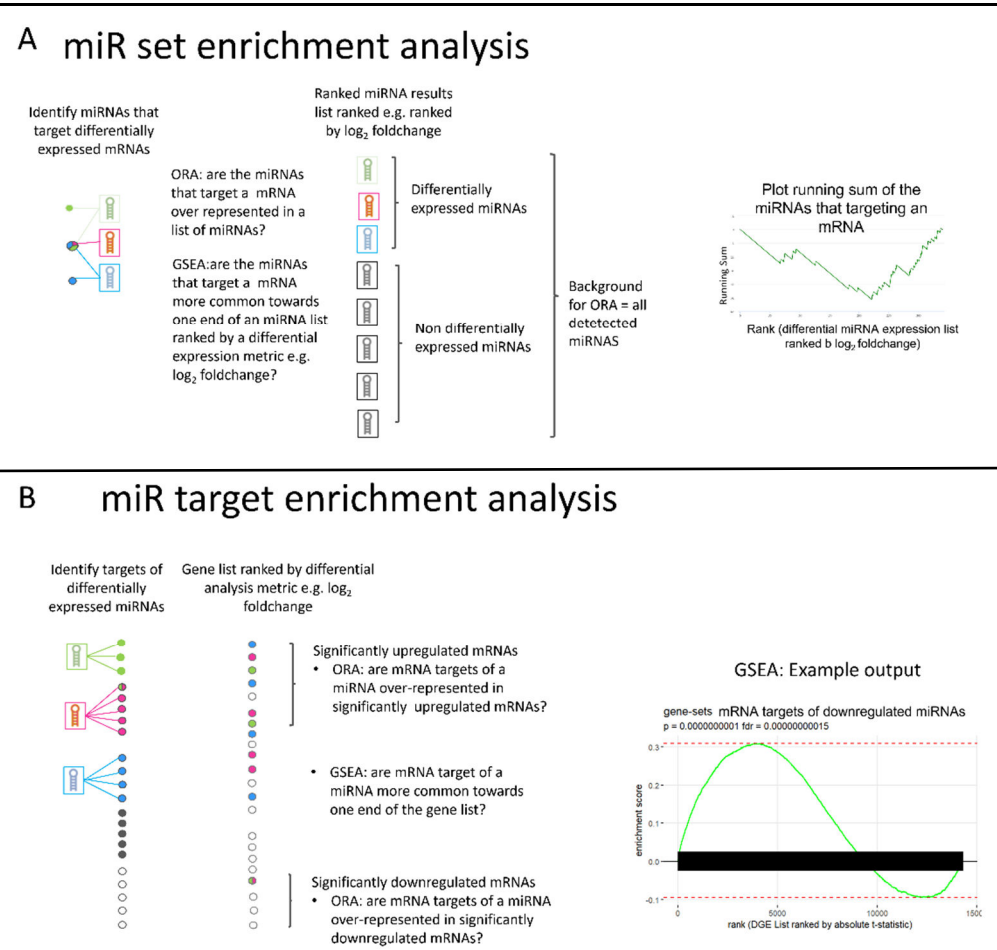
The abundance-based approach with hierarchical RNA class prioritisation was used for all studies in this thesis based on the pros and cons of all options. Non-sRNAome mapping reads were discarded to cut down on the number of features tested.

2.6 MiRNA-mRNA target interactions

Lists of experimentally validated miRNA-mRNA targets were downloaded from miRNet, collating information from three databases. miRNA-mRNA targets exclusively validated by poor-quality methods (e.g. biotinylated arrays) were filtered out.

2.7 miRNA enrichment analyses

MiRNA enrichment analyses predict whether changes in miRNA expression will have a suppressive or permissive effect on a particular mRNA/protein or pathway. Unless otherwise specified, results were deemed significant at FDR<0.05.



- **Positive enrichment**: mRNA is targeted more by a group of miRNAs (ORA) or miRNAs towards the top of a ranked list (GSEA) than expected by chance.
- **Negative enrichment**: mRNA is less targeted by a group of miRNAs (ORA) or miRNAs towards the bottom of a ranked list (GSEA) than expected by chance.

Interpretation: If the miRNA list is ranked from most upregulated to most downregulated:
positive* enrichment** suggests the mRNA's expression (or its protein) will go ***down due to more miRNA targeting,
negative* enrichment** suggests the expression of the mRNA (or its protein) would go ***up due to less miRNA targeting.

- **Positive enrichment:** miRNA(s) target more mRNAs within a group (ORA) or towards the top of a ranked list (GSEA) than expected by chance.
- **Negative enrichment:** miRNA(s) targets fewer mRNAs within a group (ORA) or towards the bottom of a ranked list than expected by chance

Interpretation: If the mRNA list is ranked from most upregulated to most downregulated, then this approach predicts that ***upregulated*** miRNAs have targets which are ***negatively*** enriched because its targets are subject to increased repression. ***Downregulated*** miRNAs have targets which are ***positively*** enriched because its targets are released from repression.

Figure 8 mRNA and miRNA enrichment analyses. Circles = genes. Circle colour indicates which miRNA(s) targets, which gene. A: miRNA set enrichment analysis: an mRNA is converted to the set of miRNAs that targets it, and this set is compared against a group of miRNAs (e.g., differentially expressed miRNAs) or list of miRNAs, e.g. ranked log_2 foldchange. B: miRNA target enrichment analysis takes a miR (or a group of miRNAs) and converts them to the mRNAs that they target, and compares them against a group of mRNAs (e.g., differentially expressed mRNAs) or list of mRNAs, e.g., ranked by log_2 foldchange.

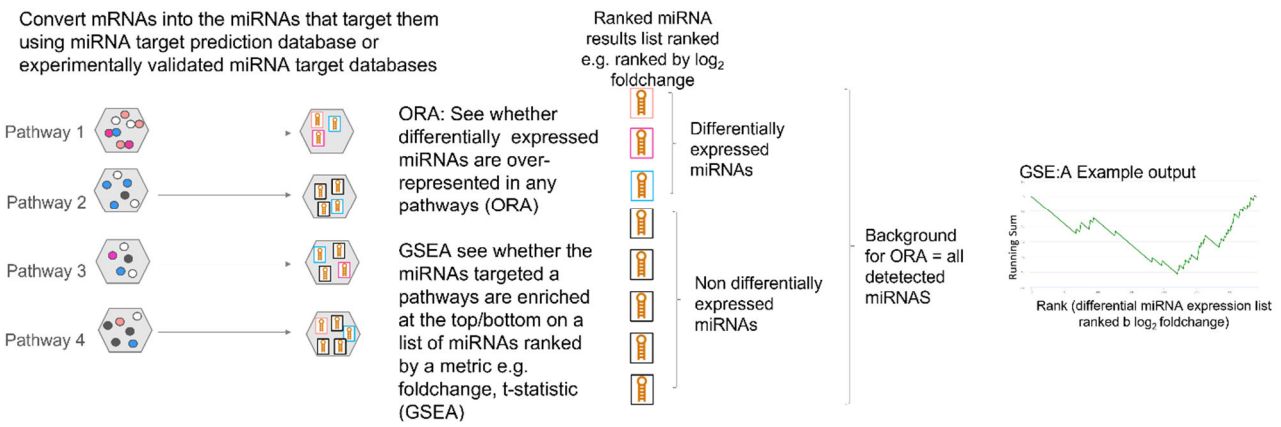
2.7.1 miRNA vs mRNAs

There are two approaches, as shown in Figure 8:

- **miRNA <u>set</u> enrichment analysis** (Figure 8A) is implemented by MIEAA 2.0 web server[337]
- **miRNA <u>target</u> enrichment analysis** (Figure 8B) is implemented in R using miRNA target interactions downloaded from miRnet, and the fsgea package[338,339].

2.7.2 miRNA pathway analysis

MiRNA pathway analysis was done using the MIEAA2 web server using a GSEA-type approach[337] (Figure 9 explains the method and interpretation). A different way of performing miRNA pathway enrichment analysis is to convert a miRNA(s) to the genes they target and perform a pathway over-representation analysis (ORA). This latter approach is common in the literature but has been criticised for generating less useful terms, with cell-cycle and cancer terms often over-represented regardless of the list of inputted miRNAs[340].



Positive enrichment: the pathway contains mRNAs targeted by a group of miRNAs (ORA) or miRNAs towards the top of a ranked list (GSEA) more than expected by chance.
Negative enrichment: the pathway contains mRNAs that are less targeted by a group of miRNAs or miRNAs towards the bottom of a ranked list than expected by chance.

Interpretation: If the miRNA list is ranked from most upregulated to most downregulated:
Positive* enrichment** suggests that the expression of mRNAs/proteins in that pathway would ***decrease due to more miRNA targeting.
Negative* enrichment** suggests that the expression of mRNAs/proteins in that pathway would ***increase due to less miRNA targeting.

Figure 9 MIEAA2 GSEA style pathway enrichment analysis. Circles = genes. Circle colour indicates, which miRNA(s) targets, which gene. Hexagons = pathways. A pathway is converted to the miRNAs that target its genes, and those miRNAs are compared against a group of miRNAs (e.g. differentially expressed miRNAs) or a list of miRNAs, e.g. ranked by a differential expression metric such as log_2 foldchange.

3 The influence of vaccination on sRNA expression in plasma in infants

3.1 Introduction

To understand whether vaccination influences extracellular microRNA expression in infants, a study was designed using samples from the EUCLIDS infant vaccine study (details below).

3.1.1 4CMenB: a model for understanding vaccine reactogenicity

The multicomponent meningococcal serogroup B vaccine (4CmenB [Bexsero®, GlaxoSmithKline Biologicals, Rixensart, Belgium]) is given at 2, 4 and 12 months of age to protect against serogroup B *Neisseria meningitidis* (MenB) infection—the commonest cause of meningococcal septicaemia in infants. The vaccine contains MenB surface proteins: Neisserial Heparin Binding Antigen (NHBA), factor H binding protein (fHbp), *Neisseria* adhesin A (NadA) and PorA from the outer membrane vesicle (OMV).

4CmenB is reactogenic. Post-vaccine fever rates double to 60% when 4CMenB is added into infant vaccine regimes, with 10% of infants developing fever above 39ºC[341]. Post-vaccination A&E attendances increased after 4CMenB was included in the infant immunisation schedule in 2015, as did invasive septic screens and parenteral antibiotics as clinicians struggled to differentiate vaccine-induced fever from serious bacterial infection[99,342]. The reasons underpinning 4CMenB reactogenicity are unclear. 4CMenB is an ideal model of a reactogenic vaccine for delineating pathways underlying vaccine-associated fever in infants.

3.1.2 The EUCLIDS study

EUCLIDS was a randomised, descriptive, open-label study exploring the effect of 4CMenB on whole-blood gene expression (EudraCT number 2014-000126-38). The study aimed to elucidate immunological mechanisms underlying increased vaccine reactogenicity when 4CMenB is included in the infant vaccine schedule.

Infants were randomized to receive their 2-month, 4 month and 12 month routine vaccines with or without 4CMenB (Figure 10, Figure 11, and Section 3.3.2). Whole blood, PBMCs and plasma were taken at baseline, 4 hours, 24 hours, three days and seven days after vaccination at 4 and 12 months of age. Though infants were enrolled at two months, blood samples were only taken around the 4 and 12 month timepoint because 4CMenB vaccination is poorly immunogenic at 2-months-old.

One hundred and eighty-one infants completed the study. Fever of 38°C or over occurred in 43.5% of participants and was more common in the 4CMenB group compared with the routine group (59.6% [95% CI: 48.6, 69.8] vs 27.3% [95% CI: 18.3, 37.8], respectively)[76]. Neutrophils (measured by hospital full blood counts) increased 4 hours post-vaccination, peaking at 24 hours. Maximal differential gene expression occurred 24 hours after vaccination[76]. Only one gene, *ENTPD7* (an ectonucleoside involved in purinergic signalling between cells), was differentially expressed between the vaccine groups.

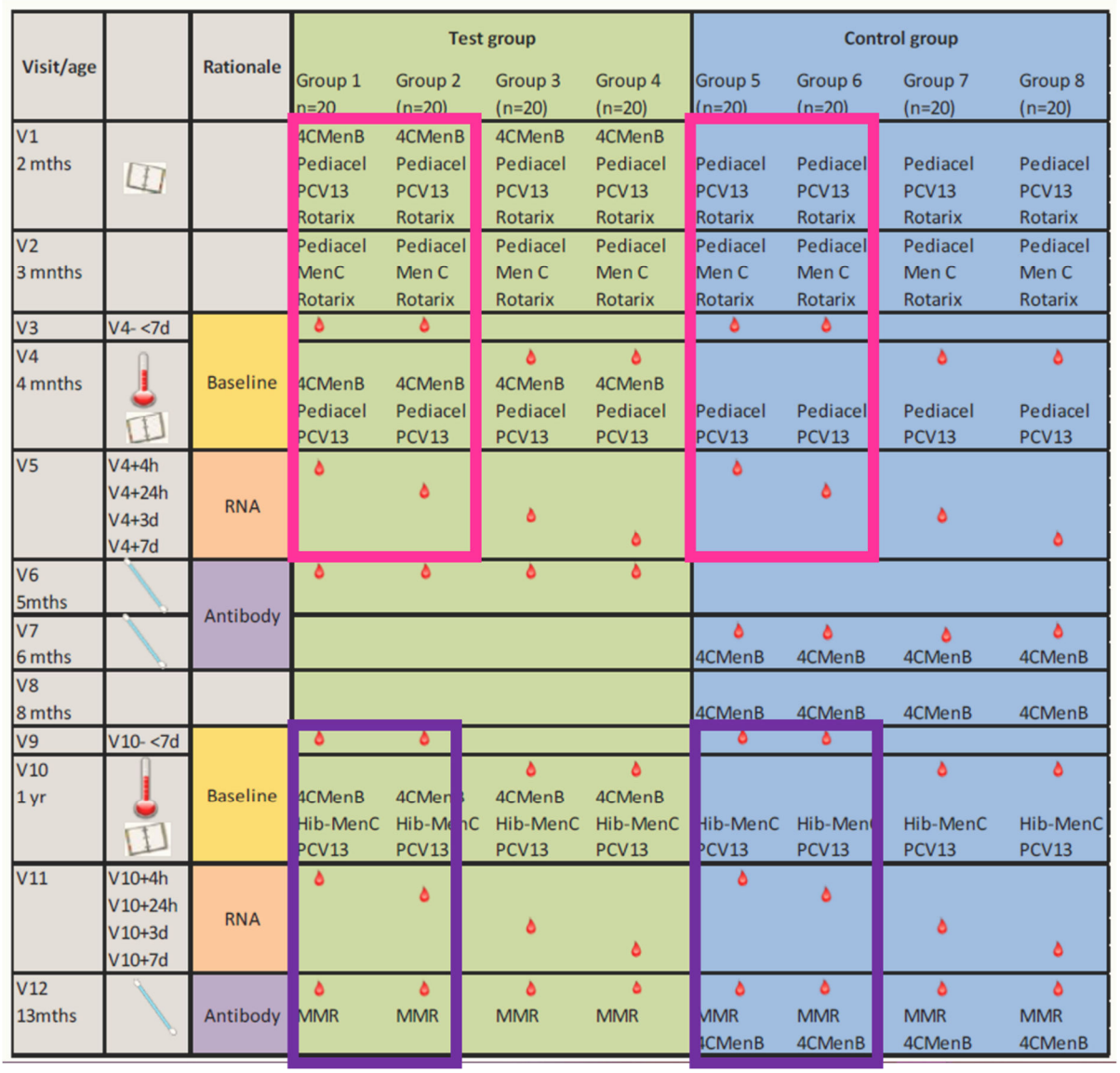


Figure 10 EUCLIDS study Overview (adapted from EUCLIDS study protocol)
At 2 months, 4-months and 12-months routine immunisations were given to the control vaccine group ("routine group"), routine immunisations plus 4CMenB were given to the test group ("4CmenB group"). Samples from Test groups 1 and 2, and Control groups 5 and 6 were used in this thesis to study changes in circulating sRNAs and proteins in the first 24 hours after vaccination. Vaccines: Pediacel (Sanofi Pasteur, DTaP-IPV-Hib, a combined 5 in 1 vaccine against diphtheria, tetanus, pertussis, poliomyelitis and Haemophilus influenzae type b). PCV13 (13-valent pneumococcus conjugate vaccine). Hib-MenC (Mentorix, GSK, a combined Haemophilus influenzae type b and Meningococcal group C conjugate vaccine). MMR (measles, mumps, rubella). Figure adapted from EUCLIDS study protocol to highlight samples used in this thesis. Pink boxes; pilot cohort, Purple boxes: Validation cohort

EUCLIDS offers a unique opportunity to explore the effect of vaccination on extracellular sRNA expression.

3.2 Chapter aims

Chapter aims:

1) Describe the influence of vaccination on plasma sRNAs
2) Evaluate whether concomitant administration of 4CMenB alongside routine immunisations influences plasma sRNAs
3) Establish whether extracellular miRNAs represent biomarkers of vaccine reactogenicity
4) Identify the source and function of differentially expressed (DE) miRNAs

3.3 Chapter Methods

3.3.1 EUCLIDS study population

Healthy infants of two Caucasian parents born at term were enrolled at 2 months of age. Exclusion criteria were: confirmed, suspected or a family history of immunodeficiency, a major congenital defect, serious chronic illness, prior receipt of blood-derived products, meningococcal serogroup B vaccination, or a previous household contact of confirmed bacterial meningitis.

3.3.2 EUCLIDS study intervention

Infants were randomized to receive their 2, 4 and 12 month routine vaccines with or without 4CMenB (Figure 10). The group that received concomitant

4CMenB are referred to as the "4CMenB" group. The group that received routine vaccines only are referred to as the "routine" group.

Routine vaccinations at 4 months of age were Pediacel and PCV13. Routine vaccinations at 12 months of age were Mentorix plus PCV-13 (Figure 10). Paracetamol was administered after vaccination at parental discretion.

3.3.3 Temperature measurement

Body temperature was measured continuously for 24 hours after vaccination using i-button technology (a probe that measures skin temperature every minute) alongside axillary temperatures at 4, 24 and 72 hours post-vaccination.

3.3.4 Measurement of sRNAs in infant plasma

3.3.4.1 Timepoints studied

sRNA expression in plasma was measured by sRNA-seq before and 24 hours after participants' 4 month and 12 month vaccinations (Figure 11), as this coincided with maximal differential gene expression in the whole blood RNA-seq dataset.

3.3.4.2 Sample selection.

Only participants with paired baseline and post-vaccination samples available at 4 months and/or 12 months were included (Figure 11 contains sample numbers).

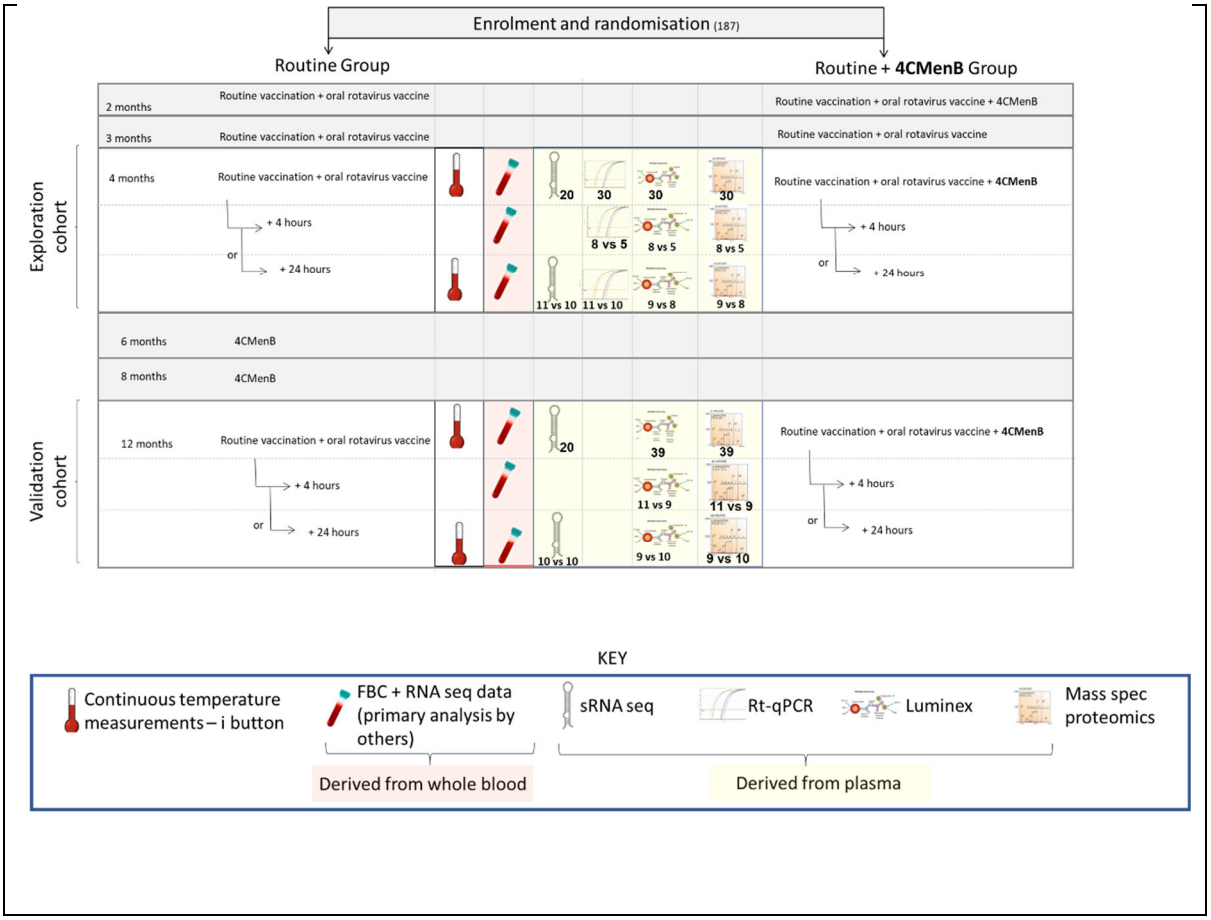


Figure 11 Overview of study samples and assays studied in this thesis. Note that 40 participants were enrolled on the main study—but not all had paired samples available—numbers underneath assays show the number of samples available for that assay (routine vs 4CMenB groups).

3.3.4.3 Plasma sample processing:

3.3.4.3.1 Plasma preparation

Plasma was isolated from blood collected in sodium heparin tubes and rendered platelet-poor by spinning at 10,000g for 10 minutes at 4°C.

3.3.4.3.2 Estimation of haemolysis

Visual inspection showed plasma samples were affected by haemolysis to different extents. Haemolysis alters the sRNA profile of plasma release[343–345]. Plasma sample absorbance at a wavelength of 414nm (haemoglobin absorbance wavelength) was measured using a nanodrop spectrophotometer as an objective

measure of haemolysis[343]. These readings were used as a covariate in the differential expression models to control for haemolysis.

3.3.4.3.3 Removal of heparin

Heparin inhibits PCR and potentially sequencing; therefore, 200μl of platelet-poor plasma was treated with one unit of heparinase I from flavobactum for two hours at room temperature before RNA extraction.

3.3.4.3.4 spike-ins

Artificial miRNA spike-ins were added to 12-month-old samples before RNA extraction at one attomole per 200μl of plasma. Spike-in amounts were based on a paper by Max et al.[346].

3.3.4.4 RNA extraction:

Total RNA was extracted from 200μl of processed plasma using Norgen's Total RNA extraction kit into a final elution volume of 50μl. RNA integrity was not measured as the relative ratio of 28S and 18S ribosomal RNA is not valid for ex-RNA. No assays were sensitive enough to measure these samples' RNA concentrations.

3.3.4.5 Sequencing of sRNA libraries

See methods section 2.2.

3.3.4.6 Statistical analysis

The 4-month-old cohort was treated as an exploratory cohort, and the 12-month-old cohort was used as a validation cohort. Although the 12-month-old time point is not a genuinely independent or replicating cohort (significant overlap between participants with samples sequenced at 4 months and 12 months, older age and different vaccines given), treating the 4 month and 12-month cohorts separately was a pragmatic way to identify robust results. To conserve resources, the 12-month cohort was run once the 4-month-old results showed promising findings.

In the 4-month-old cohort, features with an FDR value below 0.05 were deemed significant ("features of interest"). "Features of interest" were validated as differentially expressed at an unadjusted p-value cut-off below 0.05 in the 12-month-old analysis. This was done because the prior probability of differential expression of "features of interest" in the 12-month-old analysis is improved by significance in the 4-month-old data. Furthermore, fewer features are evaluated in the validation analysis reducing the multiple testing issue.

3.3.4.7 Correlates of reactogenicity

Neutrophil abundances from the EUCLIDS whole blood RNA-seq data were estimated via the CIBERSORTx pipeline (using the LM22 signature matrix)[347]. CIBERSORTx neutrophil fractions were highly correlated with full blood count neutrophil abundances (rho=0.81, $p<2.22$ e-16). Due to frequent missing full blood count data, CIBERSORTx neutrophil abundances were used in correlation analyses.

3.3.4.8 Quantitative reverse transcription PCR (RT-qPCR)

RT-qPCR was used to technically validate the sRNA-seq approach, study the kinetics of miRNA expression, and measure miRNA expression in mice plasma and liver. Probe-based Taqman assays and Qiagen LNA-based (appendix table 1) were used.

3.3.4.8.1 RT-qPCR using Locked Nucleic Acids based primer assays

3.3.4.8.1.1 RT-qPCR of RNA derived from plasma

Two microlitres of total RNA were universally reversed transcribed using the miRCURY LNA RT Kit (cat#339340). A synthetic oligonucleotide, UniSP6, was spiked into each reverse transcription reaction as a positive control. cDNA was subject to qPCR as per manufacturer's protocol.

3.3.4.8.1.2 RT-qPCR of RNA derived from tissues and cells

RT-qPCR of tissue and cell RNA was performed as above, except RNA was quantified using Qubit RNA Broad range detection assays (ThermoFisher Scientific cat#Q10210) using the Qubit 4 Fluorometer (ThermoFisher Scientific, cat#Q33238) with 10ng of RNA submitted to each reverse transcription reaction.

3.3.4.8.2 RT-qPCR using Taqman hydrolysis Probe-based primers

Two microlitres of total RNA were universally reversed transcribed, and pre-amplified using TaqMan™ Advanced miRNA cDNA Synthesis Kit (cat#A28007) as per manufacturer's instructions.

3.3.4.8.3 qPCR runs

qPCR was conducted using the StepOnePlus™ Real-Time PCR System (Thermofisher, cat#4376600) per manufacturer protocols. Assays were conducted in triplicate, and each sample's arithmetic mean cycle threshold values (Ct values) calculated. Negative controls without cDNA were included. Assays were taken forward if their amplification curves could be seen when normalised fluorescence was plotted on a linear scale, had Cts<35 and 5 cycles lower than Ct values in non-template control wells, and a single Tm peak occurred in melt curve analysis (in the case of LNA primers). Baseline Rn was manually set. Delta Rn threshold values were set at 0.1 or 0.01, based on the optimal threshold value for each primer.

3.3.4.8.4 RT-qPCR Data processing

Outlier Ct values were removed. miRNA Ct values were normalised to the Ct values of endogenous controls selected from the RNA sequencing data.

3.3.4.8.5 Selection of endogenous control for RT-qPCR.

MiRNAs present in 100% of samples were ranked in terms of average $\log_2$ fold-change and the standard deviation of their fold-change. These ranks were multiplied to give a composite rank. MiRNAs with the lowest composite rank were selected as controls. Three endogenous controls were used[348].

3.3.4.8.6 Statistical analysis

Relative miRNA expression between paired samples was expressed as $\log_2$ fold changes calculated using the -ΔΔCt method. Differential expression was tested using a paired Wilcoxon test.

3.3.5 Mouse vaccination study

3.3.5.1 Mice and vaccinations

Six-week-old C57BL/6 mice were immunised with one of four immunisations schedules, as shown in Figure 31. Immunisation was repeated three weeks later. Mice were sacrificed twenty-four hours after their second immunisations.

3.3.5.2 Mouse temperature measurement

Mouse temperature was measured 1.5, 3, 4.5, 6 and 24 hours after vaccination by infrared thermometry of the perianal region using a No Touch Thermometer NTF3000 (Braun).

3.3.5.3 Mouse samples and sample processing

Blood was stored in heparinised tubes after cardiac bleed. Heparinised tubes recapitulated blood sampling in the infant study. Samples were processed per the infant study, except RNA was extracted from 100μl of mouse plasma per manufacturer's advice. cel-miR-39 and synthetic UniSP4 (Qiagen, cat#339390) were spiked into plasma before RNA extraction as the best reference miRNAs in mice are unknown.

Whole livers were perfused with normal saline to remove excess blood and stored in RNAlater. RNA was extracted from 20mg of liver using Norgen Total RNA Purification Plus Kit (Cat#48300) designed to remove excess DNA. This was performed per manufacturer instructions, except liver was first homogenised in gentleMACS™ M Tubes (Miltenyi Biotec, cat#130-093-236) in 600μl RLT lysis buffer containing beta-mercaptoethanol.

3.3.5.4 RT-PCR of mouse plasma and liver

PCR was done using Qiagen LNA-based primers because it has a validated primer for a commonly used tissue sRNA normaliser—snRNA U6[76]. A bridging experiment on infant study plasma samples confirmed that LNA-based primers detect miR-122-5p downregulation (Appendix Figure 5).

RT-PCR was done using 7 potential housekeeping genes; the reference miRNAs used in the infant study, two spike-in miRNAs (cel-miR-39 and the synthetic UniSP4) and two commonly used housekeeping miRNAs for mouse serum; mmu-miR-30b-5p and mmu-miR103a-3p, and the best ones chosen empirically from the RT-PCR data.

3.4 Results

3.4.1 Sample demographics

Paired plasma was available at baseline and 24 hours after vaccination in 21 and 20 participants at 4 and 12 months of age, respectively. Table 11 and Appendix Table 2 contain a summary and individual participant demographics,

respectively. Thirteen participants had paired samples at both 4 and 12 months of age.

Table 11 Demographics of participants with samples submitted for sRNA-seq

	Vaccine Group		χ^2 p-value
	Routine	Routine + 4CMenB	
Number (%)			
four-months	11 (52%)	10 (48%)	*NA*
12months	10 (50%)	10 (50%)	*NA*
Female (%)			
four-months	7 (64%)	6 (60%)	0.86
12months	6 (60%)	5 (50 %)	0.96
Fever ≥ 38ºC (%)			
four-months	2 (18%)	5 (50%)	0.12
12months	4 (40%)	6 (60%)	0.37
Fever missing data			
four-months	0 (0%)	1 (10%)	*NA*
12months	1 (1%)	0 (0%)	*NA*
Maximal fever recorded within 24 hours of vaccination; median (IQR)			
four-months	37.7°C (37.3°C-37.8°C)	38.2°C (37.8°C-38.6°C)	*NA*
12-months	37.9°C (37.6°C - 38.2°C)	38.2°C (37.5°C - 39.1°C)	*NA*
Paracetamol administration (%)			
four-months	4 (36%)	4 (40%)	0.86
12-months	1 (10%)	4 (40%)	0.12

3.4.2 sRNA-seq of plasma prior to and 24 hours after vaccination in 4-month-old infants.

See Appendix 9.2.3 for library size statistics. There were 2,796 features tested in the final analyses (Table 12). Appendix 1 contains QC plots of the data. No outlier samples were identified.

Table 12 Number of features included in the EUCLIDS differential expression analysis

RNA class	Number of features
miRNA	237
Offset miRNAs (miRNA hairpins)	26
tRNA	21
snoRNA	17
snRNA	5
yRNA	4
Protein coding	2,111
lncRNA	705
vaultRNA	3
Reverse complement (antisense to any class of RNA)	468
Total features	**2,796**

3.4.2.1 miRNAs are a small proportion of sRNA sequences in plasma

Figure 12 shows the distribution of reads across feature types, broken down by vaccine status. The most abundant features were tRNA-dfs (median 40%, IQR:31%-48%) followed by rRNA fragments (median 21%, IQR:12%-32%). miRNAs comprised a small proportion of features (median 5%, IQR:3%-7%). yRNA fragments only made up 6% of reads (IQR 4.9-7.6), but given that only 4 yRNA genes exist, they were amongst the most abundant at a gene-wise level. Four per cent of reads map to the genome, but not to a feature; of these, approximately 50% (i.e., 2% of reads) lie opposite (i.e., are antisense) to a known feature—these are prefixed with "reverse complement" in the results. Supplementary Table 15 shows the relative abundance of RNA types across the studies in this thesis.

3.4.2.2 Haemolysis influences the composition of the sRNAome.

Haemolysis influenced plasma sRNA expression profiles at the global, RNA biotype, and gene-wise levels (Figure 13, Table 13, Appendix Table 3, Appendix Table 4); thus, haemolysis was controlled for in differential expression analyses. Vaccine status was not associated with haemolysis (p=0.4).

3.4.2.3 Vaccination changes the composition of the sRNAome and tRNAs and miRNA expression in particular

After vaccination, the relative abundance of tRNA and miRNA reads increased and decreased, respectively (Table 13, Figure 12A). This was also reflected at the feature-wise level (Figure 12B). Since reads compete for sequencing space, global upregulation/downregulation in one feature influences the abundance of other features. Differential expression analyses were therefore conducted on miRNA and non-miRNA reads separately. This "within RNA biotype analysis" approach prevents miRNAs from appearing downregulated secondary to tRNA upregulation and/or vice versa.

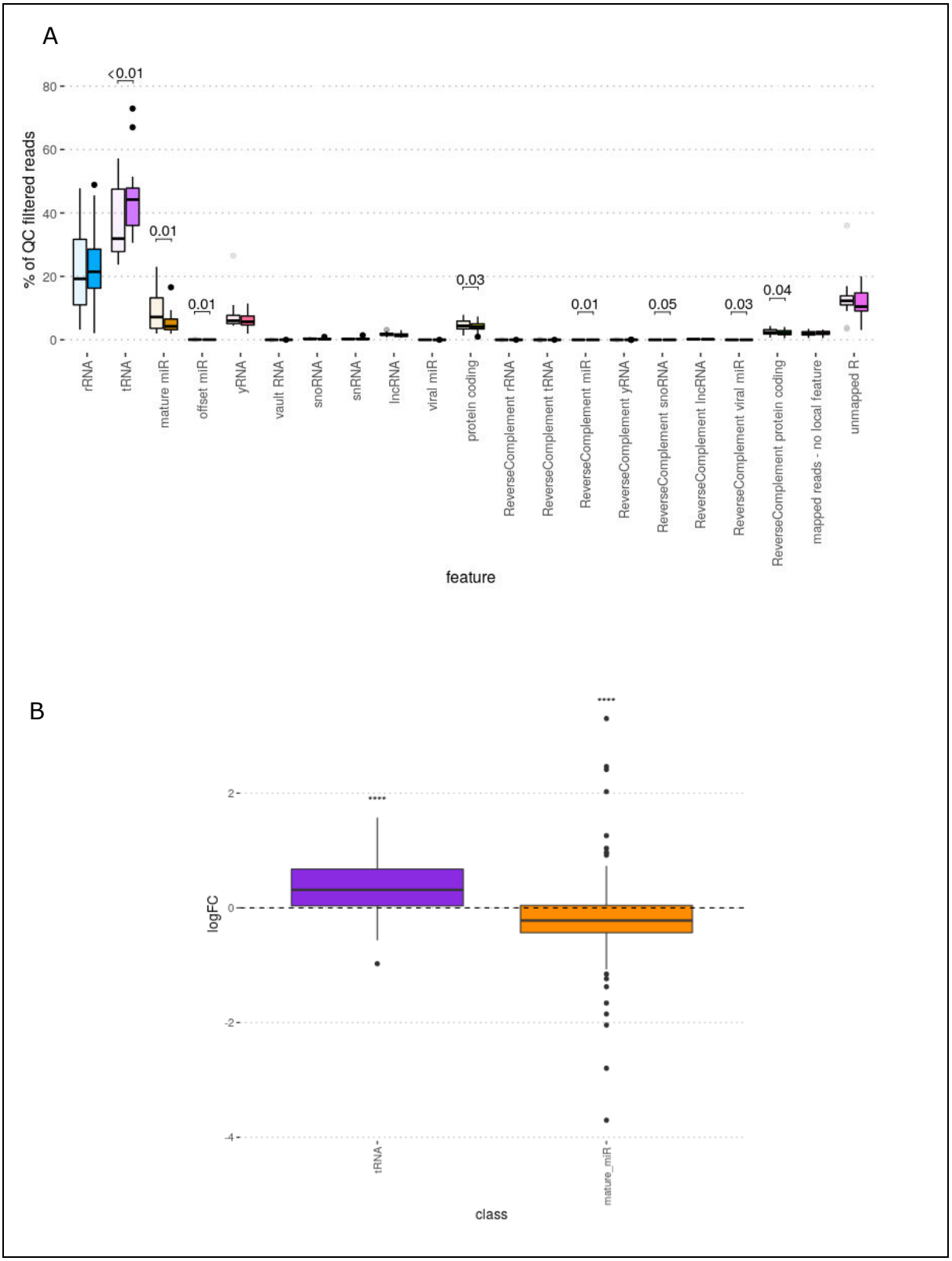


Figure 12 Global changes in the expression of sRNAs after vaccination in 4-month-old infants

A. Boxplot summarizing the sample-wise distributions in the percentage of reads mapping to each feature type. Light boxplots and circle points represent pre-vaccine samples, and dark boxplots represent post-vaccine samples. Significant

(Figure 12 caption continued)
p-values are presented derived from a mixed-effects model where the percentage of all reads mapping to each RNA biotype of RNA was modelled using vaccine status (as a fixed effect), haemolysis (as a fixed effect) and participant ID (as a random effect), as additive effects, i.e.,
% of all reads mapping to each RNA biotype of RNA~ vaccine status+ haemolysis + 1|participant ID

B. *Boxplot of the log_2 fold changes of tRNA derivatives and mature miRNAs. ****p<0.0001*

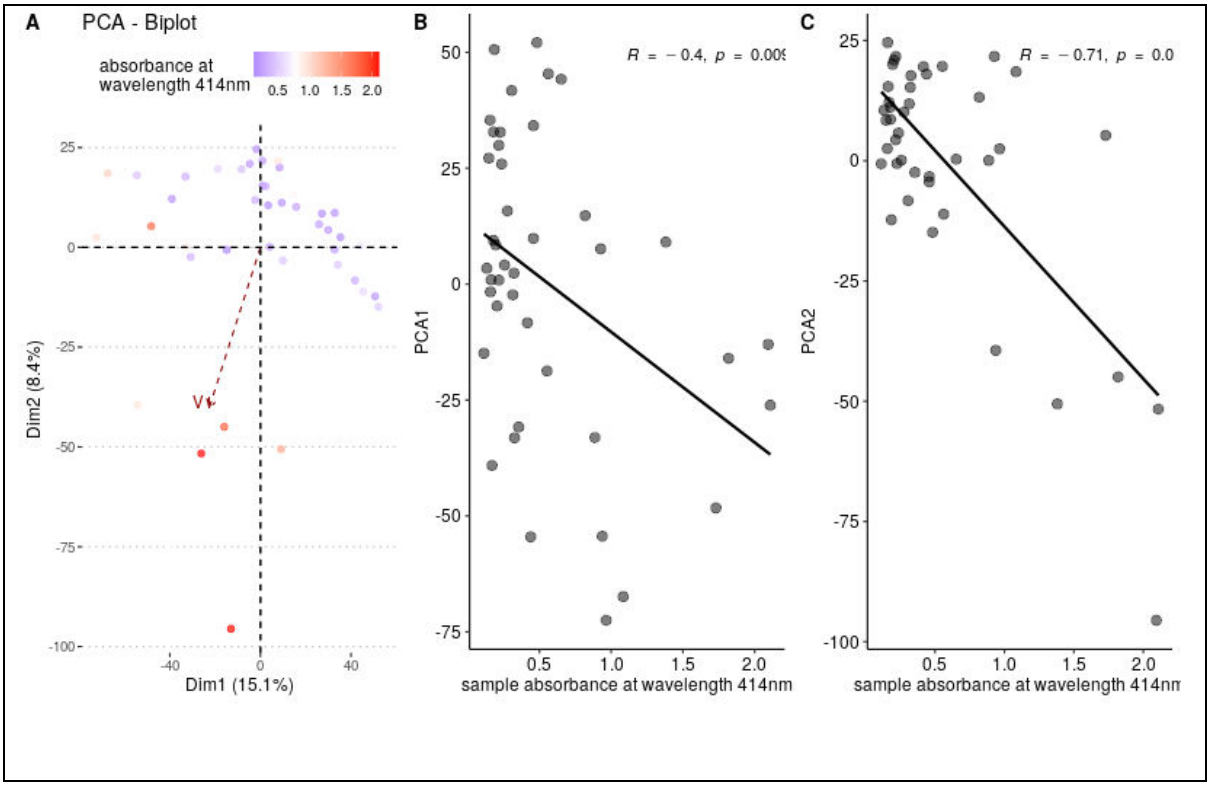


Figure 13 The influence of haemolysis on the global expression of sRNAs in 4-month-old infants

A. *PCA plot based on the expression (normalized log_2 cpms) of all features. Each dot represents a sample. Colour relates to sample absorbance at λ 414 nm (wavelength of haemoglobin). The more red a sample is on the plot, the greater the absorbance. The higher the absorbance, the higher the haemoglobin content of the samples and the greater the degree of haemolysis. More heamolysed samples plot towards the bottom left-hand of the plot, indicating that haemolysis influences both PCA1 and PCA2*

B. *Correlation between PC1 and heamolysis. (Pearson correlation presented).*

C. *Correlation between PC2 and heamolysis. (Pearson correlation presented).*

Table 13 association testing between % feature type abundance and vaccine (both groups) status and haemolysis. Bold figures highlight results with p<0.05. Red cells: significantly higher post-vaccination, blue cells significantly lower post-vaccination. Linear mixed-effect model = % mapping to feature ~ vaccine status + absorbance at λ 414nm + 1|participant number.

Feature	Paired Wilcoxon (p-value for association with vaccine status)	Linear mixed-effect model	
		(P-value for association with haemolysis)	(P-value for association with vaccine status)
lncRNA	0.539	**0.009**	0.147
mature miR	**0.005**	**0.000**	**0.006**
offset miR	**0.016**	0.183	**0.005**
rRNA	0.812	**0.036**	0.984
snoRNA	**0.035**	**0.011**	0.093
snRNA	0.374	0.125	0.179
tRNA	**0.032**	**0.000**	**0.002**
vault RNA	0.272	0.517	0.140
viral miR	0.393	0.558	0.162
yRNA	0.191	0.139	0.090
protein-coding	0.393	**0.000**	**0.030**
ReverseComplement lncRNA	0.562	**0.028**	0.197
ReverseComplement miR	**0.032**	0.560	**0.008**
ReverseComplement protein-	0.432	**0.000**	**0.035**
ReverseComplement rRNA	0.838	**0.005**	0.604
ReverseComplement snoRNA	0.203	**0.000**	**0.046**
ReverseComplement tRNA	0.433	0.816	0.421
ReverseComplement viral	0.185	**0.005**	**0.031**
ReverseComplement yRNA	0.584	0.357	0.482
unmapped R	0.393	**0.008**	0.090
mapped reads - no local	0.785	**0.000**	0.371

3.4.2.4 *MicroRNAs are differentially expressed post-vaccination, and this is more marked when 4CMenB is given*

PCA showed that vaccination globally influences miRNA expression (Figure 14). miRNAs were significantly differentially expressed when vaccine groups were analysed together (8/237) and separately (4CMenB group: 8/237, routine

group: 3/237) (Figure 14, Figure 15, Figure 16). Two miRNAs, hsa-miR-122-5p and hsa-miR-483-5p, were DE in both vaccine groups. Differential expression was more pronounced in the 4CMenB compared with the routine group, with a significant trend toward lower p-values in the 4CMenB group compared with the Routine group (Figure 14, Figure 16, Figure 17), indicating 4CMenB co-administration increases perturbation of circulating miRNAs. Two miRNAs (hsa-miR-145-3p and hsa-miR-338-5p) had significantly larger post-vaccination fold changes in the 4CMenB compared with the routine group (Figure 16).

3.4.2.5 MiRNAs are potential correlates of vaccine reactogenicity

Only DE sRNAs post-vaccination were tested for association with fever, cell counts and SBA titres. Post-vaccination expression of 5/8 DE miRNAs significantly associated with maximal post-vaccination temperature. Vaccine group potentially confounds associations between temperature and miRNAs because co-administration of 4CMenB is associated with higher fever. After controlling for vaccine group, 3 miRNAs were associated with fever (Figure 18). Post-vaccination expression of 5/8 and 7/8 DE miRNAs were associated with neutrophils when controlling and not controlling for vaccine group, respectively (Figure 19).

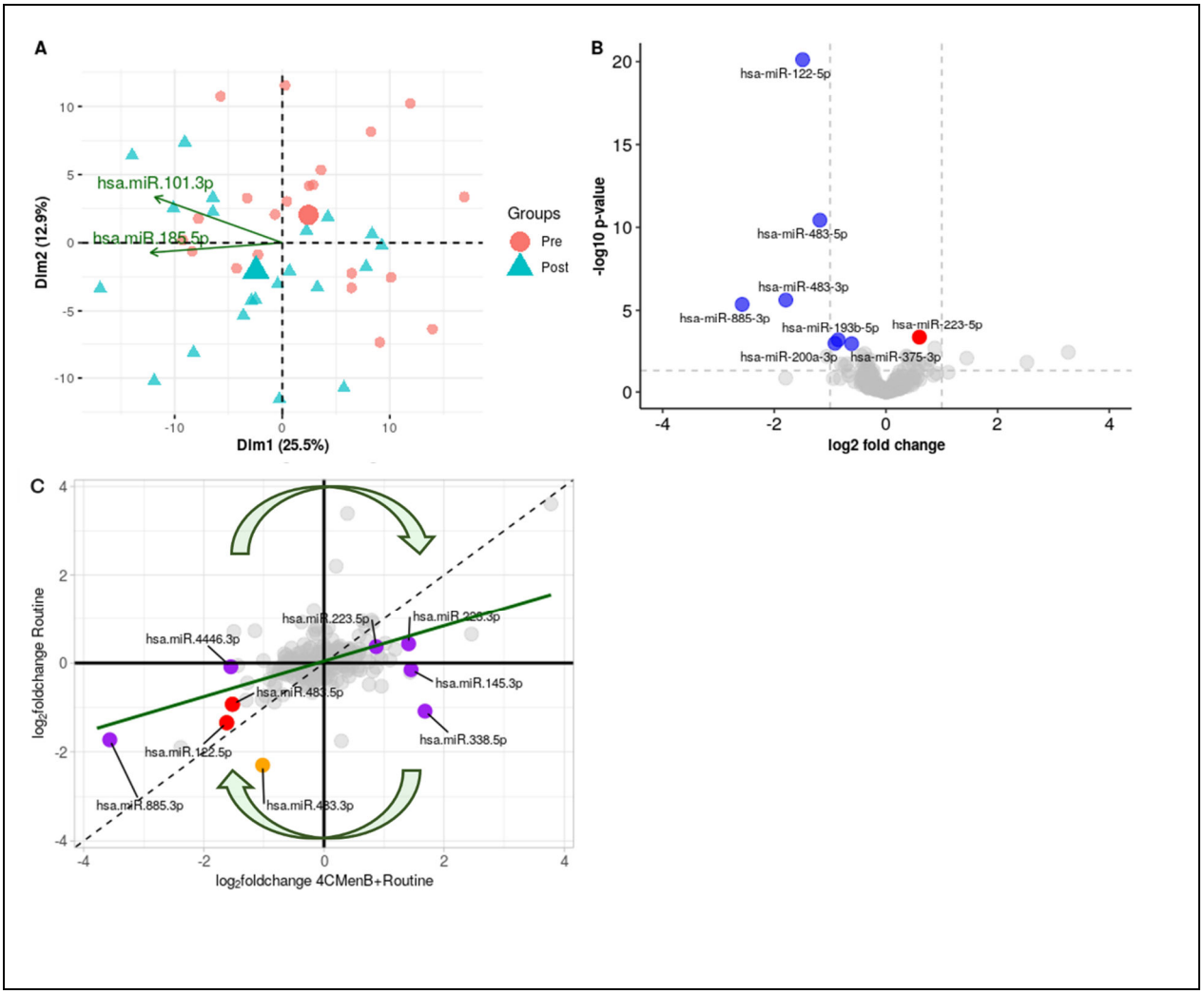


Figure 14 Influence of vaccination on miRNA expression

A. *Principle component analysis of miRNA expression data (normalized log counts per million, adjusted for haemolysis and participant ID). Each dot represents a sample. The centroids of the two groups are indicated with large symbols. The two miRNAs with the greatest contribution to the PCA are green.*
B. *Volcano plot showing results of differential expression analysis when both groups are analysed together. Red= upregulated at fdr<0.05, blue = downregulated at fdr<0.05*
C. *Cross plot showing agreement between fold changes in the MenB group (x-axis) and Routine Group (y-axis). Red dots highlight genes which were significantly DE in both vaccine groups. Purple dots highlight genes significantly DE in the MenB group only. Orange dots highlight genes significantly DE in the Routine group only. A line of best fit has been drawn in green.*

Table 14 DE miRNAs in plasma pre versus 24 hours post-vaccination

Analysis	miRNA	Log_2 Foldchange	Log CPM	P-value	FDR
Both Groups combined	hsa-miR-122-5p	-1.49	16.42	7.68E-21	<0.01
	hsa-miR-483-5p	-1.18	11.19	3.69E-11	<0.01
	hsa-miR-483-3p	-1.79	7.12	2.63E-06	<0.01
	hsa-miR-885-3p	-2.57	5.67	4.38E-06	<0.01
	hsa-miR-223-5p	0.60	9.95	4.77E-04	0.02
	hsa-miR-193b-5p	-0.86	7.44	6.96E-04	0.03
	hsa-miR-200a-3p	-0.91	6.26	1.13E-03	0.04
	hsa-miR-375-3p	-0.62	10.68	1.19E-03	0.04
Routine + 4CMenB group	hsa-miR-122-5p	-1.62	16.42	2.76E-10	<0.01
	hsa-miR-483-5p	-1.52	11.19	6.42E-08	<0.01
	hsa-miR-885-3p	-3.56	5.67	3.44E-05	<0.01
	hsa-miR-145-3p	1.45	6.36	1.43E-04	0.01
	hsa-miR-223-3p	1.41	6.98	3.13E-04	0.01
	hsa-miR-223-5p	0.87	9.95	3.48E-04	0.01
	hsa-miR-4446-3p	-1.55	6.94	5.81E-04	0.02
	hsa-miR-338-5p	1.68	6.04	1.38E-03	0.04
Routine group	hsa-miR-122-5p	-1.34	16.42	2.09E-08	<0.01
	hsa-miR-483-3p	-2.33	7.08	2.05E-05	<0.01
	hsa-miR-483-5p	-0.93	11.19	1.85E-04	0.01
Difference in the log_2 fold changes between the Routine Group and the Routine+ 4CMenB group	hsa-miR-338-5p	2.77	6.04	1.37has4	0.01
	hsa-miR-145-3p	1.60	6.36	1.76E-03	0.01

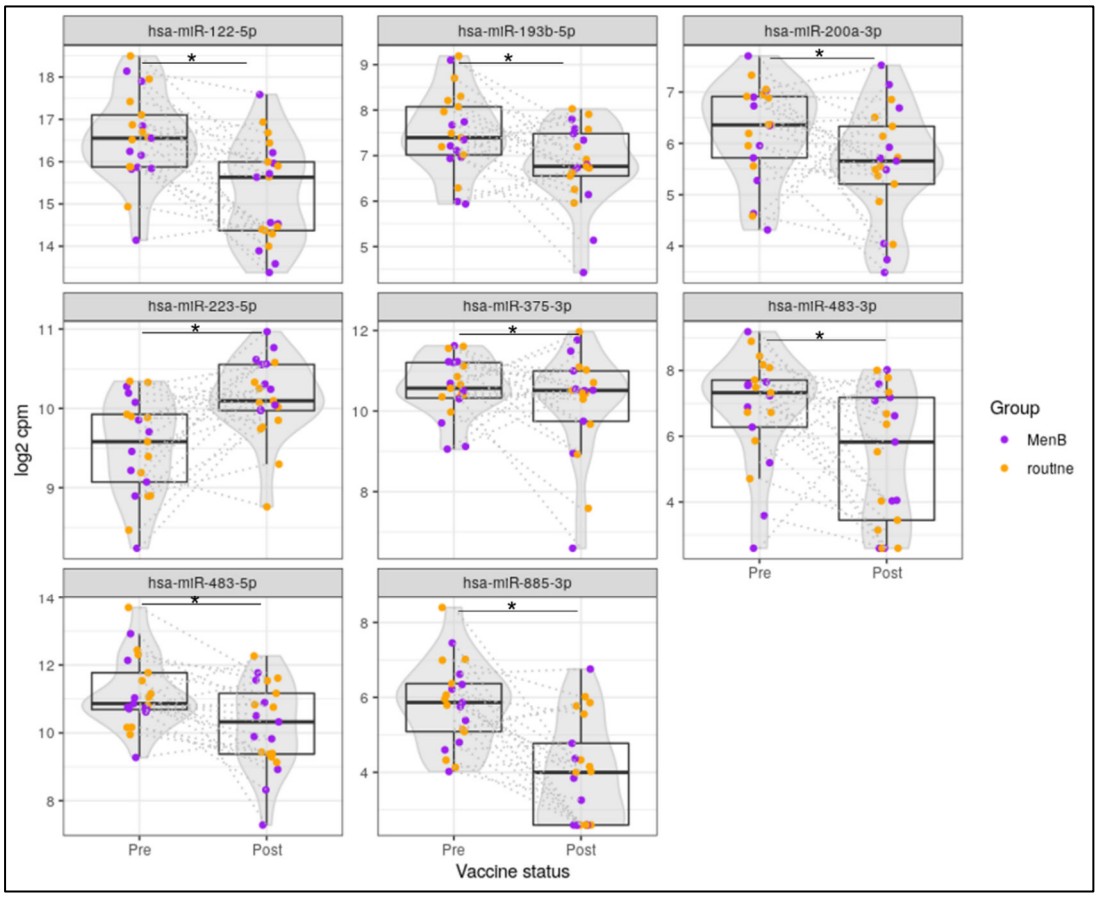


Figure 15 Dot plot showing the expression of miRNAs DE 24 hours after vaccination compared with baseline when both vaccine groups were combined.

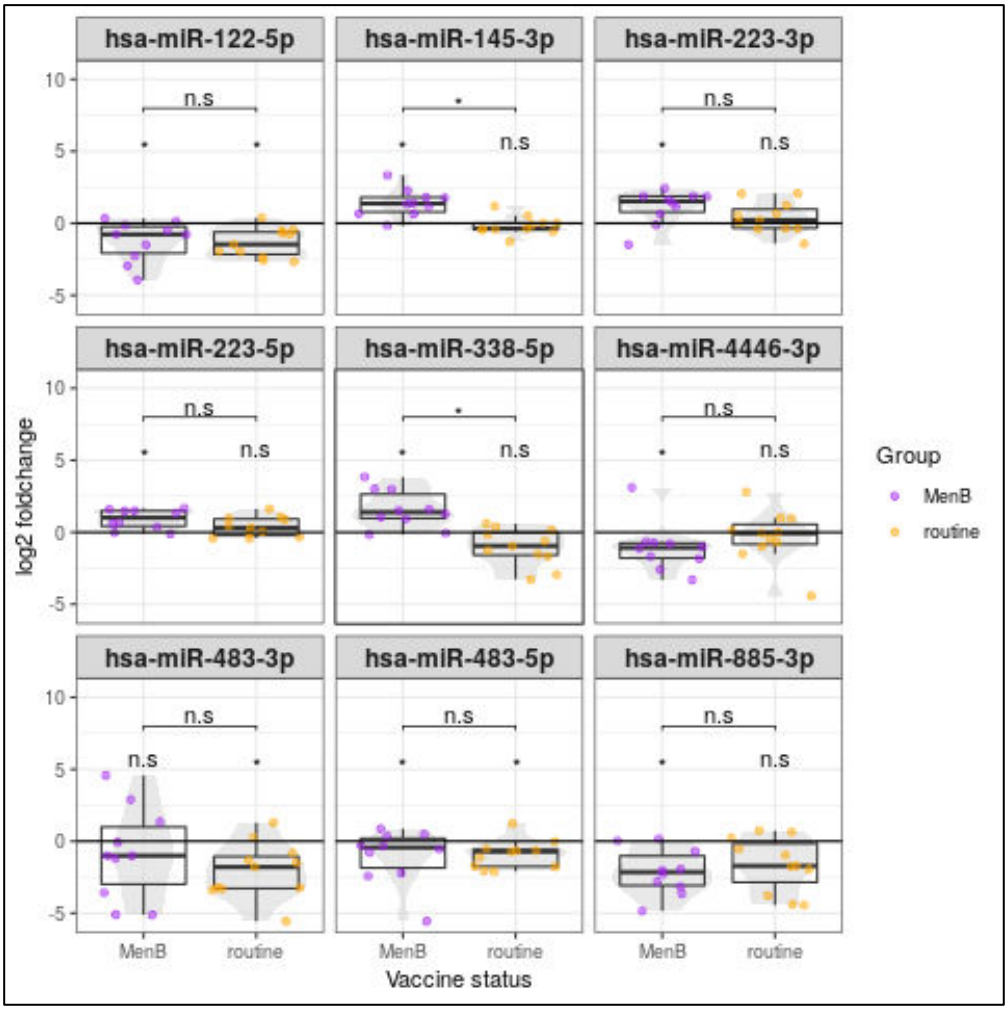


*Figure 16 Log$_2$ fold changes for miRNAs showing significant differential expression 24 hours after vaccination in the 4CMenB or Routine group. * fdr<0.05. hsa-miR-145-3p and hsa-miR-338-5p also showed significantly different fold changes between the two vaccine groups.*

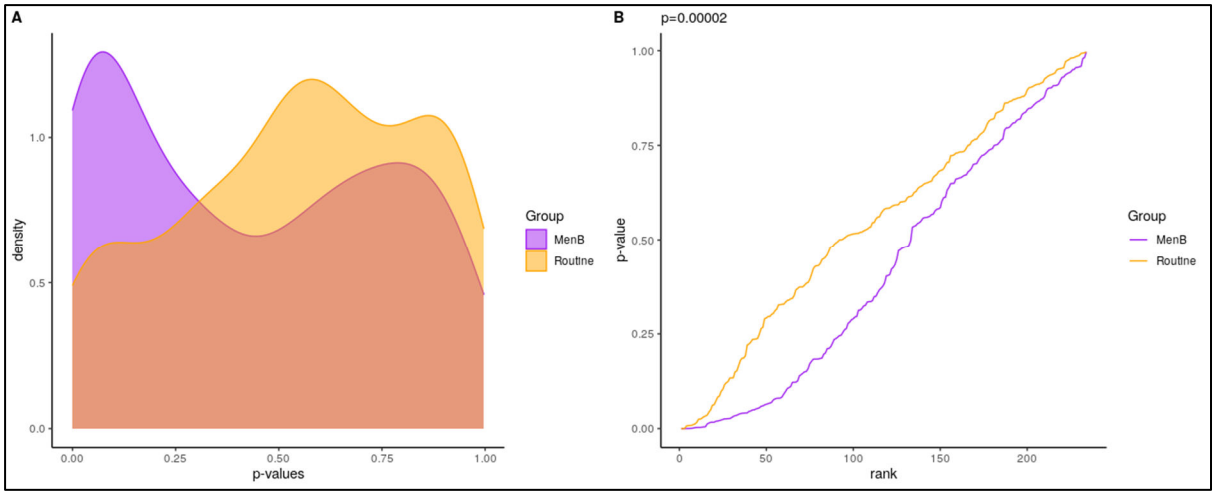


Figure 17 Distribution of p-values from pre- versus post-vaccine differential expression analysis shown in two ways.

A. The cumulative density plot shows the distribution of p-values for the differential expression analysis of the two vaccine groups. There is a trend towards lower p-values in the MenB group compared with the Routine group.

B. The graph shows the number of microRNAs passing different significance thresholds. P-value based on a KS test between the two distributions of p-values.

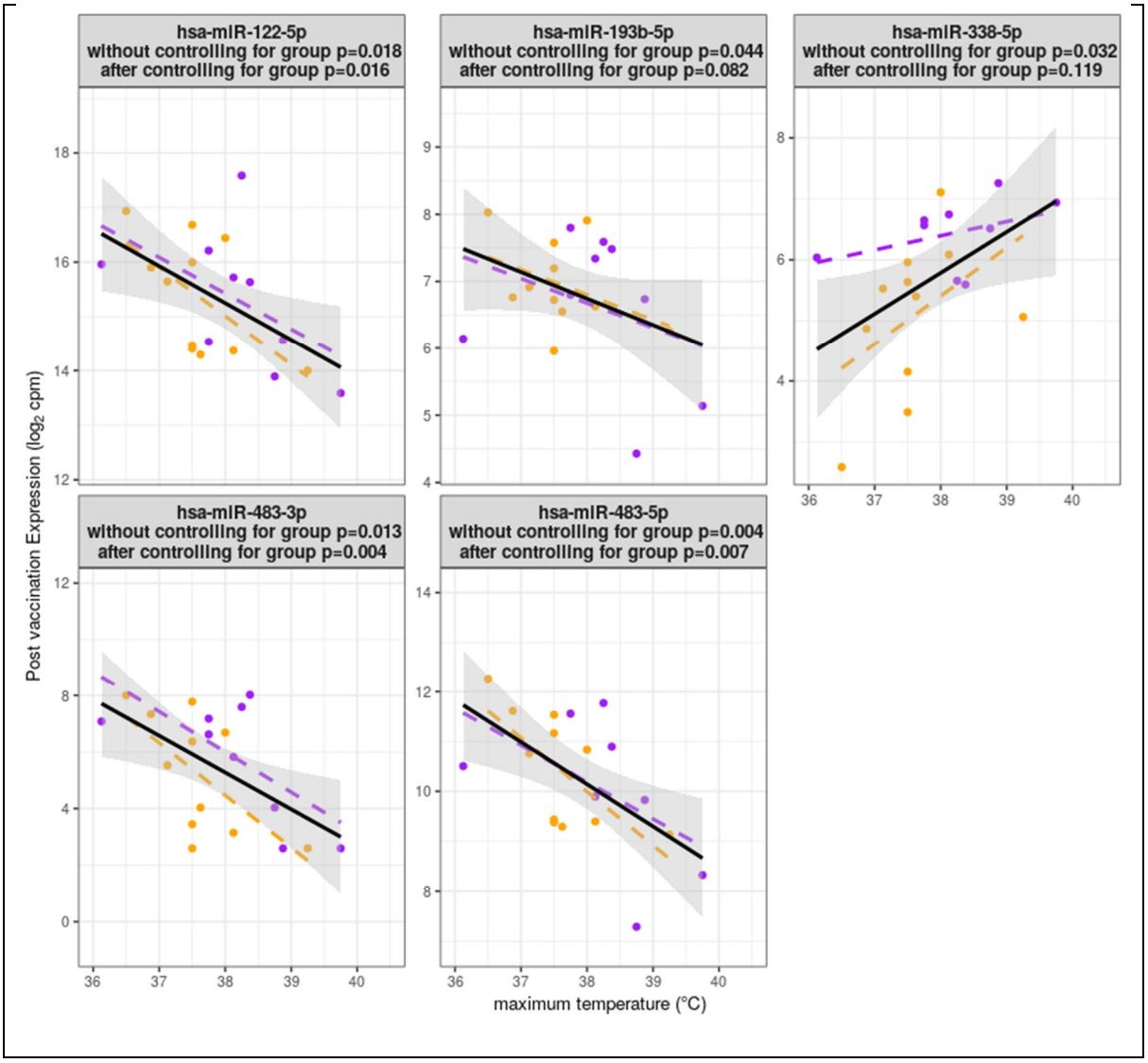


Figure 18 Correlations between post-vaccination miRNA expression and maximal post-vaccination temperature.

MiRNAs that were DE after vaccination were correlated with fever. The post-vaccination expression of 5 miRNAs was associated with maximal temperature within the first 24 hours after vaccination. Association testing was done using a linear model to control for haemolysis. Two models were run that did and did not control for vaccine group. The p-value results of both models are shown. Regression lines: Black line with confidence intervals = All participants, Orange dashed line = Routine group, Purple dashed line= 4CMenB Group.

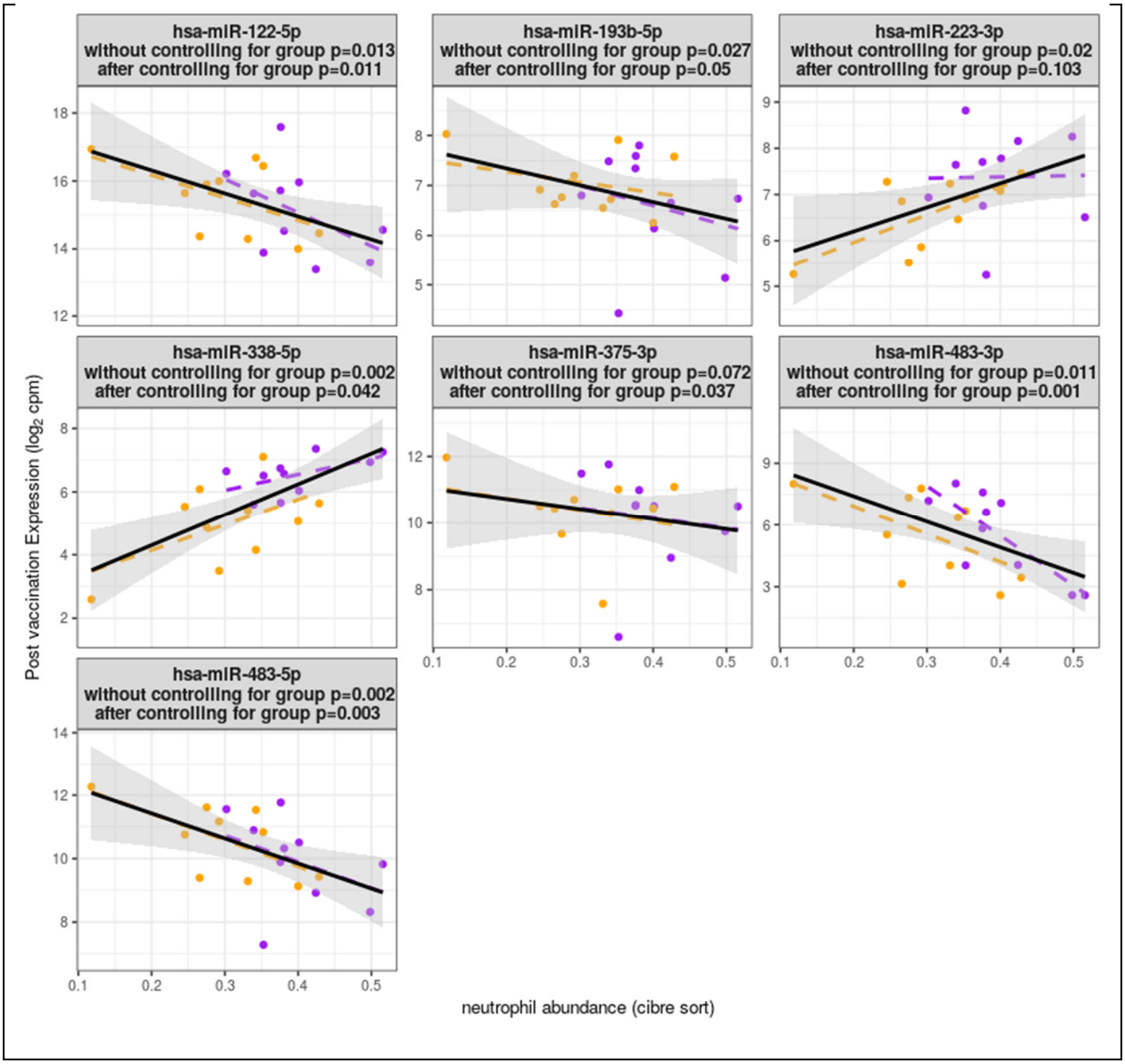


Figure 19 Correlations between post-vaccination miRNA expression and neutrophil abundance 24 hours after vaccination.
MiRNAs DE after vaccination were correlated with post-vaccination neutrophil abundance (as determined by CIBERSORTx analysis). Association testing was done using a linear model to control for haemolysis. Two models were run that did and did not control for vaccine group. The p-value results of both models are shown. Regression lines: Black line with confidence intervals = All participants, Orange dashed line = Routine group, Purple dashed line= 4CMenB Group.

3.4.2.6 Derivatives of tRNAs, snoRNAs and protein-coding genes are differentially expressed.

PCA showed good separation between pre- and post-vaccination samples. Post-vaccine samples from each vaccine group cluster separately (Figure 20).

Vaccination, altered expression of 26 sRNA-dfs (Figure 20 C, Table 15), including tRNA-dfs, protein-coding dfs and miRNA-hairpins ("offset miRNAs") and reads that mapped anti-sense to known features. Similar to the miRNA analysis, DE was more pronounced in the 4CMenB compared with the routine group (21 vs 6 sRNA-dfs, respectively, Figure 20 D&E, Table 15). Twenty-one sRNA-dfs had significantly different post-vaccination fold changes in the 4CMenB compared with the routine group, of which 13 mapped sense or antisense to mitochondrial genes (Table 15, Figure 16).

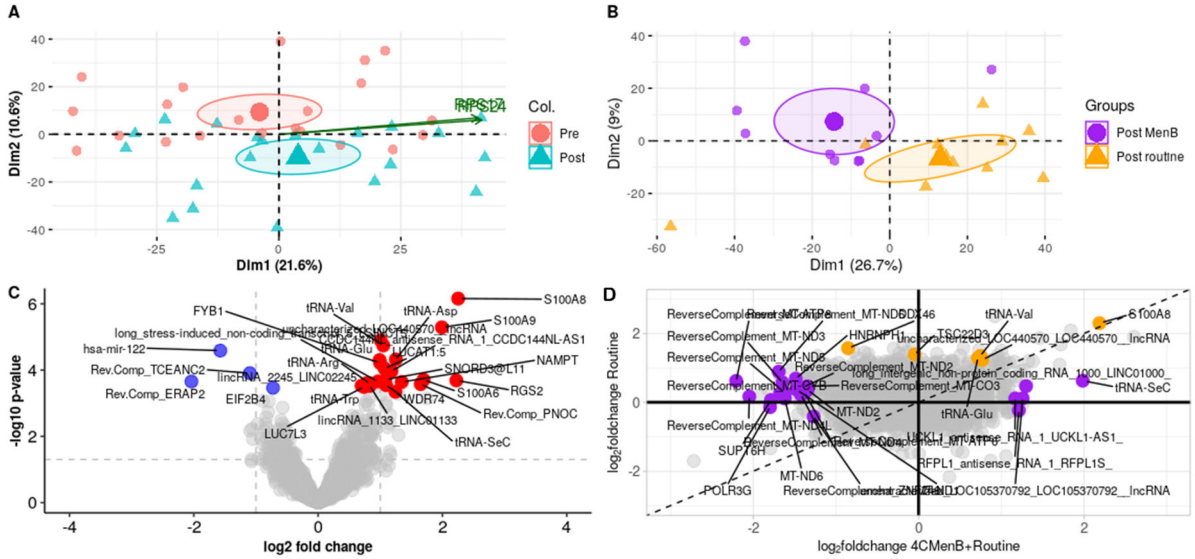


Figure 20 Influence of vaccination on sRNA-dfs *in infant plasma*

A. *PCA of sRNA-df expression data (normalized log counts per million, adjusted for haemolysis and participant ID). Each dot represents a sample. The centroids of the two groups are indicated with large symbols, and the ellipses represent 95% confidence intervals of the centroids of each cluster. The two sRNAs with the greatest contribution to the PCA are plotted in green.*
B. *PCA of sRNA-dfs expression data (normalized log counts per million, adjusted for haemolysis) of post-vaccination samples. The centroids of the two groups are indicated with large symbols, and the ellipses represent 95% confidence intervals of the centroids of each cluster. The two sRNAs with the greatest contribution to the PCA are plotted in green.*
C. *Volcano plots showing results of differential expression analysis when both groups are analysed together. Red= upregulated at fdr<0.05, blue = downregulated at fdr<0.05*
D. *Cross plot showing agreement between fold changes in the MenB + Routine group (x-axis) and Routine Group (y-axis). Purple dots highlight genes significantly DE in the MenB+Routine group only. Orange dots highlight genes significantly DE in the Routine group only. A line of best fit has been drawn in green.*

Table 15 DE sRNA-dfs, pre vs post-vaccination

Analysis	Feature name	Feature RNA biotype	logFC	logCPM	PValue	FDR
Both Groups combined	S100A8	Protein-coding	2.25	4.89	7.00E-	<0.01
	S100A9	Protein-coding	1.99	5.47	5.10E-	<0.01
	uncharacterized_LOC440570	lncRNA	1.01	5.99	1.27E-	<0.01
	tRNA-Val	tRNA	1.05	14.99	1.81E-	0.014
	hsa-mir-122	miRNA hairpin	-1.57	2.70	2.58E-	0.014
	tRNA-Asp	tRNA	1.25	7.91	4.76E-	0.014
	tRNA-Glu	tRNA	0.99	18.72	4.99E-	0.014
	FYB1	Protein-coding	1.00	3.71	6.14E-	0.025
	LUCAT1:5	lncRNA	1.07	3.58	8.94E-	0.037
	CCDC144NL_antisense_RNA_	lncRNA	1.18	7.99	1.09E-	0.044
	ReverseComplement_TCEANC	Reverse complement to a protein-coding gene	-1.10	3.18	1.24E-	0.044
	SNORD3@L11	snoRNA	1.08	4.50	1.94E-	0.050
	S100A6	Protein-coding	1.70	2.99	1.96E-	0.050
	RGS2	Protein-coding	2.22	2.68	2.05E-	0.03
	long_stress-induced_non-	lincRNA	0.96	3.07	2.20E-	0.03
	ReverseComplement_ERAP2	Reverse complement to a protein-coding gene	-2.04	2.45	2.21E-	0.03
	NAMPT	Protein-coding	1.35	2.77	2.29E-	0.03
	LINC02245	lincRNA	0.77	4.76	2.51E-	0.03
	tRNA-Arg	tRNA	0.83	6.48	2.52E-	0.03
	ReverseComplement_PNOC	Reverse complement to a protein-coding gene	1.65	3.77	2.64E-	0.03
	WDR74	Protein-coding	1.09	10.30	2.84E-	0.03
	tRNA-Trp	tRNA	0.70	3.95	2.95E-	0.03
	LINC01133	lincRNA	0.83	7.42	3.00E-	0.03
	LUC7L3	Protein-coding	0.74	3.90	3.22E-	0.03
	EIF2B4	Protein-coding	-0.72	4.25	3.43E-	0.04
	tRNA-SeC	tRNA	1.24	3.44	4.52E-	0.04

Table 13 continued

4CMenB	**ReverseComplement_MT-**	**Reverse complement to a mitochondrial protein-**	**-1.65**	**21.56**	**3.40E-**	**0.01**
	ReverseComplement_MT-CYB	Reverse complement to a mitochondrial protein-	-1.67	21.03	4.50E-	0.01
group	ReverseComplement_MT-	Reverse complement to a mitochondrial protein-	-2.21	19.55	9.80E-	0.01
	MT-ND6	Mitochondrial protein-coding	-1.62	19.41	1.05E-	0.01
	ReverseComplement_MT-	Reverse complement to a mitochondrial protein-	-2.05	18.99	1.31E-	0.01
	ReverseComplement_MT-ND6	Reverse complement to a mitochondrial protein-	-1.69	18.05	2.16E-	0.01
	ReverseComplement_MT-ND5	Reverse complement to a mitochondrial protein-	-1.62	17.28	3.23E-	0.01
	tRNA-SeC	tRNA	1.99	16.64	4.51E-	0.01
	POLR3G	Protein-coding	-1.80	15.96	6.47E-	0.02
	ReverseComplement_MT-ND4	Reverse complement to a mitochondrial protein-	-1.42	15.75	7.22E-	0.02
	ReverseComplement_MT-	Reverse complement to a mitochondrial protein-	-1.38	15.20	9.70E-	0.02
	MT-ND1	Mitochondrial protein-coding	-1.47	14.89	1.14E-	0.02
	SUPT6H	Protein-coding	-1.78	14.79	1.20E-	0.02
	ReverseComplement_MT-ND2	Reverse complement to a mitochondrial protein-	-1.41	14.73	1.24E-	0.02
	UCKL1_antisense_RNA_1	lncRNA	1.17	14.57	1.35E-	0.02
	ReverseComplement_ZNF746		-1.27	14.48	1.42E-	0.02
	Uncharacterized_LOC105370	lncRNA	1.26	13.99	1.84E-	0.03
	ReverseComplement_MT-CO3	Reverse complement to a mitochondrial protein-	-1.34	13.67	2.17E-	0.03
	RFPL1_antisense_RNA_1	lncRNA	1.21	13.54	2.34E-	0.03
	LINC01000	lincRNA	1.30	13.47	2.42E-	0.03
	HNRNPH1	Protein-coding	-1.48	12.49	4.09E-	0.05
	MT-ND2	Mitochondrial protein-coding	-1.34	12.47	4.14E-	0.05
Routine	tRNA-Val	tRNA	1.35	15.00	3.51E-	0.04
	DDX46	Protein-coding	1.57	2.66	5.80E-	0.04
Group	Uncharacterized_LOC440570	lncRNA	1.23	5.99	7.35E-	0.04
	TSC22D3	Protein-coding	1.39	2.99	8.92E-	0.04
	tRNA-Glu	tRNA	1.29	18.73	1.02E-	0.04
	S100A8	Protein-coding	2.30	4.87	1.04E-	0.04

Table 13 continued

Difference in the log_2 fold changes between the Routine Group and the Routine+ 4CMenB group	**ReverseComplement_MT-**	**Reverse complement to a mitochondrial protein-**	**-2.58**	**7.00**	**1.10E-**	**<0.01**
	ReverseComplement_MT-ND3	Reverse complement to a mitochondrial protein-	-2.22	7.36	3.30E-	<0.01
	ReverseComplement_MT-	Reverse complement to a mitochondrial protein-	-2.83	5.93	1.66E-	<0.01
	ReverseComplement_MT-CYB	Reverse complement to a mitochondrial protein-	-1.94	10.87	6.20E-	<0.01
	HNRNPH1	Protein-coding	-2.17	3.35	7.26E-	<0.01
	DDX46	Protein-coding	-2.42	2.66	8.05E-	<0.01
	ReverseComplement_MT-ND2	Reverse complement to a mitochondrial protein-	-1.83	7.46	2.19E-	<0.01
	ReverseComplement_MT-ND5	Reverse complement to a mitochondrial protein-	-1.89	10.82	2.44E-	<0.01
	ReverseComplement_MT-	Reverse complement to a mitochondrial protein-	-2.21	4.93	2.56E-	<0.01
	ReverseComplement_MT-CO3	Reverse complement to a mitochondrial protein-	-1.81	8.12	2.57E-	<0.01
	MT-ND1	Mitochondrial protein-coding	-1.84	10.18	3.20E-	<0.01
	ReverseComplement_MT-	Reverse complement to a mitochondrial protein-	-1.65	8.58	4.51E-	<0.01
	MT-ND6	Mitochondrial protein-coding	-1.73	9.41	5.13E-	<0.01
	ReverseComplement_MT-ND4	Reverse complement to a mitochondrial protein-	-1.64	9.93	6.32E-	<0.01
	MT-ND2	Mitochondrial protein-coding	-1.70	9.65	8.55E-	<0.01
	RFPL1_antisense_RNA_1	lncRNA	1.44	3.54	1.04E-	<0.01
	SUPT6H	lncRNA	-1.85	2.84	1.53E-	<0.01
	POLR3G	Protein-coding	-1.66	3.36	5.10E-	0.01
	UCKL1_antisense_RNA_1	lncRNA	1.06	4.50	1.01E-	0.03
	uncharacterized_LOC1053707	lncRNA	1.14	2.98	1.12E-	0.03
	TSC22D3	Protein-coding	-1.44	2.99	1.17E-	0.03

3.4.2.7 SRNA-df expression correlates with vaccine reactogenicity and immunogenicity.

Post-vaccination expression of 22 and 21 sRNA-dfs was significantly associated with fever and neutrophils, respectively. No sRNA-dfs DE after vaccination was associated with SBA titres. If the post-vaccination expression of all sRNA-dfs is interrogated for association with SBA titres (not just sRNA-dfs DE after vaccination), two sRNA-dfs, TWIST-2 and a feature reverse complement to HK2, were associated with SBA titres.

3.4.3 RT-qPCR data validation of miRNA results from the sRNA-seq data

3.4.3.1.1 Selection of miRNAs for RT-PCR validation

Due to cost and time constraints, only the most significantly upregulated and downregulated miRNAs were chosen as a representative sample of DE miRNAs. miRNAs with wet-lab validated primers available for purchase using LNA or Taq-man primers were prioritised. miRNAs selected for RT-qPCR normalisation were picked empirically from the sRNA-seq data—see results appendix section 9.2.5.

3.4.3.2 RT-PCR validates differential expression of miRNAs post-vaccination at 24 hours but not at 4 hours.

RT-PCR was run on paired samples at baseline and 4 and 24 hours post-vaccination. The most stably expressed candidate housekeeping genes determined by Bestkeeper were miR128-3p, hsa-miR-320b, and hsa-miR-378a.

All miRNAs selected for technical validation by RT-qPCR were confirmed DE 24 hours post-vaccination (Figure 21), as was the significant relationship between 24-hour post-vaccination expression of hsa-miR-122-5p and hsa-miR-438-5p and maximal post-vaccination fever and neutrophil abundance (Figure 22). No miRNAs were DE 4 hours post-vaccination (Figure 21A).

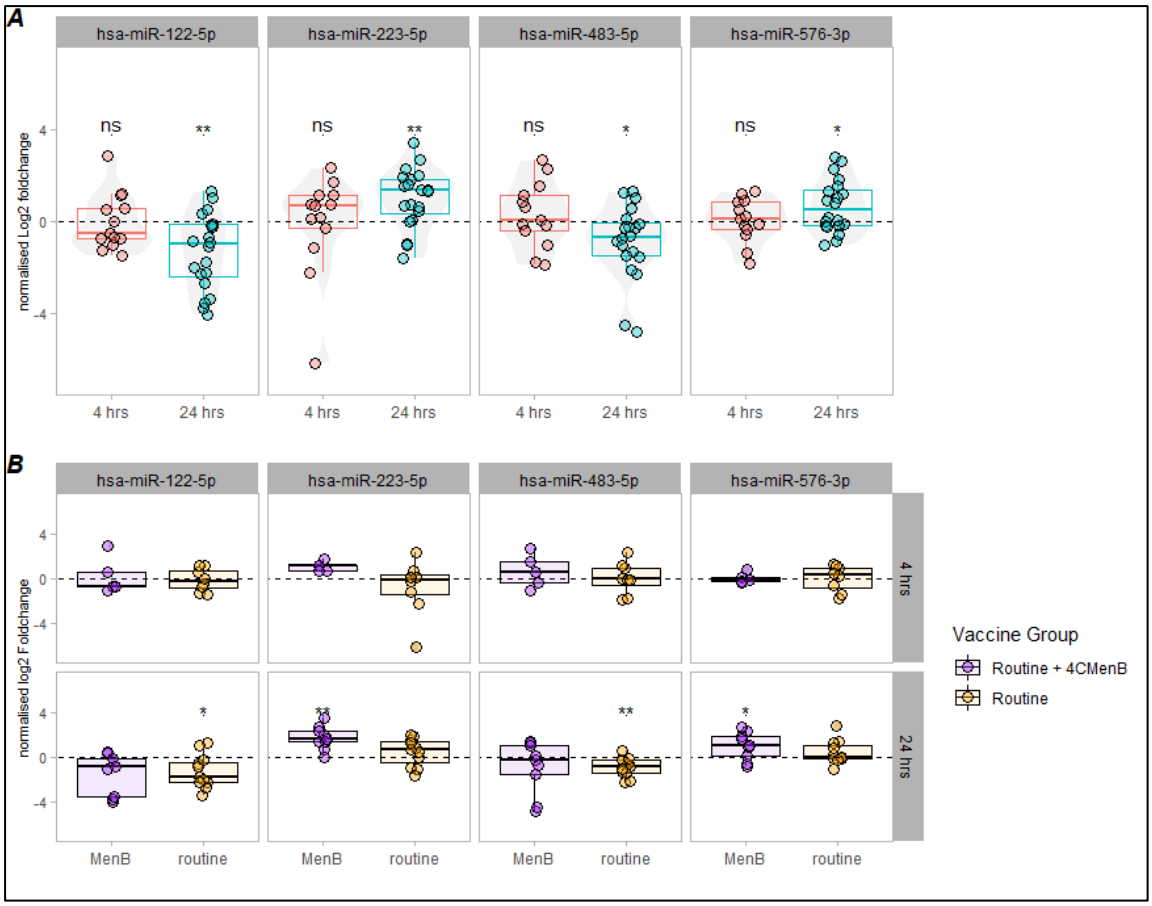


Figure 21 RT-qPCR of miRNAs in infant plasma samples before and after vaccination
A. Paired log_2 fold changes before versus 4 hours (blue) and 24 hours (red) post-vaccination.
B. Paired log_2 fold changes pre versus 4 hours (upper panel) and 24 hours (lower panel) post-vaccination, split by vaccine group.
** p<0.05, **p<0.01, n.s. not significant.*

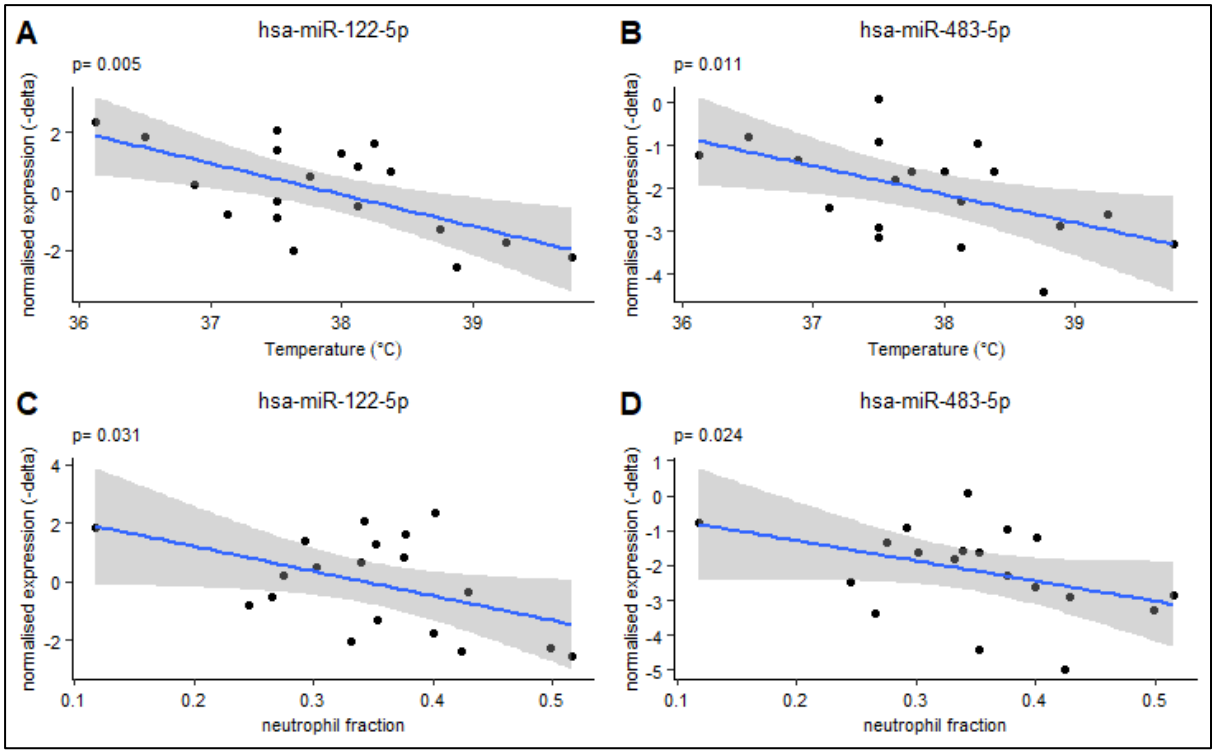


Figure 22 Correlation between post-vaccination miRNA expression and markers of reactogenicity

A. *Correlation between post-vaccination expression of hsa-miR-122-5p* and an infant's maximum temperature in the first 24 hours after vaccination***
B. *Correlation between post-vaccination expression of hsa-miR-483-5p* and an infant's maximum temperature in the first 24 hours after vaccination***
C. *Correlation between post-vaccination expression of hsa-miR-122-5p* and neutrophil count****
D. *Correlation between post-vaccination expression of hsa-miR-483-5p* and neutrophil count****

**Normalised expression has been plotted as -delta, i.e* $-1 \times (C_{t\,miR-122-5p} - (\sqrt[3]{Ct_{miR-122-5p} + Ct_{miR-122-5p} + Ct_{miR-122-5p}}))$ *so that samples with higher expression post-vaccination lie towards the top of the plot, making the graphs easier to comprehend.*

***Maximum temperature in the first 24 hours after vaccination as measured by ibutton.*

**** Neutrophil count derived by CIBERSORTx algorithm from the whole blood RNA sequencing*

The p-values represent the statistical significance of the relationship between post-vaccination miRNA (normalised by the geometric mean expression of reference miRNAs) and reactogenicity readout (i.e., temperature or neutrophil fraction) after correction for haemolysis. Evaluated using a linear model:

Normalised miRNA expression ~ reactogenicity readout + plasma absorbance at wavelength 414

3.4.4 sRNA-seq of plasma before and 24 hours after vaccination in 12-month-old infants.

The RT-qPCR results proved the sRNA-seq data generated reliable findings; thus, the sRNA-seq experiment was extended to measure plasma sRNA expression pre- and post-vaccination in 12-month-old infants of EUCLIDS (see Figure 10 and Figure 11 for study design).

3.4.4.1 There are no global changes in the expression of miRNAs and tRNAs after 12-month vaccines

The sRNA feature expression profile at 12 months and 4 months were similar (Figure 23 and Figure 12). Few spike-in reads were detected (0-10 reads per sample), rendering them non-useful for normalisation.

The expression of miRNAs was again positively associated with sample haemolysis (Table 16). Unlike at 4 months of age, no global changes occurred after vaccination in tRNA-dfs or miRNAs (Figure 23 Table 16).

*Table 16 Association testing between % feature type abundance and vaccine status and haemolysis. Bold figures highlight results with p<0.05. Red cells: significantly higher post-vaccination, blue cells significantly lower post-vaccination. Linear mixed-effect model = % mapping to feature ~ vaccine status + absorbance at λ 414nm + 1|participant number. * also significant in the 4-month-old analysis.*

Feature	Linear mixed-effect model	
	(P-value for association with	(P-value for association with vaccine
lncRNA	0.07	0.66
mature miR	**0.04***	0.48
offset miR	**0.01***	0.85
rRNA	0.12	**0.01**
snoRNA	0.71	**0.04**
snRNA	0.38	**0.03**
tRNA	0.10	0.46
vault RNA	**0.01**	0.31
viral miR	0.61	0.36
yRNA	0.47	0.89
protein-coding	0.12	0.38
ReverseComplement lncRNA	**0.05***	0.45
ReverseComplement miR	0.35	0.18
ReverseComplement protein-	0.10	0.37
ReverseComplement rRNA	**0.02***	0.34
ReverseComplement snoRNA	0.62	0.31
ReverseComplement tRNA	0.48	0.40
ReverseComplement viral	0.21	0.05
ReverseComplement yRNA	0.53	0.33
unmapped R	0.07	0.11
mapped reads - no local	**0.00***	0.91

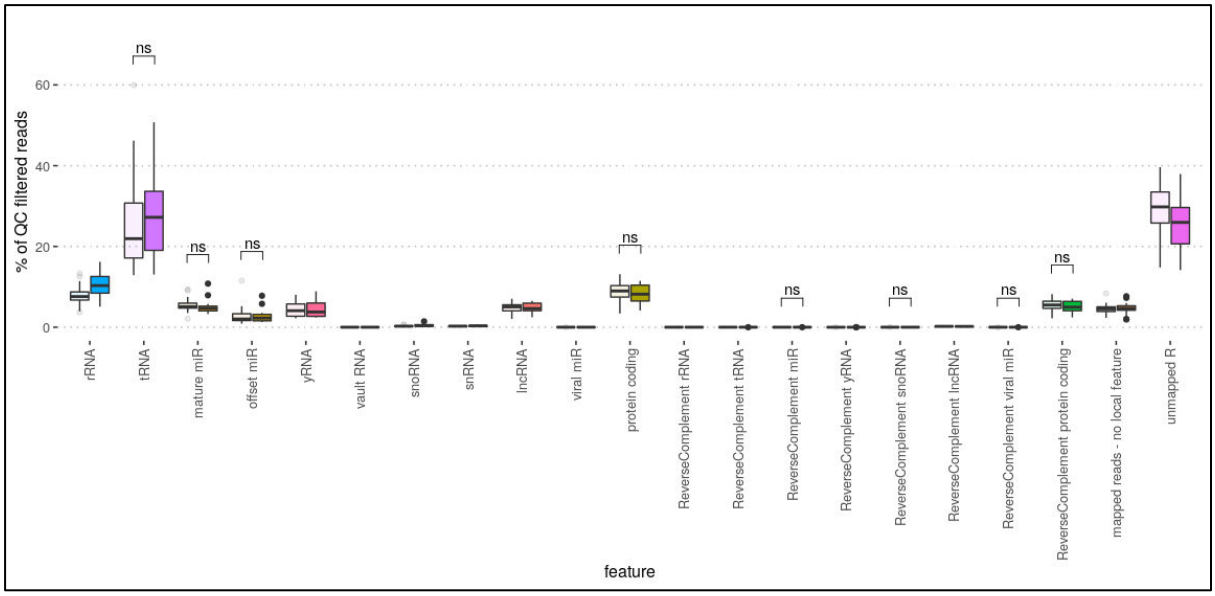


Figure 23 Global changes in the expression of sRNAs after vaccination in 12-month-old infants. Boxplot summarizing the sample-wise distributions in the percentage of reads mapping to each RNA biotype type. Light boxplots and circle points represent pre-vaccine samples, and dark boxplots and triangle points represent post-vaccine samples. P-values were derived from a mixed-effects model where the percentage of all reads mapping to each RNA biotype of RNA was modelled using vaccine status (as a fixed effect), haemolysis (as a fixed effect) and participant ID (as a random effect) as additive effects, i.e.,
% of all reads mapping to each RNA biotype of RNA~ vaccine status+ haemolysis + 1|participant ID. Significance testing was only done on biotypes, which had shown pre versus post-vaccination differences at the 4 month timepoint.

3.4.4.2 Vaccination induces differential expression of miRNAs in 12-month-old infants validating the 4-month-old data

PCA and differential expression analysis were conducted solely on miRNAs to mirror the 4-month-old cohort analysis. As with the 4-month cohort, vaccination globally influenced miRNA expression at 12 months (Figure 24A). Fewer miRNAs were DE in the 12-month-old infants compared with 4-month-olds (2 vs 8, respectively, Figure 14, Figure 24 B). Four miRNAs, hsa-miR-122-5p, hsa-miR-483-5p, hsa-miR-885-3p, and hsa-miR-223-5p were validated as DE at 12-month-old (Square points Figure 24B).

When the vaccine groups were analysed separately, only hsa-miR-1246 was DE in the routine group. Hsa-miR-122-5p and hsa-miR-223-5p were validated as DE in the 4CMenB group. Unlike the 4-month data, vaccine-induced miRNA perturbation was not statistically more pronounced in the 4CMenB group, and there were no significant differences in the post-vaccination fold changes between the vaccine groups.

3.4.4.3 Differential expression of hsa-miR-122-5p, hsa-miR-885-3p and hsa-miR-223-5p is validated when the analysis is restricted to a genuinely independent 12-month-old cohort.

One limitation of using the 12-month-old cohort to validate results is the participant overlap at 4 and 12 months. Seven participants had a sample available at 12 months only: these participants represent a genuinely independent validation cohort. Restricting differential expression analysis to these seven participants again confirmed changes in post-vaccination expression of hsa-miR-122-5p, hsa-miR-885-3p and hsa-miR-223-5p (Figure 25). This finding provides robust evidence that the vaccination regimens in this study change the plasma levels of these three miRNAs in infants.

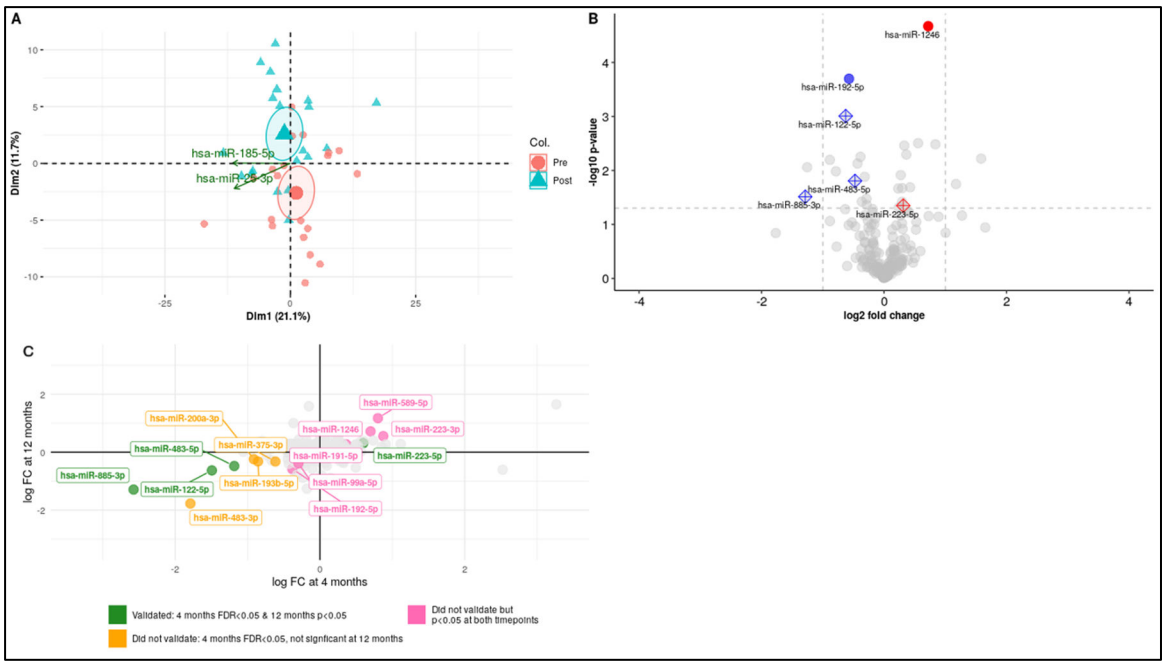


Figure 24 Influence of vaccination on miRNA expression in 12-month-old infants at baseline versus 24 hours

A. *PCA of miRNA expression data (normalized log counts per million, adjusted for haemolysis and effect of the individual). Each dot represents a sample. The centroids of the two groups are indicated with large symbols. The two miRNAs with the greatest contribution to the PCA are green.*
B. *Volcano plots showing results of differential expression analysis when both groups are analysed together. Red= upregulated, blue = downregulated, filled dots: FDR< 0.05 at 12-months, Cross squares: validated DE miRNAs i.e FDR< 0.05 at 4months and p<0.05 at 12-months*
C. *Cross plot comparing log_2 fold changes 24-hours post-vaccination at 4-months versus 12-months. miRNAs that show some evidence of differential expression in one or both cohorts are highlighted. Validated miRNAs are those with FDR< 0.05 at 4 months and p<0.05 at 12 months. miRNAs with p<0.05 in both cohorts are also highlighted.*

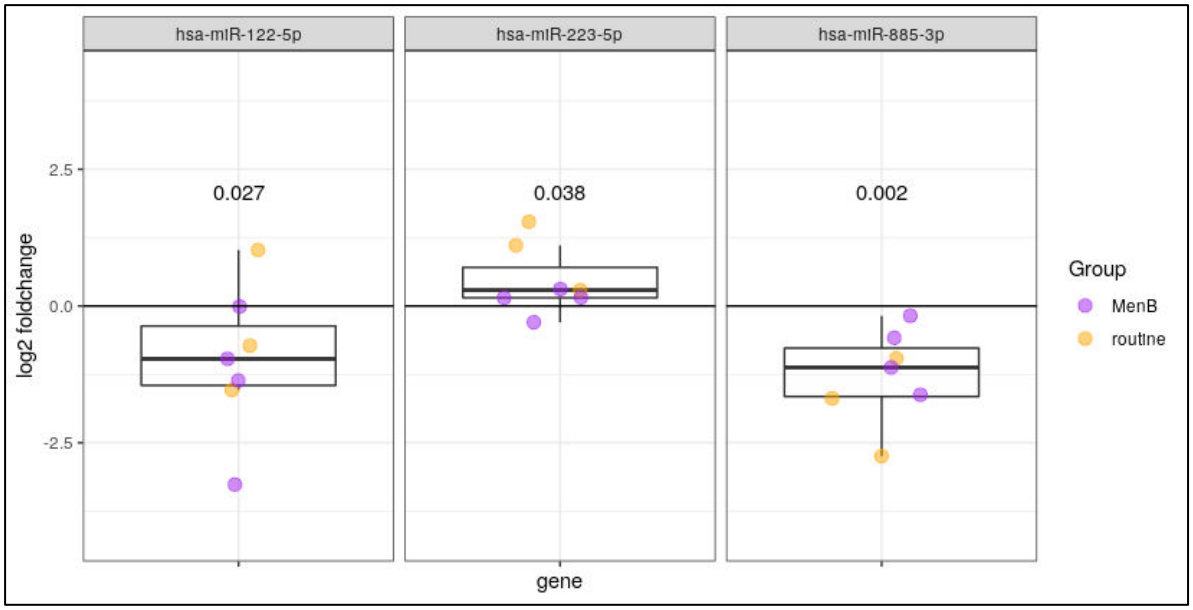


Figure 25 Differential expression results of miRNAs in 12-month-old participants representing a genuinely independent validation cohort. The numbers above each boxplot are p-values. All of these miRNAs were DE at 4 months of age with FDR< 0.05

3.4.4.4 12-month-old data confirms miR-122-5p, miR-223-3p and miR-375-3p as correlates of vaccine reactogenicity

There were fewer significant correlations between post-vaccination miRNA expression and vaccine reactogenicity in the 12-month-old data compared with the 4-month-old data. Post-vaccination expression of hsa-miR-122-5p, hsa-miR-223-3p and hsa-miR-375-3p were confirmed as correlates of post-vaccination neutrophil abundance. In the genuinely independent validation cohort, this correlation was significant for hsa-miR-375-3p (Figure 26). No miRNA correlated with maximum fever.

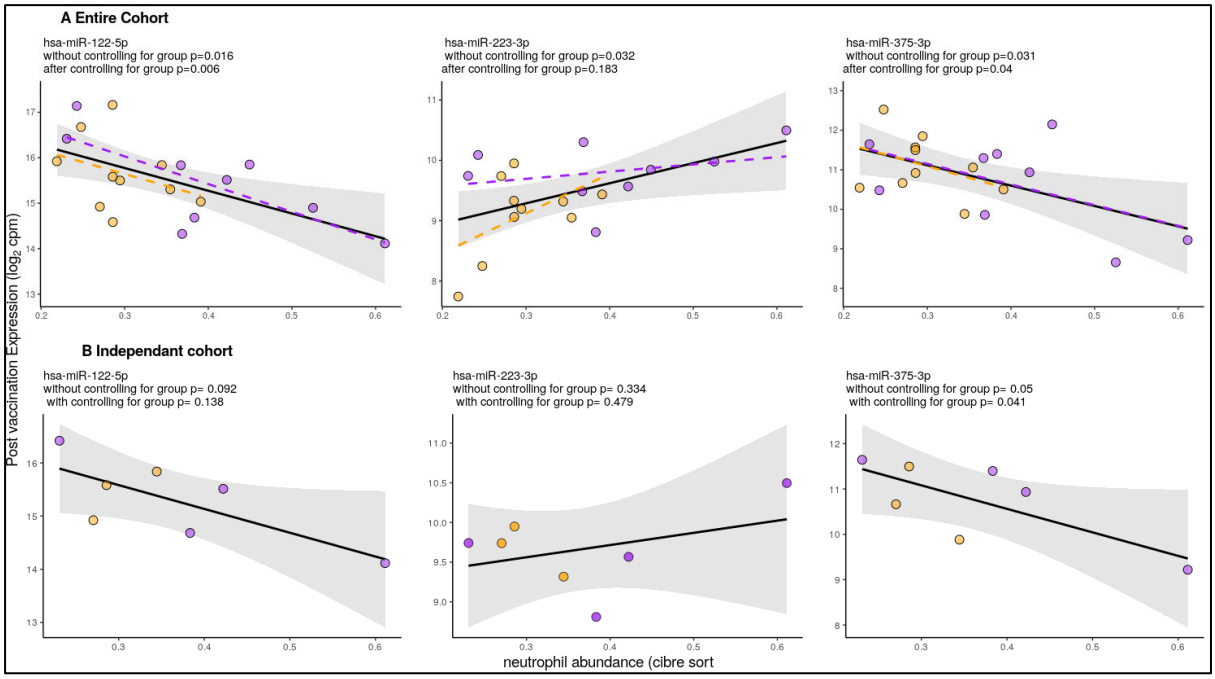


Figure 26 Correlation between post-vaccination expression of miRNAs and neutrophil abundance in the 12-month-old cohort.
A: Correlations in the entire 12-month-old cohort
B: Correlations in the independent subset of the 12-month-old cohort.

3.4.4.5 In 12-month-old infants, derivatives of tRNAs, snoRNAs and protein-coding genes are differentially expressed, validating the 4-month-old infant findings.

At 12 months, PCA showed poorer separation between pre- and post-vaccine samples than in the 4-month-old data. Post-vaccination samples from the 4CMenB group no longer clustered separately from the routine group (Figure 27A).

Eight out of 10 sRNA-dfs DE after vaccination in the 4-month-old data were validated in the 12-month cohort (Figure 27B), including tRNA-dfs, protein-coding-dfs and miRNA hairpins. In the genuinely independent cohort, two sRNA-df features (S100A9 and linc0113) were DE ($p<0.05$). In the 4CMenB cohort,

5/22 sRNA-dfs were validated as DE (hsa-mir-122 hairpin, EIF2B4, tRNA-Asp, CCD144NL-AS1 and S100A9). In the routine group, no sRNA-dfs validated as differentially expressed. No sRNA-dfs were validated as having significantly different pre vs post vaccination foldchanges between the vaccine groups.

Correlation between participants' pre versus post-vaccination log_2 fold change in sRNA-df expression at 4 months and 12 months was not strong (Figure 28, Appendix figure 3); however, there was strong agreement in the direction of log_2fold changes at 4 months and 12 months, especially for sRNA-dfs with $p<0.05$ in both cohorts (95% agreement, 95% CI: 84%-99%, $p<0.01 \times 10^{-7}$). This supports the concept that sRNAs are DE 24 hours after vaccination and suggests only ten features were validated because of inadequate power (else agreement would be random, i.e., ~50%).

No correlations between sRNA-df expression and maximal fever were validated at 12 months.

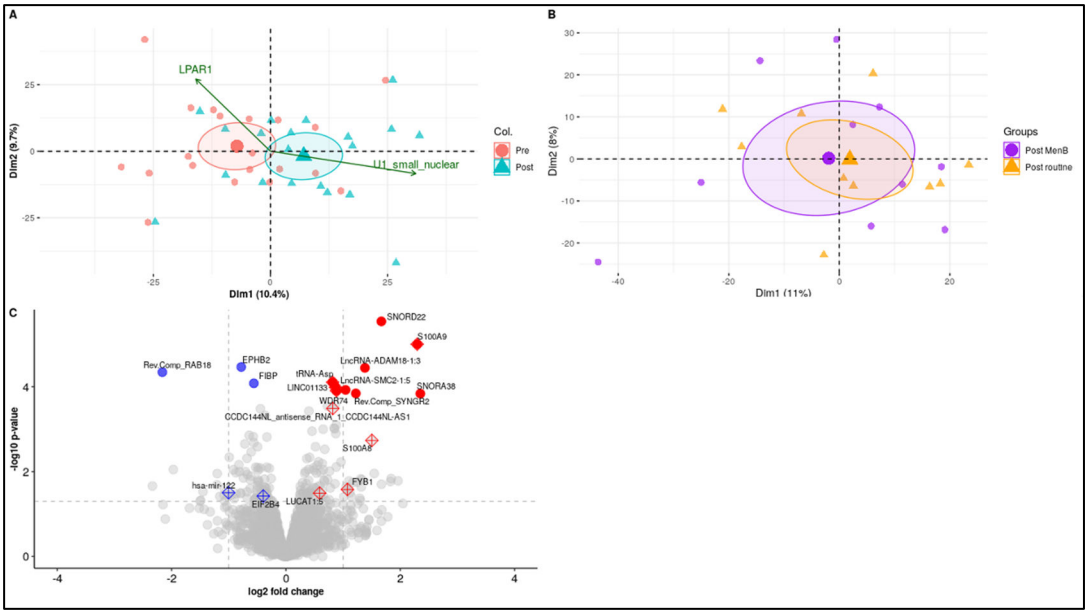


Figure 27 Influence of vaccination on sRNA-df expression in 12-month-old infants at baseline versus 24 hours

A. *PCA of sRNA-df expression data (normalized log counts per million, adjusted for haemolysis and effect of the individual). Each dot represents a sample. The centroids of the two groups are indicated with large symbols. The two sRNA-dfs with the greatest contribution to the PCA are plotted in green.*
B. *PCA of sRNA-dfs expression data (normalized log counts per million, adjusted for haemolysis) of post-vaccination samples. The centroids of the two groups are indicated with large symbols, and the ellipses represent 95% confidence intervals of the centroids of each cluster.*
C. *Volcano plots showing results of differential expression analysis when all participants are analysed together: filled points FDR< 0.05 at 12 months, squares: validated DE sRNA-dfs, i.e. FDR< 0.05 at 4 months and p<0.05 at 12 months. Circles: DE at 12 months only.*

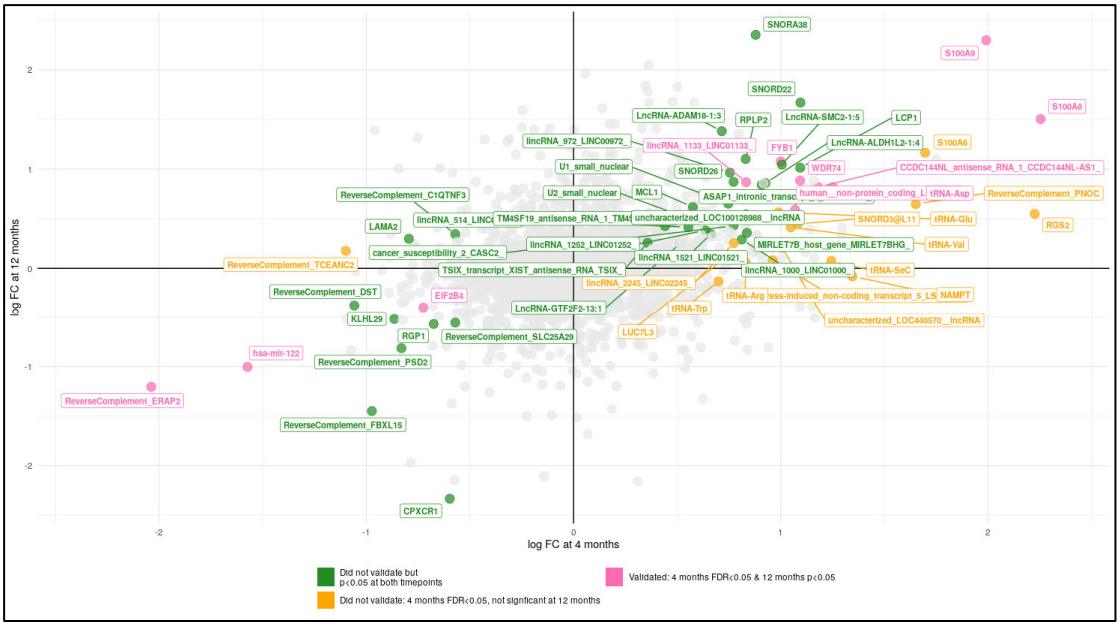


Figure 28 Cross plot comparing log_2 fold changes 24-hours post-vaccination at 4 months versus 12 months. sRNA-dfs that show some evidence of differential expression in one or both cohorts are highlighted. Validated sRNA-dfs had FDR< 0.05 at 4 months and p<0.05 at 12 months. sRNA-dfs with p<0.05 in both cohorts are also highlighted. sRNA-dfs in the upper left and lower right quadrants have disagreeing directions of fold changes in the 4 month and 12-month datasets.

3.4.5 Identifying the cellular and exRNA source of differentially expressed miRNAs

3.4.5.1 Changes in circulating exRNA carriers pre versus post-vaccination are inferred by a deconvolution algorithm

Vaccine-associated global changes in sRNA biotypes in the 4-month-old participants may reflect changes in circulating exRNA vehicle abundance (e.g., AGO, microvesicles, exosomes, HDL, apoptotic bodies). A deconvolution algorithm (XDec) was applied to explore this possibility. XDec deconvolutes samples into a set of exRNA carrier profiles that it then compares with previously described exRNA carriers (referred to as CT1, CT2, CT3A, CT3B, CT3C, and CT4 in Murillo et al.[349]). XDec identified three ex-RNA carrier profiles in the EUCLIDS

samples: carriers 1, 2, and 3. Carrier 1 has a similar sRNA profile to CT1, CT2 and CT3C, which Murillo et al. assigned as low-density vesicles, HDL, and an uncharacterised exRNA carrier. Carrier 2 is similar to CT3, which Murillo et al. ascribed RNA-protein complexes, and carrier 3 is similar to CT4, ascribed as extracellular vesicles of variable density (see heat map in Figure 29A and exRNA carrier description in Murillo et al.[349]).

Vaccination was associated with a relative increase in carrier 1 (~low-density vesicles) and a decrease in carrier 2 (~RNA-protein complexes) at 4 and 12 months of age (Figure 29B).

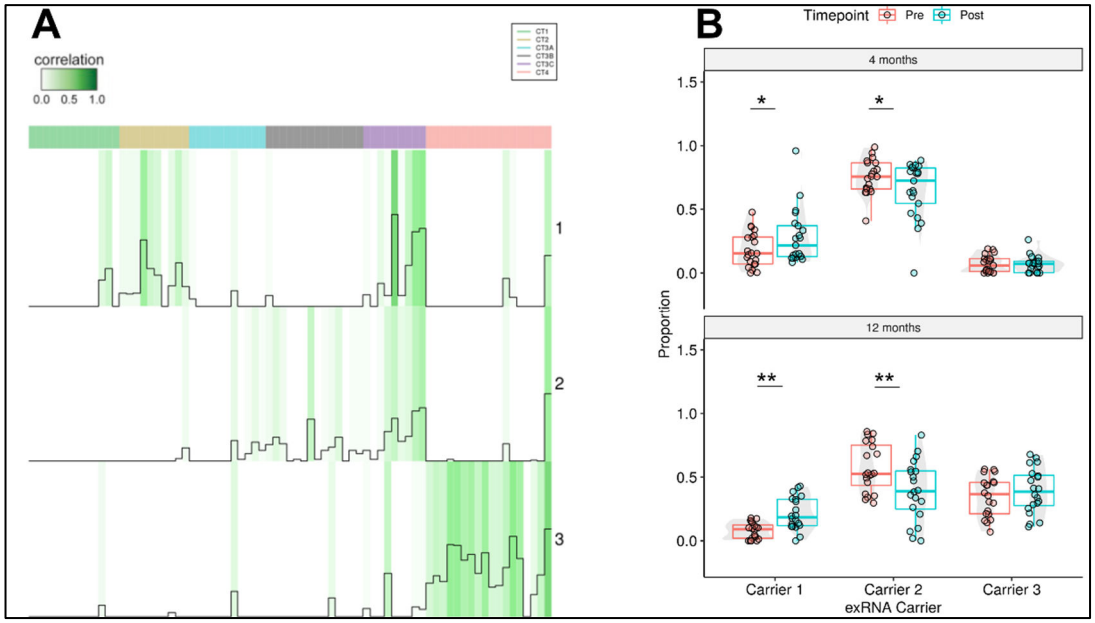


Figure 29 XDEC exRNA carrier deconvolution results.
A. Heatmap auto-generated by the XDec deconvolution algorithm. The strength of colour represents the correlation between each exRNA carrier(1:3) identified in the EUCLIDS dataset (rows) and the reference sample profiles of each of the carrier types (CT1:CT4) previously identified in the XDec paper[349] (columns). Note that exRNA carrier 3 in the EUCLIDS data strongly correlates to the CT4 profiles of reference samples (pink bar) which has a similar profile to the CT4 exRNA carrier described as "extracellular vesicles of variable density".

*B. Relative proportion of each exRNA carrier contained within each sample grouped by timepoint and cohort. (paired Wilcoxon *p<0.05. **p<0.001).*

3.4.5.2 Cellular sources of differentially expressed miRNAs

Cell expression atlases were used to infer possible cellular sources of DE miRNAs (Table 17). The liver was the top cell type for downregulated miRNAs (hsa-miR-122-5p, hsa-miR-483-5p, hsa-miR-483-3p, hsa-miR-885-3p). The top cell type for upregulated miRNAs was leukocytes (hsa-miR-223-5p, hsa-miR-200a-3p, hsa-miR-145-3p, hsa-miR-223-3p hsa-miR-338-5p), particularly CD15+ (a myeloid-specific marker) cells. MIEAA2 enrichment analysis using all miRNAs (ranked by their signed likelihood ratio based on their pre- versus post-vaccination expression) and miRNA:mRNA pairings from the miRNATissueAtlas2

showed significant positive enrichment in leukocytes and depletion in liver and muscle. Depletion in liver-expressed miRNAs was validated in the 12-month-old data (Appendix Table 5).

Table 17 Cellular distribution of differentially expressed miRNAs according to miRNA expression atlases.
** based on the highest median expression. ** note blood cells not included in tissue atlas*

miRNA	**miRNATissueAtlas2*****[350]	**Blood atlas***[311]	**FANTOM5 database**[311]		
	Top cell type (s) according	**expressed cell type (bold = major cell type)**	**Top three cell types that the miRNA is expressed in**	**Cell type(s) miRNA is enriched in**	**Cell type(s) miRNA is depleted in**
hsa-miR-122-5p	Liver	Not detected	Hepatocyte Polarised intestinal cell IPS derived from CD19+ B-cells	Hepatocyte, endodermal cell, epithelial cell	mesodermal cell, squamous epithelial cell
hsa-miR-483-5p	Liver	Not detected	Amniotic Epithelial cell Placental epithelial cell Chorionic membrane cell	nil	nil
hsa-miR-483-3p	Liver, pleura, heart	Not detected		Extraembryonic cell, epithelial cell of amnion, placental epithelial cell	hematopoietic cell, leukocyte, myeloid leukocyte

Table 15 (continued)

hsa-miR-885-3p	Submandibular gland, liver	Not detected	Pineal gland Fibroblast Hepatocyte	nil	nil
hsa-miR-223-5p	bone	**CD15+** CD14+, CD19+, CD56+	CD14+ monocytes Neutrophils Mast cell CD 19+ B-cells	nil	nil
hsa-miR-193b-5p	Muscle and adipocyte	Not detected	Chorionic membrane cell Placental epithelial cell Amniotic membrane cell	nil	nil
hsa-miR-200a-3p	Trachea, submandibular gland, kidney	CD14+, **CD19**+, CD56+	Renal Proximal tubular cell Renal cortical Epithelial cell Intestinal epithelial cell	endo-epithelial cell, epithelial cell, respiratory epithelial cell	leukocyte, hematopoietic cell, myeloid leukocyte
hsa-miR-375-3p	Pancreas	Not detected	Pituitary gland Pineal gland IPS derived from CD19+ B lymphocytes	epithelial cell, endodermal cell, endo-epithelial cell	blood vessel endothelial cell, endothelial cell, hematopoietic cell
hsa-miR-145-3p	Bladder, artery, stomach	**CD15+**	Synoviocyte Adipocyte Mesenchymal stem cell	nil	nil
hsa-miR-223-3p	bone	CD4+, CD8+, CD14+, **CD15+,** CD18+, CD56+	Neutrophils CD14+ monocytes Mast cell	leukocyte, hematopoietic cell, myeloid leukocyte	epithelial cell, muscle cell,endo-epithelial cell

Table 15 (continued)

hsa-miR-4446-3p		Not detected	CD3 + T-lymphocytes CD19 + B Lymphocytes Fibroblast	leukocyte, hematopoietic cell, lymphocyte	mesodermal cell, neutrophil
hsa-miR-338-5p	Brain	CD14+, **CD15+**, CD56+	Spinal cord CD14+ monocytes Neutrophils	nil	nil

3.4.6 The function and targets of differentially expressed miRNAs

MIEAA2 GSEA analysis (methods chapter Figure 9) showed enrichment in 89 gene ontology biological process (GO:BP) terms pre versus post-vaccination in the 4-month-old participants. Top hits related to the immune response, lipid metabolism, gene regulation and muscle. Coagulation, cardiac, neuronal, ion transport, and cell signalling pathways were also identified (Figure 30, Appendix Table 5). Twenty-three of these pathways were validated in the 12-month-old data (labelled pathways in Figure 30).

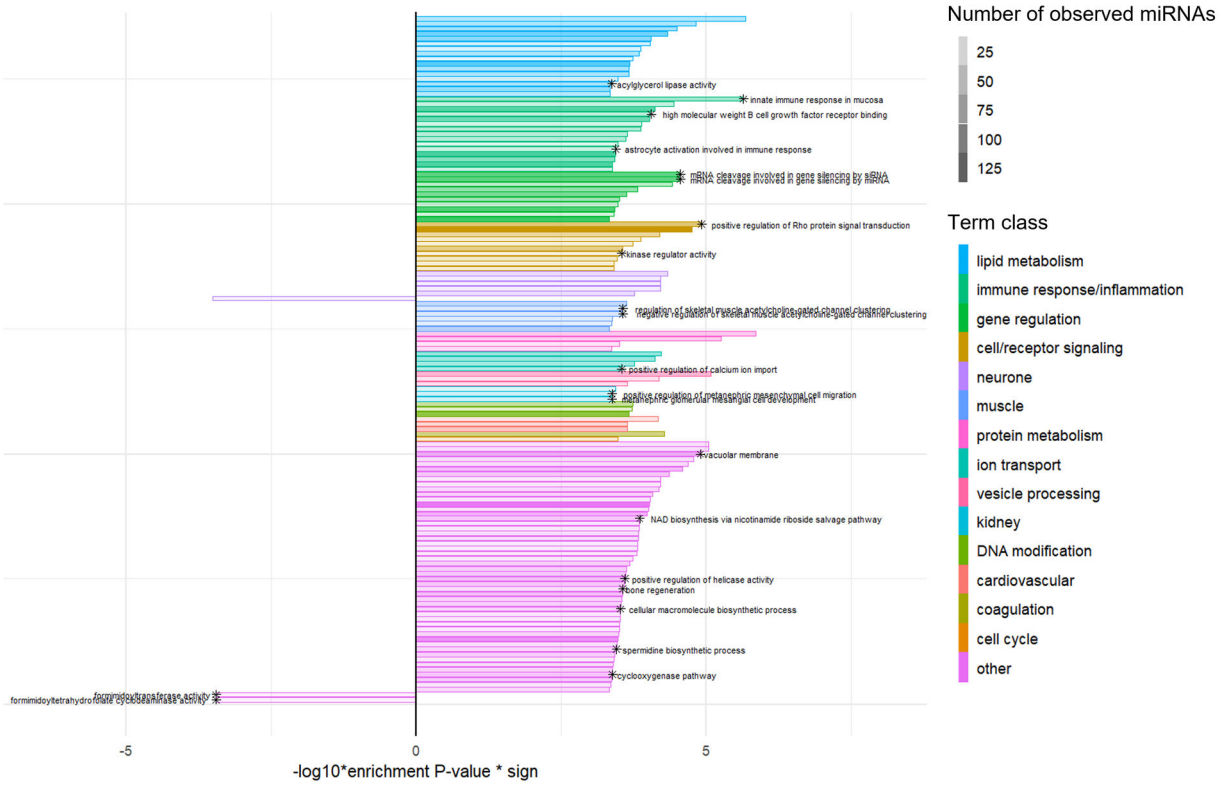


Figure 30 Gene set enrichment results for the 4-month-old cohort using all miRNAs. GSEA analysis was performed on the MIEAA2 web program using GO genesets from annotations derived over miRTarBase. Colour represents term class. Intensity of colour bar represents term size. Starred-labelled points were validated in the 12-month-old cohort—all terms passing FDR<0.05 can be found in Appendix Table 5.

3.4.7 A mouse vaccination model reveals that a vaccine-induced reduction of plasma miR-122-5p is a cross-species phenomenon linked to vaccine-related temperature rises.

Collaborating with others working on a mouse vaccination model enabled investigation of whether vaccine-induced changes in circulating miRNA could be verified. Only miR-122-5p was adequately detected in the mice plasma by RT-PCR. Liver samples were available for all mice, but plasma samples were only available for a subset of mice due to technical challenges in obtaining samples (Figure 31).

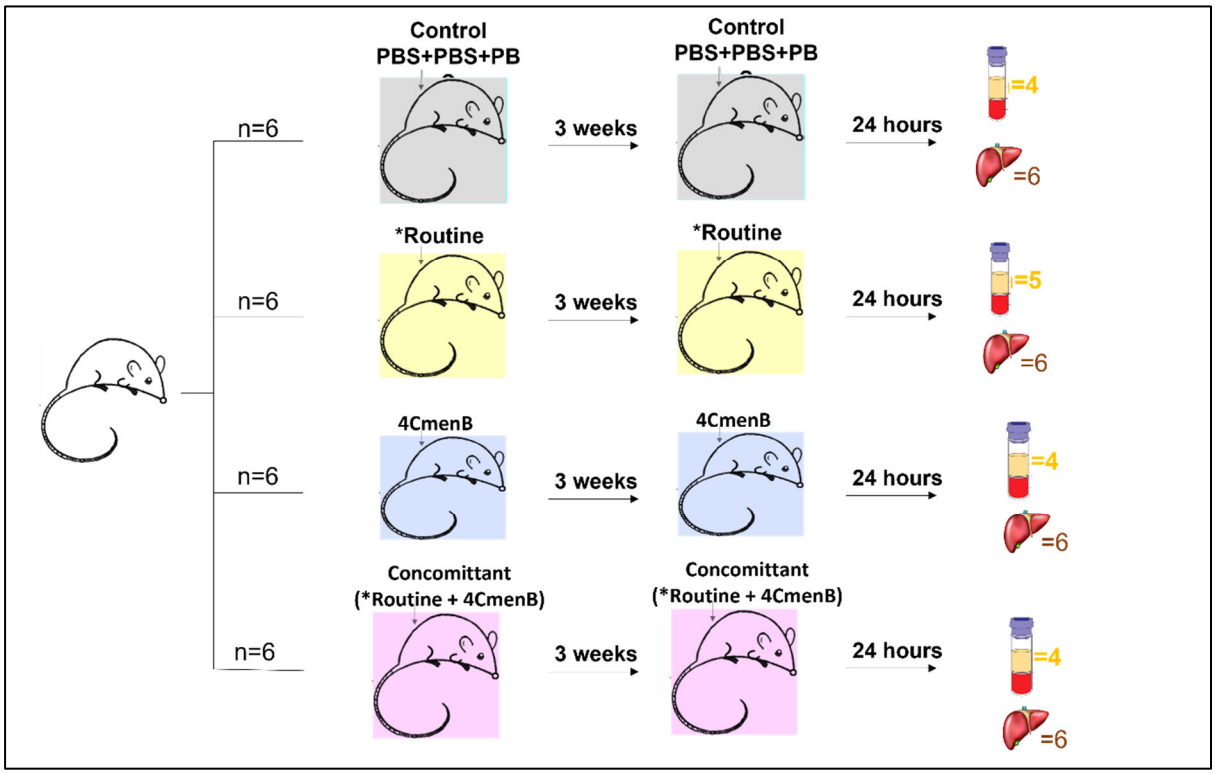


Figure 31 Experimental design to determine whether vaccination influences miRNA expression in mouse plasma and liver. The number of samples per mouse condition is illustrated next to plasma and liver icons. Routine vaccines are those that were given to 4-month-old infants.
***Control group** = PBS + PBS + PBS;*
***Routine group** = Pediacel + Prevnar + PBS;*
***4CMenB group** = 4CMenB + PBS + PBS;*
***Concomitant group** = 4CMenB + Pediacel + Prevnar*

3.4.7.1 Concomitant vaccination reduces plasma miR-122-5p expression in mice, correlating with post-vaccination temperature.

After vaccination, Mice receiving 4CMenB alongside routine vaccines had significantly lower plasma expression of miR-122-5p (Figure 32A) than controls. In addition, maximal body temperature in the 24 hours after vaccination negatively correlated with plasma miR-122 expression (R -0.6, p=0.0067, Figure

32B). The results remained unchanged regardless of which reference genes (spike-in genes or miR-30b-5p and miR-103a-3p) were used (Appendix Figure 7).

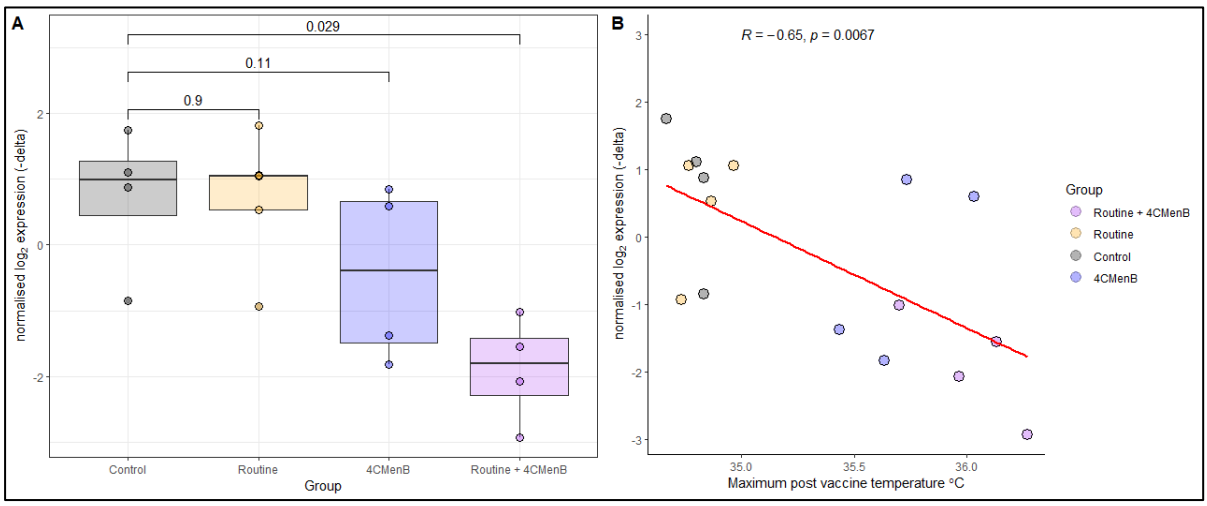


Figure 32 Expression of miR-122-5p in mice plasma 24 hours after vaccination

A. Normalised expression of miR-122-5p plasma 24 hours after vaccination in each vaccine group. Significance values for comparisons were tested using the Wilcoxon rank-sum test.

B. Scatterplot of the maximum temperature of a mouse in the 24 hours after vaccination versus the normalised expression of miR-122-5p. The correlation was tested using the Pearson correlation coefficient. Line of best fit is plotted in red.

3.4.7.1.1 Vaccination does not induce a change in hepatic miR-122-5p expression

There was no difference in the expression of miR-122-5p in whole mouse liver 24 hours after vaccination and no correlation with maximum temperature.

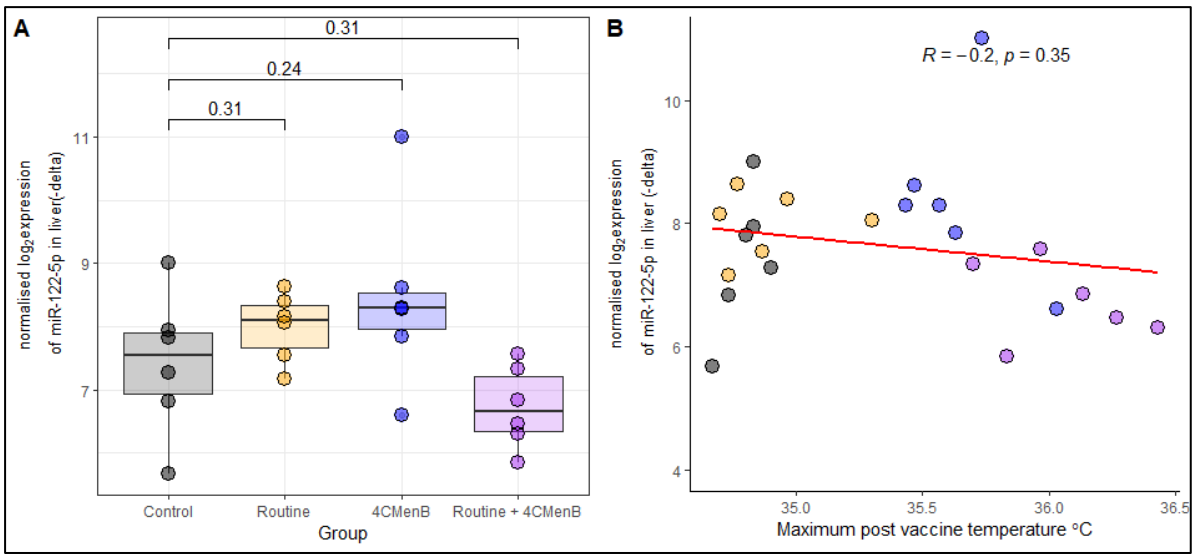


Figure 33 Normalised expression of miR-122-5p in mouse liver 24 hours after vaccination in each vaccine group. Significance values for comparisons were tested using the Wilcoxon rank-sum test. miR-122-5p expression is normalised to snRNA U6

3.5 Discussion

3.5.1 Summary of key results

This is the first study showing plasma miRNA, and sRNA-df expression is altered 24 hours after vaccination. miRNAs likely arising from leukocytes (probably myeloid cells) and liver were generally upregulated and downregulated after vaccination, respectively. Furthermore, downregulation of miR-122-5p was recapitulated in a mouse vaccination model. Changes in plasma sRNA expression were more profound in young infants and after concomitant 4CMenB administration. Significant correlations existed between post-vaccination plasma miRNA expression and fever in infants at 4 months of age and were mirrored in the mouse model but not seen in the 12-month-old infants. In addition, expression of hsa-miR-122-5p, hsa-miR-223-3p and hsa-miR-375-3p correlated with post-vaccination neutrophil abundance. Each finding is discussed below.

3.5.2 miRNAs comprise a small proportion of sRNA features in plasma

The sRNA profile of this study is similar to other sRNA-seq studies of plasma and serum with tRNA-dfs the most abundant RNA biotype[349]. This contrasts with intracellular profiles where miRNAs predominate and is in keeping with asymmetric cellular release, uptake or circulatory degradation of sRNA biotypes.

3.5.3 Vaccination influences miRNA expression in plasma and is associated with vaccine reactogenicity.

This study provides robust evidence that vaccination changes circulating miRNA expression; firstly, miRNAs were DE after children's 4-month-old vaccines, and RT-PCR technically confirmed this; secondly, differential expression was confirmed in 12-month-old participants, including in the seven infants that represented a genuinely independent 12-month-old cohort; and thirdly, miR-122-5p was downregulated in mouse plasma 24-hours post-vaccination when 4CMenB was given alongside routine vaccines.

RT-PCR suggested that the time course of differential expression of miRNAs was in keeping with the mRNA dataset, with changes most obvious 24 hours after vaccination.

The present study supports a relationship between markers of vaccine reactogenicity (fever and neutrophil abundance) and plasma miRNA expression—this relationship was stronger in the 4-month cohort and present in the mouse study. At 4 months, perturbation in miRNA expression was more significant

when the reactogenic 4CMenB was co-administered. Possibly 4CMenB alone induces perturbation in miRNA expression, or simply administration of an extra vaccine strengthens the magnitude of the immune response. Like the 4-month-old infant results, post-vaccination mouse temperature negatively correlated with miR-122-5p. No previous studies link miRNA expression and vaccine reactogenicity in animals or man.

No miRNAs correlated with group B meningococcal SBA titres at 4 months of age. However, this test was limited by the low number of samples with SBA titres (7 participants). Interestingly, miRNAs whose baseline expression best correlated with SBA titres were miR-let7b, miR-125b, miR-146b and miR-150. These miRNAs are highly expressed in lymphocytes and frequently associated with infection[35]. None of these miRNAs passed multiple-testing threshold, so over-interpretation of an association should be avoided. RT-qPCR on the baseline samples from the other EUCLIDS groups (i.e., groups 1,3,4,5,7,8 in Figure 10) could be run to determine whether baseline miRNA is associated with SBA titres.

3.5.4 Where, why, and how of differentially expressed miRNAs

Changes in miRNA expression after vaccination presumably reflect perturbation of innate immune responses. Is there a biological purpose for miRNA perturbation, or is it a bystander effect of immune processes, e.g., changes in blood cell composition? To address this matter, three questions must be answered: what is the cellular source of these miRNAs? What mechanism underlies their differential expression? What do these miRNAs target?

3.5.4.1 exRNA carriers and DE miRNAs

Changes in plasma miRNA expression could be due to altered levels of a specific exRNA carrier, as different ex-RNA carriers contain different compositions of RNA biotypes[349]. For example, HDL levels are altered by inflammation[351]. Changes in circulating exRNA carriers could explain vaccine-induced global changes in tRNA-dfs and miRNAs in 4-month-old infants. XDec deconvolution identified three principal exRNA carriers in the plasma samples, of which there was a relative increase in carrier 1 (which has a profile similar to low-density vesicles) and a decrease in carrier 2 (which has a similar profile to ribonucleic acid protein complexes) after vaccination. Changes in exRNA carrier type were more marked at 12 months, which was surprising given sRNA differential expression was more marked at 4 months. This could be because DE miRNAs are due to changes within an exRNA carrier type rather than a change in exRNA carrier composition. To dissect this further, it would be interesting to separately assay the DE miRNAs in each exRNA carrier in the plasma samples.

3.5.4.2 Cell sources of DE miRNAs

Muscle, leukocytes and the liver are likely sources of DE miRNAs as these tissue types participate in immune responses, and DE miRNAs are enriched in leukocytes and depleted in liver and muscle. The relationship with muscle is supported by literature showing plasma and serum levels of downregulated miRNAs (hsa-miR-122-5p, hsa-miR-483-5p, hsa-miR-483-3p, hsa-miR-885-3p) are linked to liver damage and disease[352–355].

The liver is particularly likely to be the source of hsa-miR-122-5p as this miRNA comprises up to 70% of the hepatic miRNA milieu, and the liver is the top cell type for miR-122-5p expression across all atlases[356,357]. Countering this, the mouse model did not show a change in hepatic miR-122-5p after vaccination. There are multiple possible explanations. Firstly technical issues could have precluded detection of differential expression in the liver, for example, sub-normalisation by the reference snRNA U6 (though sensitivity analysis using other potential housekeeping genes did not change the results). Secondly, low power from the small sample size may have precluded detection of differential expression; however, this seems unlikely as the trend is not convincing. Thirdly, timepoint may be an issue; a study of LPS injection in mice showed changes in serum miR-122-5p occurred at 24 but not 4 hours after vaccination, whereas the opposite was true within liver tissue[358]. Fourthly, if a small fraction of hepatocyte hsa-miR-122-5p is released into plasma, there may be no perceptible change in intrahepatic hsa-miR-122-5p levels. Fifthly, changes in hsa-miR-122-5p plasma levels may arise from a change in miRNA expression from a different organ. Though the liver is the major producer of miR-122-5p, the miRNA is identifiable in other organs and cell types, including the spleen, gut mucosa, kidney, respiratory system, keratinocytes, and CD19+ B-cells, albeit sometimes under inflammatory conditions[358–366]. Finally, a drop in miR-122-5p may be due to uptake of liver-derived miR-122-5p by non-hepatic cells—this is supported by enrichment analysis in Chapter 5. Future iterations of the mouse experiment should include more mice, a 4-hour post-vaccine timepoint and measurement of miR-122-5p in other organs, particularly the spleen and blood cells.

Perturbation in miRNA expression may also be due to leukocyte activity as leukocytes were identified as a top cellular source for hsa-miR-223-3p, hsa-miR-223-5p, hsa-miR-200a-3p, hsa-miR-145-3p, hsa-miR-338-5p and hsa-miR-4446-3p) particularly in CD15+ cells—a myeloid-specific marker.

Using the cell atlases to infer cellular sources of miRNAs has limitations. Firstly miRNA cell expression atlases sometimes disagree—for example, the FANTOM5 expression atlas reports hsa-miR-4446-3p as enriched in leukocytes, but the blood atlas did not detect it[311,367]. Secondly, not all tissues and cell types are represented in the atlases—for example, hsa-miR-223-5p is highly abundant in platelets, which are missing from the FANTOM5 atlas[368]. Thirdly, the top cellular location of miRNAs is not necessarily the source of DE miRNAs. Future studies should aim to isolate and identify the extracellular vesicles carrying DE miRNAs in infant plasma. It may be possible to elucidate the parent cells of those miRNAs through assays of the cell surface markers on those extracellular vesicles.

3.5.4.3 The mechanism underlying vaccine-mediated differential expression of miRNAs.

Why vaccination changes plasma miRNA expression is uncertain. Changes could reflect a shift in blood cell composition or alteration in miRNA expression within cells, which passively leak out, or could reflect altered active secretion or uptake by cells: all scenarios are supported in the literature[160,161,369–371].

Vaccine reactogenicity has been linked to both LPS and peptidoglycan. Models of LPS stimulation show changes in intracellular levels of miR-122-5p, miR-223-5p, and miR-885-3p; however, there is very little literature on peptidoglycan and miRNA expression[358,372–374]. Given the liver may be the primary source of changes in plasma miR-122-5p and miR-885-3p, such changes are likely stimulated by cytokines released elsewhere rather than TLR agonist action directly on hepatocytes; therefore, literature on *in-vitro* models of LPS and peptidoglycan stimulation may be less relevant to the miRNA findings. Interestingly, no TLR4 or LPS pathways were enriched in the miRNA MIEAA2 gene set analysis; however, the GO:BP term "positive regulation of toll-like receptor 2 pathway" was enriched for targeting at 4 months but not 12 months. TLR2 is the receptor for peptidoglycan.

Changes in cellular composition should be considered a potential contributing factor in differential expression, particularly given miR-223-3p and hsa-miR-338-5p correlate with neutrophil counts after vaccination. An increase in leukocyte-derived miRNAs post-vaccination could “dilute” the relative abundance of non-leukocyte-derived miRNAs, potentially explaining the drop in miR-122-5p plasma levels. However, two things counter this possibility; firstly, unnormalized RT-qPCR results also show a drop in miR-122-5p, suggesting an absolute (not just relative) reduction in miR-122-5p levels (note cDNA submission to RT-qPCR is not normalised to plasma RNA content); secondly, the absence of baseline correlations between leukocytes and miRNAs do not suggest that changes in miRNA expression represent changes in cell composition.

Changes in miRNA release may reflect an alteration in intracellular miRNA expression; however, cell miRNA excretion does not always mirror intracellular miRNA expression. Instead, selective sorting occurs possibly as a miRNA waste-disposal process or intercellular communication[160,161]. Measuring intracellular and extracellular miRNA expression before and after stimulation of leukocytes and liver *in-vitro* could resolve this question. Regardless of the excretion mechanism, plasma miRNAs may play a functional role in innate immunity particularly given intercellular transmission of miR-122-5p, and miR-223-5p has been shown to play a role in response to infection and immunomodulation[168,174,358,375].

3.5.5 Role of differentially expressed miRNAs within the immune system

Chapter 1 details the key role miRNAs play in leukocyte development, differentiation, and effector functions. The role of miR-122-5p is worth exploring because it is one of the most abundant plasma miRNAs, was the most DE miRNA at 4 months, correlated with post-vaccination neutrophil counts, and its downregulation is cross-species (mouse model results). MiR-122-5p plays a role in the innate immune response of hepatocytes in response to direct targeting by pathogens[376,377]. Tissue and circulating miR-122-5p is DE in several infectious and pro-inflammatory states in humans, including hepatitis B and C, varicella, inflammatory bowel disease, psoriasis, and other autoimmune diseases[35,358,365,366,378,379]. MiR-122-5p has been identified paradoxically as over and under-expressed in plasma in sepsis (though this may relate to liver injury rather than immunology)[244,247,380,381]. Studies show miR-122-5p released from cells has biological consequences; for example, the LPS mouse study referred to

above showed an increase in liver-derived miR-122-5p plasma release induced anaemia by negatively regulating renal erythropoietin, and a study by Collison et al. showed extracellular miR-122-5p promotes lung injury during viral infections[358,382]. miR-122-5p modulates critical components of the inflammatory response by directly or indirectly targeting LR4, SOCS1, MIP-2, and PACT and by promoting cytokine secretion in stimulated whole blood[377,382–385].

The predominant expression of miR-122-5p by hepatocytes suggests the liver could be involved in vaccine responses. It is plausible that changes in intrahepatic miR-122-5p levels could relate to induction of hepatically derived acute phase proteins[386]. The failure of the mouse model to identify intrahepatic changes in miR-122-5p expression after vaccination counters this notion, but there are limitations to that model (discussed above).

miR-223-5p is also worth prioritising for follow-up, as it is often identified in studies of infection and inflammation and is mainly expressed in myeloid-derived leukocytes and platelets. MiR-223-5p is often described as an anti-inflammatory miRNA; however caution is warranted when ascribing broad, overly simplistic functions to miRNAs as they may play pleiotropic roles depending on the cellular, temporal and transcriptomic context[387,388].

3.5.6 Targets of differentially expressed miRNAs

DE miRNAs in this study have the potential to collectively target hundreds of mRNAs, some of which are known to be involved in the immune response; for

example, miR-223-3p targets IL6. The targets of the miRNA DE in this study will be covered in more detail in Chapter 5.

3.5.7 Vaccine-induced changes in sRNA-dfs

In the 4-month-old data, a global reduction in miRNAs occurred alongside an increase in tRNA-dfs. Competition between reads for finite sequencing space on the flowcell makes it impossible to determine whether these changes are due to variations in miRNAs, tRNAs or both. The mechanism underpinning global difference RNA-biotype expression is unclear. ANG cleavage of tRNAs during cellular stress responses increases tRNA-dfs production. Endoplasmic reticulum stress response in Jurkat pre-T-cell line globally increases t-RNA fragments and decreases miRNAs mirroring the changes in the 4-month-old data[371]. Evidence that changes in tRNA-dfs abundance after vaccination might have an immunological purpose comes from studies showing tRNAs are selectively sorted into exosomes during T-cell activation, and one of the DE tRNAs derived from tRNA-Glu is induced by cellular infection with respiratory syncytial virus[389].

Global differences in miRNA and tRNA expression were not replicated in the 12-month cohort. This could reflect age/vaccine regime-related differences between the cohorts or reflect a chance finding (i.e. type 1 error) in the 4-month cohort.

sRNA-dfs arising from lincRNAs and protein-coding genes were DE in the 4 months and 12-month cohorts. sRNA-dfs derived from mitochondrial genes were DE in the 4-month data but only in the 4CMenB group. One explanation could be differences in sample quality between the vaccine groups, given a high amount

of mitochondrial RNA is associated with poor-quality samples (i.e. dead and dying cells) in single-cell RNA-seq. However, these are cell-free samples; the level of haemolysis was similar between the groups, and there was no association between mitochondrial sRNA-dfs and white cell counts. Another explanation is that downregulation of mitochondrial sRNA-dfs in the 4CMenB group reflects changes in cellular metabolic activity when 4CMenB is co-administered to younger infants. However, this hypothesised age-specific effect is confounded by batch effects; therefore, the study needs repeating and samples from both cohorts processed and analysed in tandem to determine if it is true.

The biological significance of DE sRNA-dfs is contentious, as there is less evidence they have a biological function. These sRNA-dfs may represent degradation fragments of genes, which are differentially expressed; however, the expression profile of reads mapping to these sRNA-dfs suggests otherwise, as they show a distinct peak at loci within the genes (Appendix Figure 29) rather than a random distribution along the gene typical of degradation fragments. The expression of these DE sRNA-dfs is correlated with their mRNA counterpart in the whole blood mRNA sequencing data in Chapter 5.

3.6 Conclusion

Infant vaccination is a model of controlled stimulation of the innate immune system in infants and may model their early response to infection; thus, the changes seen here may be part of a generalised innate immune response that is potentially conserved across species in the case of miR-122-5p. Differential

sRNA expression may be induced by a systemic response to vaccination, e.g., mediated by cytokine release from the vaccine site. In addition, the association between vaccination and miRNAs downregulated by the liver suggests a possible link between the hepatic acute phase response and ex-miRNA release. Based on this supposition, Chapter 4 looks at the expression of plasma proteins post-vaccination in these infants. To determine whether there appears to be a relationship between plasma expression of miRNAs and gene regulation after vaccination, the miRNA, protein and mRNA datasets are integrated in Chapter 5.

4 The Influence of Vaccination on the Infant Plasma Proteome

4.1 Introduction

There is a paucity of studies investigating how vaccination influences plasma protein expression. Measuring plasma proteins can provide insights into cross-talk and signalling between cells and organs involved in the immune response and can shed light on what is happening in organs and tissues that participate in the immune response outside blood, e.g., liver, muscle and endothelium. This is particularly important for understanding how and why vaccination induces changes in plasma miRNA expression in the EUCLIDS study, some of which are highly expressed in liver and leukocytes. Furthermore, it is worth identifying discordance between mRNA and protein expression in EUCLIDS as instances could be linked to post-transcriptional gene regulation by sRNAs.

Plasma samples generated in the EUCLIDS study were used to investigate the influence of vaccination on the plasma proteome.

4.2 Chapter aims

Chapter aims are:

1) Describe the influence of vaccination on plasma protein abundance
2) Evaluate whether concomitant administration of 4CMenB alongside routine immunisations influences the plasma proteome

3) Establish whether plasma proteins represent biomarkers of vaccine reactogenicity
4) Characterise the pathways in which DE proteins act.

4.3 Chapter Methods

4.3.1 Timepoints studied

As summarised in Figure 11, protein abundance was measured before and four and 24 hours after participants' 4 month and 12 month vaccinations. Twenty-four hours was chosen so that protein and sRNA-seq datasets could be integrated and coincides with maximal differential expression in the mRNA data. Four hours post-vaccination was chosen to identify early changes that may initiate and orchestrate changes seen at 24 hours.

4.3.2 Sample selection

All participants with a baseline and post-vaccination sample at 4 and/or 12 months were included in the proteomics study (Figure 11 contains sample numbers at each time point).

4.3.3 Mass spectrometry

Mass spectrometry was used to measure acute-phase proteins because it does not require pre-specification of the proteins of interest and picks up the top several hundred plasma proteins, making it ideal for studying high-abundance proteins made by the liver, e.g., acute-phase proteins.

Platelet-poor plasma was submitted to the Target Discovery Institute, University of Oxford, UK, for sample preparation, liquid chromatography with tandem mass spectrometry (LC-MS/MS), and raw data processing.

4.3.3.1 LC-MS/MS on plasma from infants at 4 months of age

Plasma was subject to top 12 protein depletion (Pierce™ Top 12 Abundant Protein Depletion Spin Columns, Thermo Scientific cat#85165) to increase the number of proteins detected. Fibrinogen, α1-acid glycoprotein, α1-antitrypsin, haptoglobin, α2-macroglobulin, IgA, albumin, IgG, apolipoprotein A-I, IgM, apolipoprotein A-II and transferrin were depleted. This was necessary because a data-dependent acquisition mode was used in which highly abundant, potentially non-informative peptides suppress detection of less abundant peptides.

Plasma samples were trypsin digested. The digest was subject to LC-MS-MS on the timsTOF Pro instrument (Brucker) using parallel accumulation serial fragmentation, and 15-minute gradients. A pooled sample comprising all plasma samples was run after every eight samples as a QC step to identify instrument drift, which can be accounted for during raw data processing.

4.3.3.2 LC-MS/MS on plasma from infants at 12 months of age

Key differences compared with the 4-month cohort samples: plasma samples were not subject to top-12 protein depletion because LC-MS/MS was run in data-independent acquisition mode (DIA). In DIA, all peptides are subject to

identification, so abundant peptides do not hinder the detection of less abundant ones. Thirty-minute gradients were used. The pooled sample was fractionated (via centrifugation), with each fraction run separately through LC-MS/MS to aid resolution of peptide identification. This information can infer the identity of ambiguously identified peptides in the other plasma samples using a "matching by spectra" approach.

4.3.3.3 Data processing

4.3.3.3.1 Raw data processing

Raw mass spectrometry data from the 4-month cohort was processed using MaxQuant software in label-free quantification (LFQ) mode and its Andromeda search engine[390]. Raw mass spectrometry data from the 12-month cohort was processed using DIA-NN software[391].

4.3.3.3.2 Data QC

Peptide group intensities were received from the TDI. Peptide groups flagged as "identified by site", "reverse mapping peptides", and "contaminants" were removed (except for hepatocyte growth factor).

4.3.3.3.3 Noise filtering

Peptide groups present in at least 50% of samples in at least one timepoint were kept.

4.3.3.3.4 Imputation

Missing values are common in LC-MS/MS data and are usually imputed. They can occur because a peptide group was under the limit of detection; such instances are termed "missing not at random" (MNAR) or because a peptide group was present but not identified, this is known as "missing completely at random" (MCAR). MCAR values occur when a peptide is not selected for identification because it was outcompeted by more abundant peptides simultaneously entering the MS machine or because its MS2 spectrum was ambiguous.

It is impossible to know whether a missing value is MNAR or MCAR. A heuristic approach must be employed so the value can be imputed appropriately. Missing values of protein groups present in fewer than two-thirds of samples were presumed MNAR and imputed as the lowest intensity value detected for that protein group. Missing values of protein groups present in more than two-thirds of samples were presumed MCAR and imputed using a k-nearest neighbour approach.

Outlying samples were identified using PCA and box plots of the $\log_2$ LFQ intensities and removed from the analysis.

4.3.3.3.5 Differential abundance analysis

The LFQ intensity data from the 4-month-old cohort is inherently normalised by the LFQ approach. Data from the 12-month-old cohort were normalised using median-centred normalisation.

The number of protein groups detected and their intensities can vary markedly from sample to sample due to technical factors, e.g. efficiency of top-12 protein depletion and trypsin digestion. Samples missing highly abundant protein groups (defined as protein groups present in 90% of samples) are likely to be of lower quality. The number of highly abundant protein groups present in each sample was used as a sample quality metric and included in the differential abundance analysis.

Differential abundance analysis was done as per section 2.2.5.1. Participant number was modelled as a random effect; timepoint was modelled as a fixed effect and sample quality metric was used as a continuous variable[321]. Results are reported using the protein name that best summarises the protein group.

4.3.4 Luminex

Luminex is a targeted, sensitive method suitable for low-abundance plasma proteins like cytokines. It allows multiplex measurement of up to 30 proteins from 25μl of plasma, making it economical in terms of time and sample volume.

4.3.4.1 Selection of proteins for study

Proteins were selected for measurement based on EUCLIDS mRNA-seq results, an analogous mouse vaccine study, and biological knowledge of cytokines and chemokines important in innate immune responses. One panel assayed low abundant cytokines and chemokines (Merck Millipore, catalogue number

HCYTOMAG-60K-24C). Another panel measured CRP, L-selectin and haptoglobin (Merck Millipore, catalogue number HCVD3MAG-67K-03). Supplementary Table 2 lists measured cytokines.

4.3.4.2 Sample processing

Platelet-poor plasma samples were processed per kit protocol, except plasma-bead incubation was done overnight at 4°C for 12-month-old cohort samples. Samples were run in duplicate and randomly allocated to wells across two plates. Paired samples were run within the same plate.

4.3.4.3 Data processing

4.3.4.3.1 Data QC

Proteins not present in at least 50% of samples within one or more time points were filtered out. Measurements below the limit of detection for a protein were offset to half the lowest detected concentration for that protein.

4.3.4.3.2 Differential abundance analysis

Concentrations were $\log_2$ transformed. A t-test (paired where possible) tested for differential abundance. Results with an FDR below 0.05 were deemed significant.

4.3.5 Clinical CRP measurement

Plasma was sent to the John Radcliffe Hospital, Oxford, UK, for measurement of CRP via clinical-grade quantification.

4.3.6 Pathway enrichment

ORA analysis was done via the STRING database website (https://string-db.org/) to identify over or under-represented pathways among DE proteins. An ORA approach was used because the number of proteins measured was too small for GSEA. ORA also allowed Luminex and mass spectrometry results to be combined. Results with an FDR of below 0.05 were deemed significant at 4 months and validated as DE at 12 months below a raw p-value of 0.05.

4.4 Results

4.4.1 Sample demographics

Paired plasma samples were available in 21/40 and 20/40 participants at 4 and 12-month-old, respectively. Appendix Table 2 contains sample demographics.

4.4.2 Mass spectrometry measurement of plasma proteins prior to and 4 and 24 hours after vaccination

In the 4-month cohort, 166 proteins were retained in the dataset after QC and filtering. Five samples were excluded due to being outliers (Appendix Figure 21 and Appendix Figure 22 contain QC plots).

4.4.2.1.1 Proteins are differentially expressed 24 but not 4 hours post-vaccination, and this is more marked when 4CMenB is co-administered

In the 4-month cohort, PCA shows global differences between pre- and post-vaccination samples (Figure 34), best seen in PC2, where there is a progression from baseline to 4 to 24 hours. The same pattern was seen in PC1 of the 12-month data (Figure 34 and Figure 36A). No proteins were DE between baseline and 4 hours in either cohort. Twenty-three proteins were DE 24 hours after vaccination in the 4-month cohort (Figure 34B, Figure 35, Supplementary Table 3), of which 9 were validated in the 12-month cohort (Figure 36B, Figure 37, Supplementary Table 4)

Samples were hierarchically clustered by sample expression of proteins that had been DE at baseline versus 24 hours in the 4 and 12-month-old cohorts. In both cohorts, the samples separated into two clusters; cluster 1 (left-hand cluster) was majority composed of 24-hour samples, while cluster 2 (right-hand cluster) contained mostly baseline vaccine samples (Figure 38).

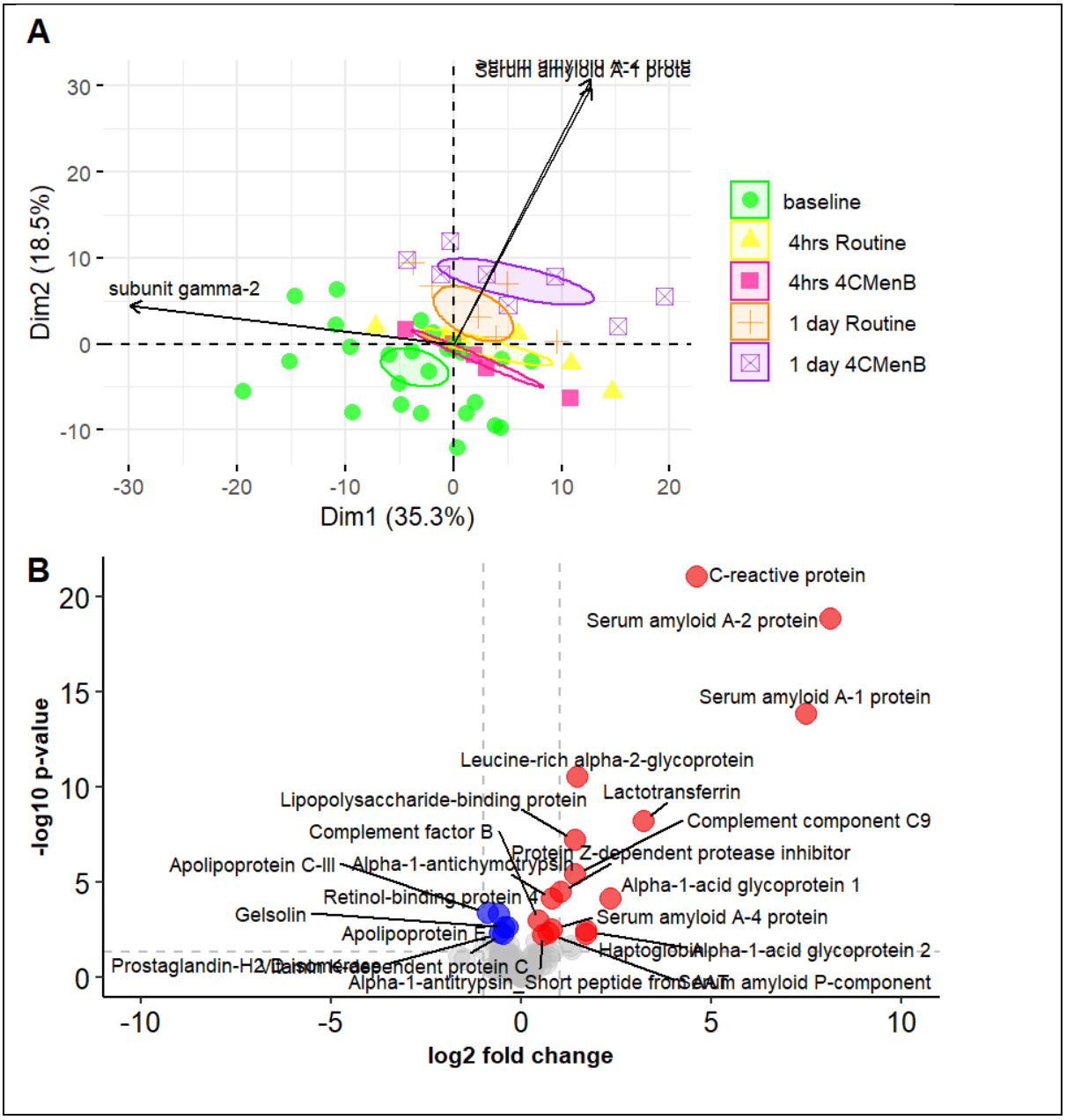


Figure 34 Influence of vaccination on protein abundances

A. *PCA of mass spectrometry protein abundance data (adjusted for participant ID). Each dot represents a sample. The centroids of the groups are indicated with prominent symbols, and the ellipses represent 95% confidence intervals of the centroids of each cluster. The two proteins with the greatest contribution to the PCA components are plotted.*
B. *Volcano plot showing results of differential expression analysis when both groups are analysed together. Red= upregulated at fdr<0.05, blue = downregulated at fdr<0.05*

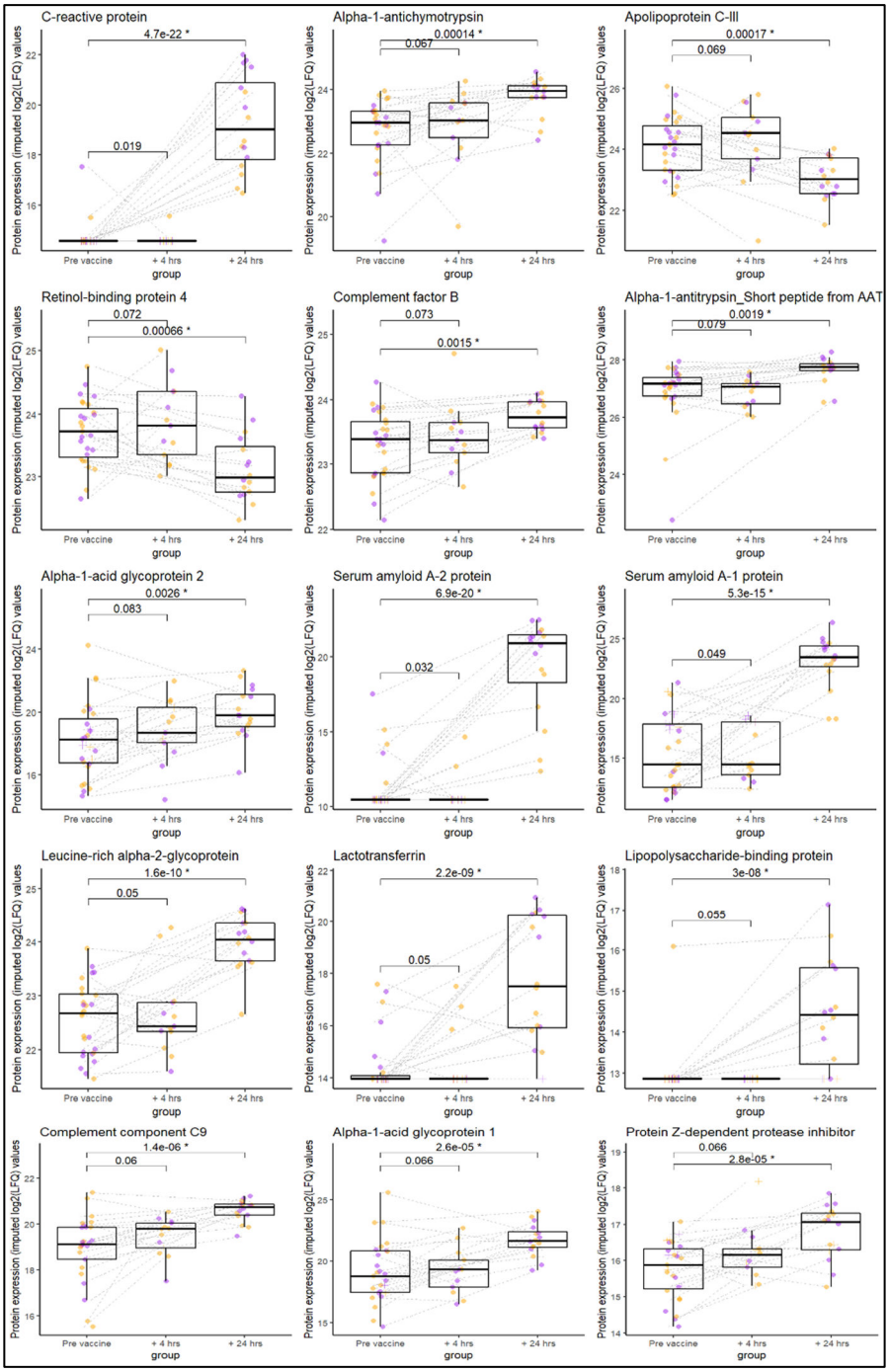


Figure 35 Dot plot showing the expression of DE proteins 24 hours after vaccination compared with baseline when both vaccine groups were combined. Raw p-values are shown. Starred p-value = FDR < 0.05. Purple= 4CMenB group, orange = routine vaccine group. Circles = measured intensity, crosses= imputed values

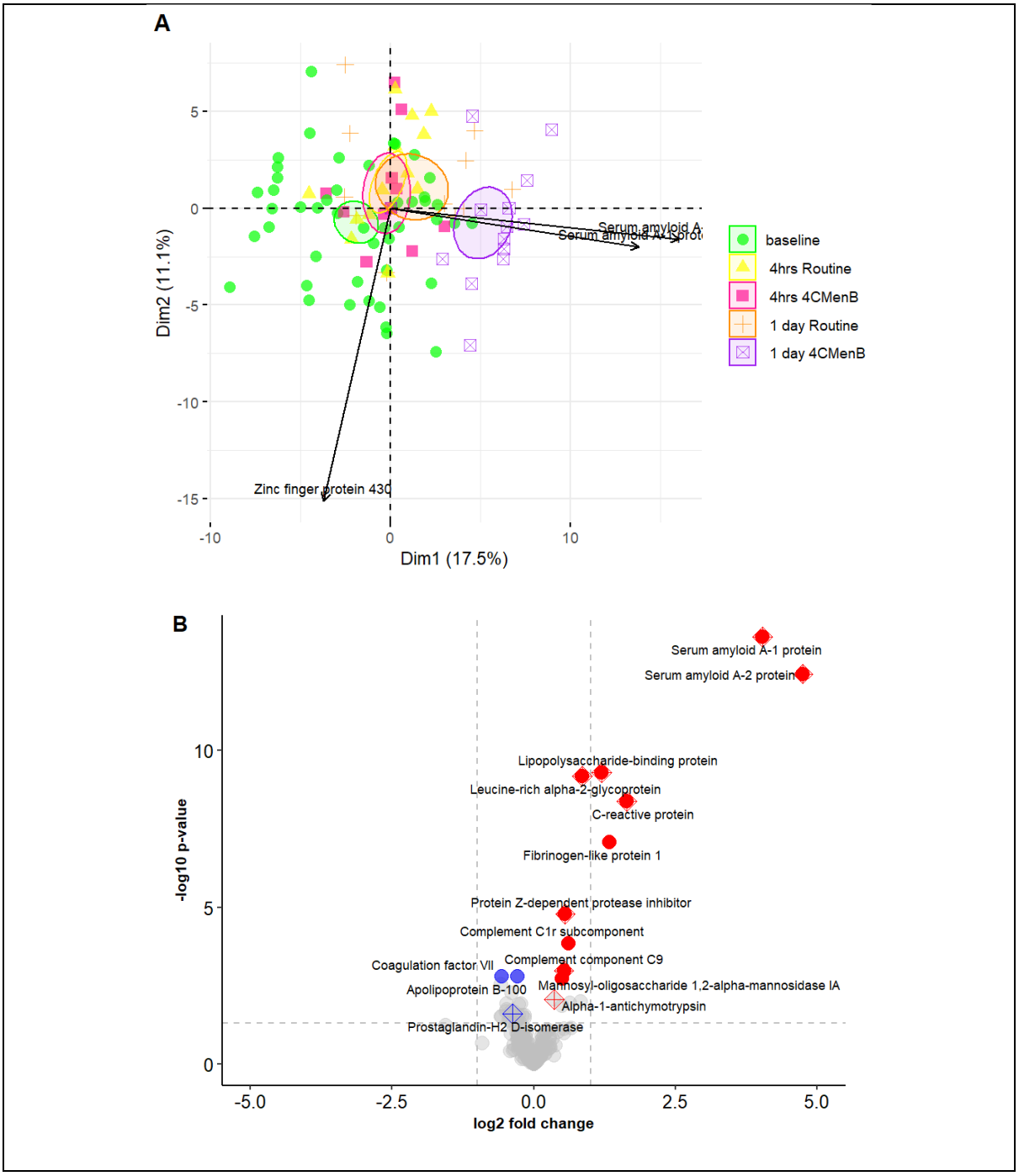


Figure 36 Influence of vaccination on protein abundance in 12-month-old infants
A. PCA of mass spectrometry protein expression data (adjusted for participant ID). Each dot represents a sample. The centroids of the groups are indicated with prominent symbols, and the ellipses represent 95% confidence intervals of the centroids of each cluster. The two proteins with the greatest contribution to the PCA components are plotted.
B. Volcano plots showing results of differential abundance analysis when both groups are analysed together. Filled dots: FDR< 0.05 at 12 months, Cross squares: validated DE proteins, i.e., FDR< 0.05 at 4 months and p<0.05 at 12 months.

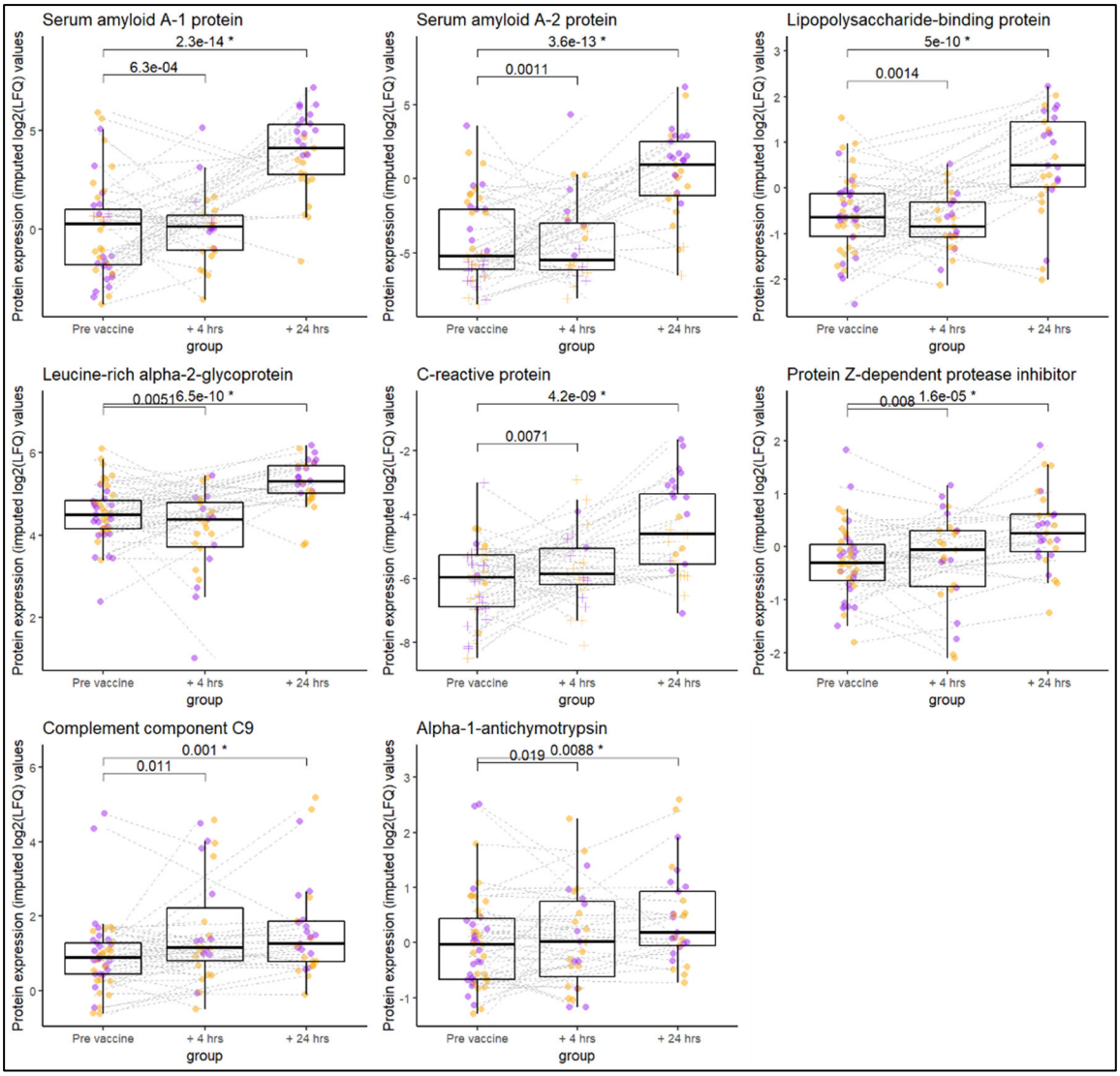


Figure 37 Dot plots of proteins validated as DE expressed after vaccination in the 12-month cohort. Raw p-values are shown. Starred p-value = post-vaccine timepoint that validated. Purple= 4CMenB group, orange = routine vaccine group. Circles = measured intensity, crosses= imputed values

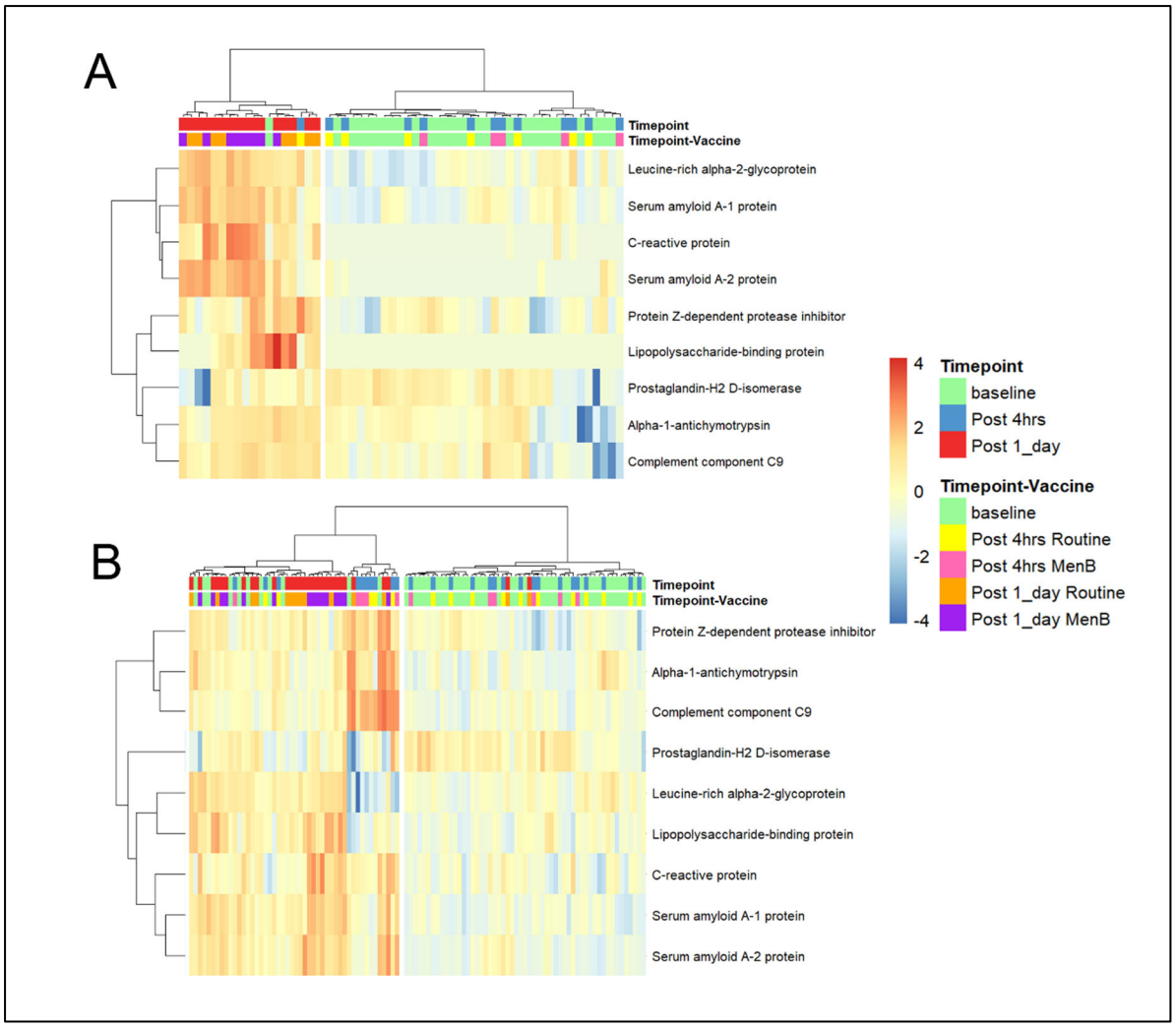


Figure 38 heatmap of proteins validated as DE 24 hours after vaccination. Samples from 4-month-old infants (A) and 12-month-old infants(B) are shown. The dendrogram represents the results of the unsupervised hierarchical clustering analysis with the tree cut at the height of 2. All samples taken 24 hours after vaccination from the 4CMenB group cluster on the left.

4.4.2.1.2 Concomitant administration of 4CMenB influences plasma protein abundance

Twenty-four hours post-vaccination samples from the routine group cluster separately from the 4CMenB group in PC2 in the 4-month cohort (Figure 34B) and PC1 in the 12-month cohort (Figure 36B). At 4 months, more proteins were DE in the 4CMenB group compared with the routine group (15 vs 8 proteins,

respectively, Supplementary Table 3, Figure 39) despite the groups' equal size and absolute $\log_2$ fold changes were greater in the 4CMenB group (Wilcoxon rank-sum test, p=0.00022). These findings were recapitulated in the 12-month cohort (Figure 41).

At 4 months, three proteins (Apolipoprotein C-I, Apolipoprotein A-I, and Apolipoprotein A-III) and 2 proteins (CRP and Serum amyloid A-2 protein) were more highly expressed in the 4CMenB compared with the routine group at four and 24 hours after vaccination, respectively (Supplementary Table 3, Figure 40). At 12 months, no proteins were validated as DE between vaccine groups at 4 hours; however, CRP and Serum amyloid A-2 protein were again higher in the 4CMenB group compared with the routine group 24 hours after vaccination Supplementary Table 4, Figure 41, Figure 39).

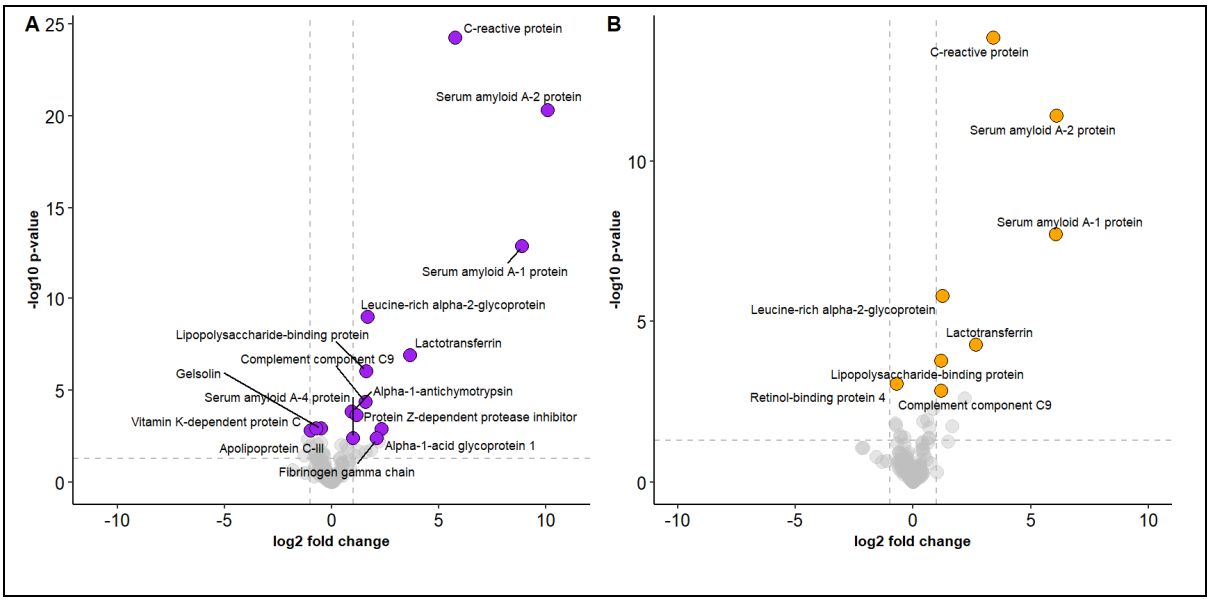


Figure 39 Volcano plots showing results of the 4 month cohort differential abundance analysis when both groups are analysed separately. A: 4CmenB + routine group. B: routine group. Labelled points with black outlines had FDR< 0.05.

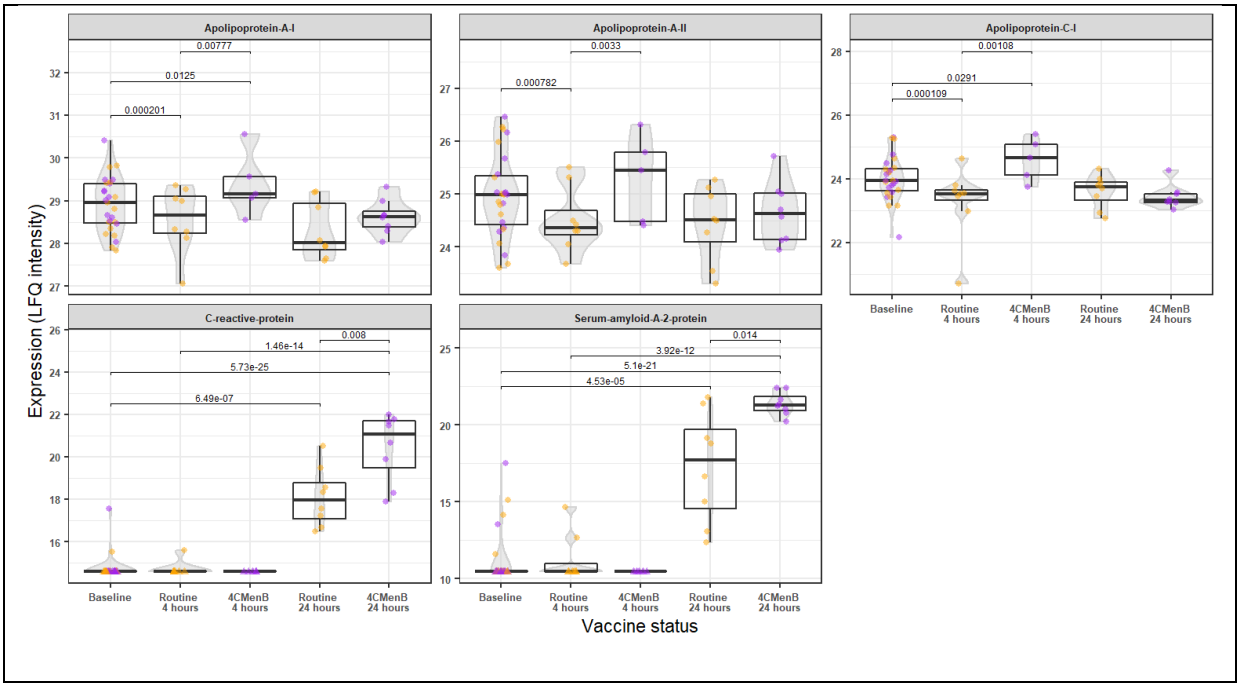


Figure 40 Proteins showed differential abundance between vaccine groups at 4 hours (upper plots) and 24 hours (lower plots) after multiple testing adjustments. Raw p-values are presented in the 4 month cohort. Only p-values under 0.05 are included in the plot.

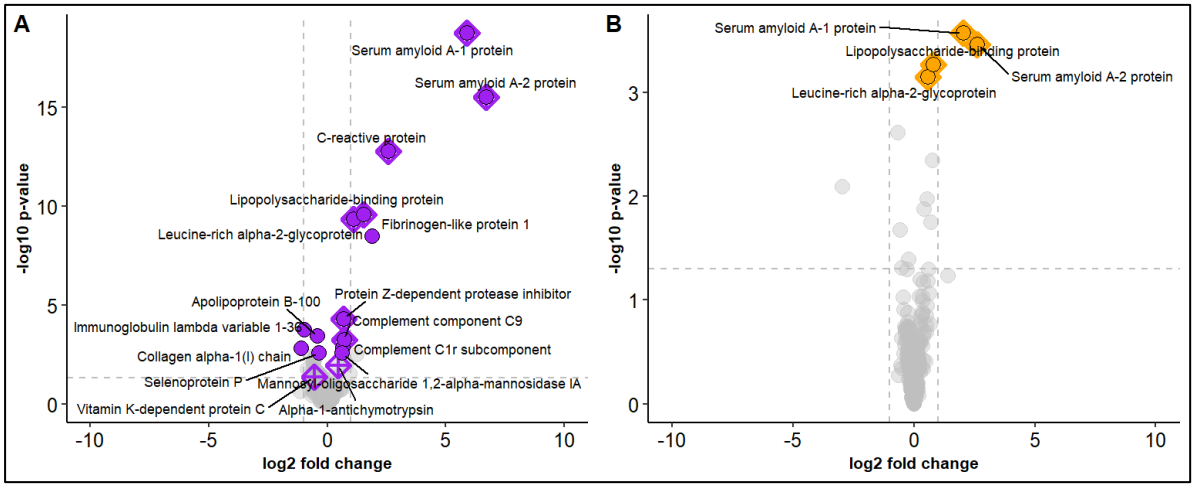


Figure 41 Volcano plots showing results of differential abundance analysis when both groups are analysed separately in the 12 month cohort. (A) Results of the 4CmenB + routine vaccinations group. (B) results of routine vaccine group. Filled points: FDR< 0.05 at 12 months, Cross squares: validated DE proteins, i.e. FDR< 0.05 at 4 months and p<0.05 at 12 months

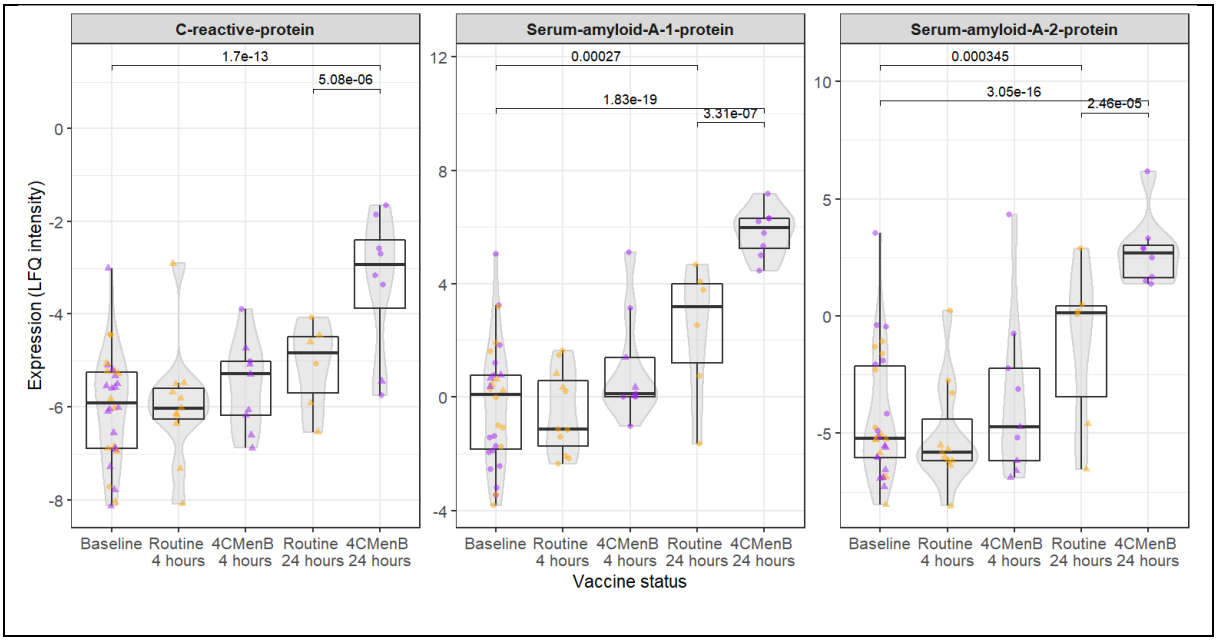


Figure 42 Proteins DE between vaccine group in the 12-month-old cohort. The numbers represent raw p-values. Samples from infants that received concomitant 4CMenB = purple. Samples from infants that received routine vaccines only = orange. P-values < 0.05 are shown.

4.4.2.1.3 Plasma proteins as correlates of vaccine reactogenicity

In the 4-month cohort, post-vaccination levels of six proteins (CRP, serum amyloid A1, serum amyloid A2, gelsolin and leucine-rich alpha-2 glycoprotein) correlated with neutrophil count (CIBERSORT abundances); however, only the correlation with serum amyloid A2 persisted after multiple testing correction (Figure 43). Gelsolin significantly negatively correlated with maximal fever post-vaccination. However, none of these results were validated in the 12-month-old cohort.

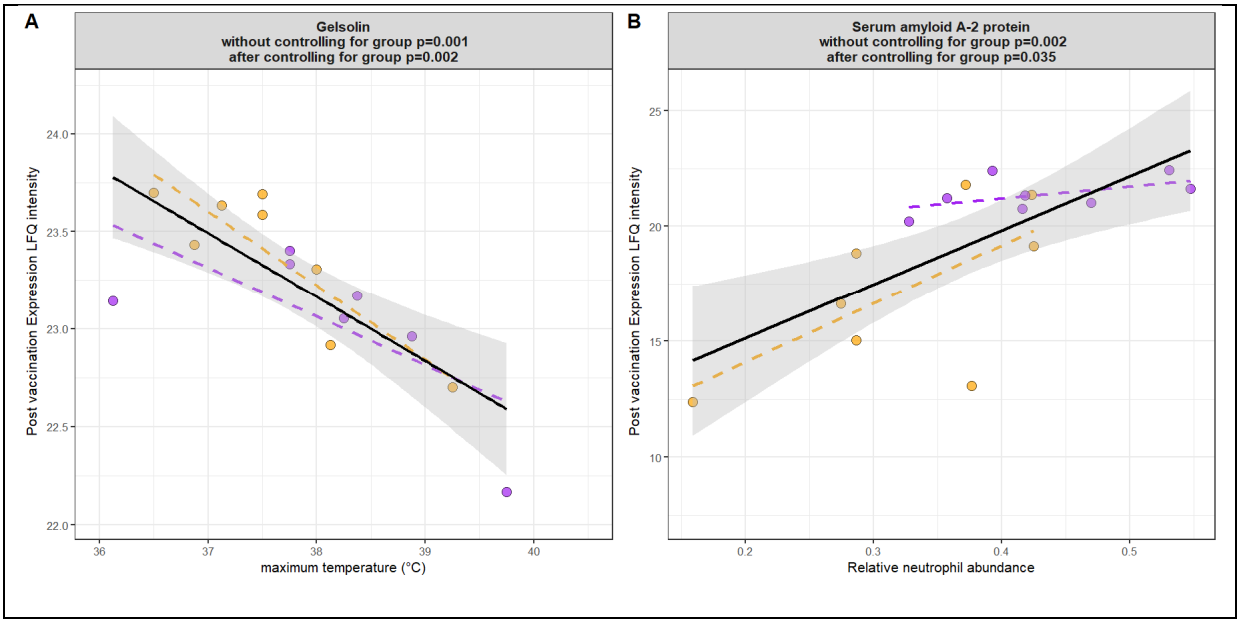


Figure 43 Relationship between plasma protein abundance 24 hours post-vaccination and markers of vaccine reactogenicity. Correlation between gelsolin (A) post-vaccination expression and an infant's maximum temperature in the first 24 hours after vaccination. Correlation between expression of serum amyloid A2 at 24 hours post-vaccination (B) and post-vaccination neutrophil abundance. Neutrophil count derived by CIBERSORTx algorithm from whole blood RNA sequencing. The p-values represent the statistical significance of the relationships between post-vaccination protein expression and marker of reactogenicity with and without controlling for the vaccine group using a simple linear model.

4.4.3 Luminex assay of plasma proteins prior to and 4 and 24 hours after vaccination

All 26 analytes were adequately detected and included in the analysis. Only cytokines DE at 4 months were evaluated at 12 months; however, the plate containing CRP, Haptoglobin and L-selection failed due to high background (the company ascribed this to a batch issue with a kit component). CRP was therefore measured using a clinical assay at the John Radcliffe hospital—there was an excellent correlation between Luminex and lab CRP at 4 months (Supplementary figure 3).

4.4.3.1.1 Vaccination induces changes in plasma cytokines and chemokine concentrations at 4 and 24 hours post-vaccination, and this is more marked when 4CMenB is co-administered.

At 4 months of age, samples cluster by timepoint in PC2 of the PCA, progressing from baseline to 4 hours to 24 hours (Figure 44A). 24-hour samples from the 4CMenB group lie at the end of that spectrum (Figure 44B). A similar separation was seen in the 12-month-old data (Figure 46A).

At 4 months of age, the 4-hour samples from the 4CMenB group cluster separately from the baseline samples in dimensions 3 and 4, and this is primarily driven by the levels of GRO (GRO-alpha). This separation by vaccine group was not evident at 12 months of age in the top 4 dimensions (Figure 46A B&C), suggesting an ameliorated effect size.

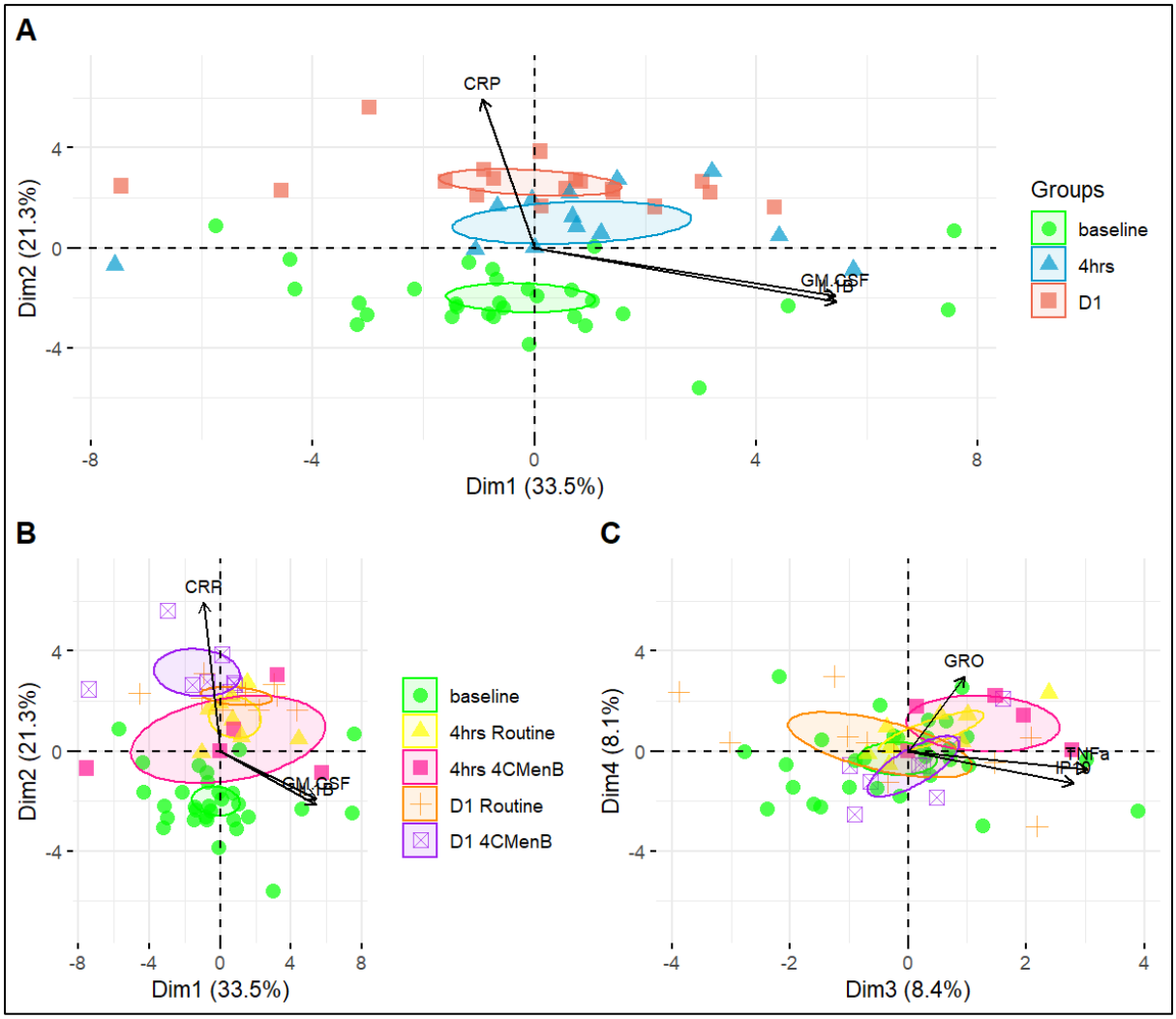


Figure 44 Influence of vaccination on cytokine, chemokine and acute-phase proteins as measured by Luminex.
PCA of Luminex expression data (adjusted for participant ID) coloured by timepoint (A) and timepoint-vaccine group (B, C) for PC components 1,2 (A, B) and 3,4 (C). Each dot represents a sample. The two proteins with the greatest contribution to the PCA components presented are plotted in green. The centroids of the groups are indicated with prominent symbols, and the ellipses represent 95% confidence intervals of the centroids of each cluster.

GRO (GRO-alpha) was the only cytokine DE 4 hours after vaccination at 4 months of age. It was not differentially expressed at 12 months of age. Eight cytokines (CRP, G-CSF, Haptoglobin, IL1-RA, IL5, IL6, L-selectin and TGFa) were DE 24 hours after vaccination (Figure 45, starred p-values) in the 4-month

cohort. Of these 8, 6 were successfully measured in the 12-month cohort (haptoglobin and L-selectin failed QC), and all were validated as DE.

CRP, G-CSF, IL-1 RA and IL6 levels were significantly higher in the 4CMenB group 24 hours after vaccination compared with the routine group at 4 months of age, but this was only true for CRP at 12 months of age.

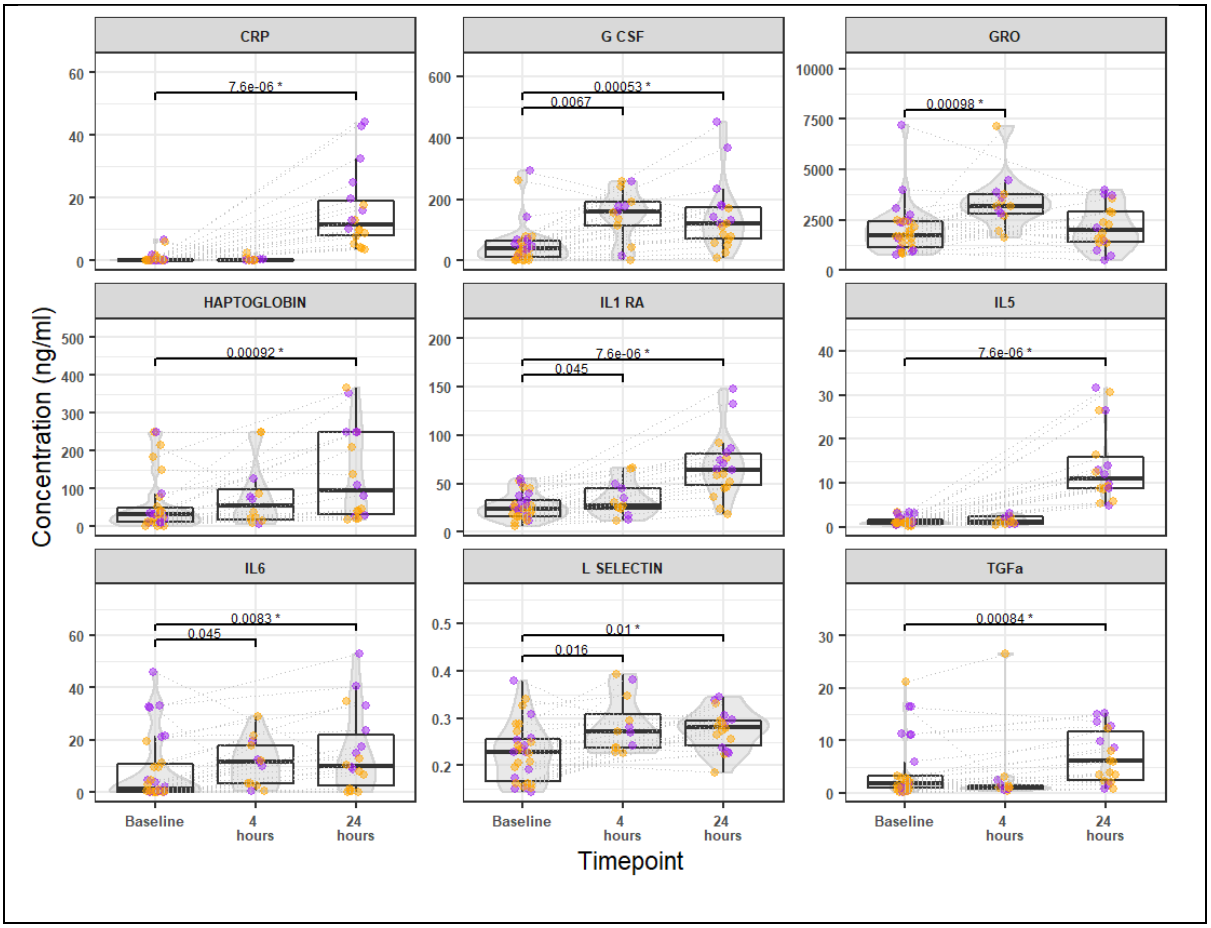


*Figure 45 Dot plots of DE proteins in 4-month-old infants. Numbers represent raw p-values. Samples from infants that received concomitant 4CMenB = purple. Samples from infants that received routine vaccine = orange. * denotes fdr < 0.05 Only p-values < 0.05 are shown.*

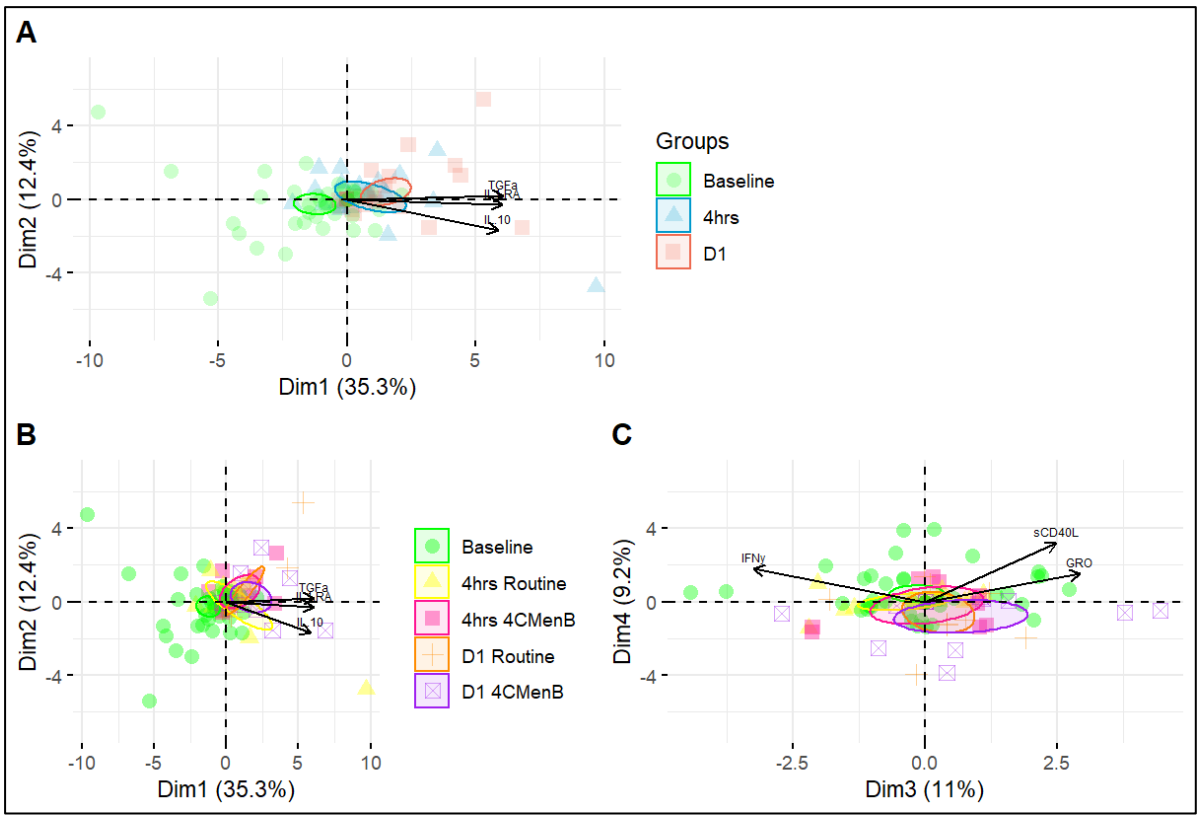


Figure 46 Influence of vaccination on cytokine, chemokine and acute-phase proteins as measured by Luminex in 12-month-old infants.
PCA of Luminex expression data (adjusted for participant ID) coloured by timepoint (A) and timepoint-vaccine group (B, C) for PC components 1,2 (A, B) and 3,4 (C). Each dot represents a sample. The two proteins with the greatest contribution to the PCA components presented are plotted. The centroids of the groups are indicated with prominent symbols, and the ellipses represent 95% confidence intervals of the centroids of each cluster.

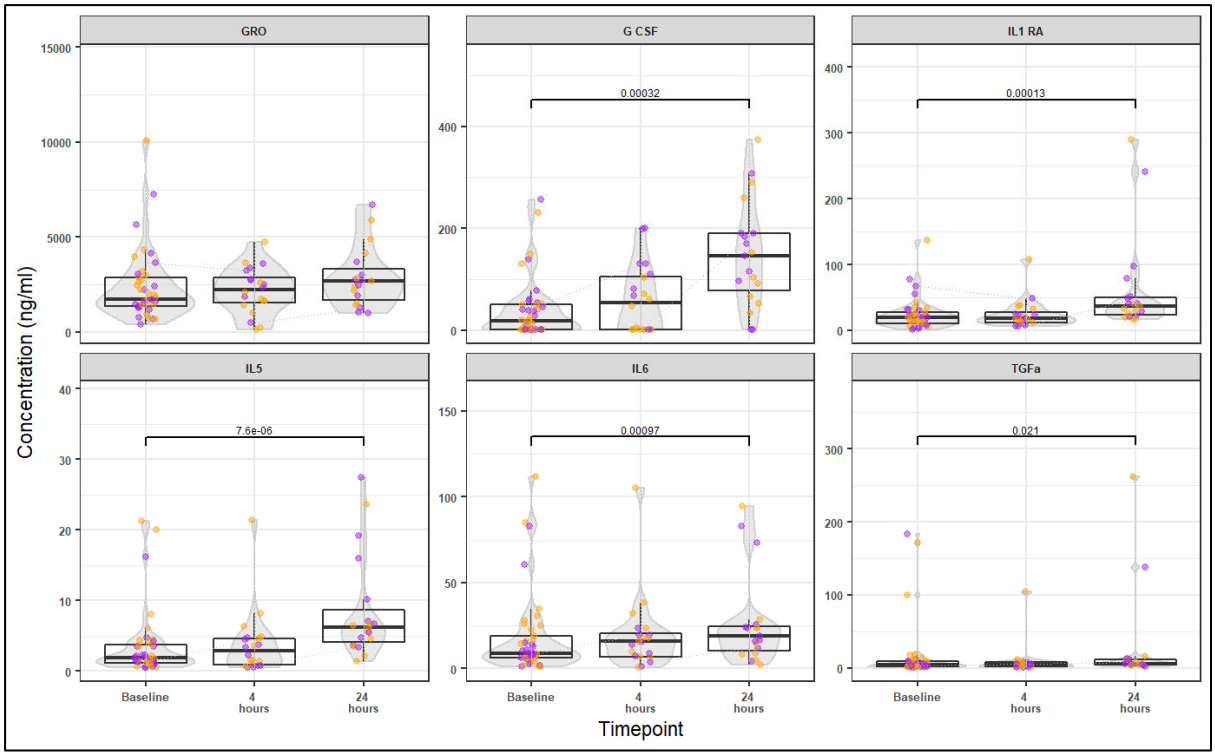


*Figure 47 Dot plots of DE proteins in the 12-month-old infants as measured by Luminex. Numbers represent raw p-values. Samples from infants that received concomitant 4CMenB = purple. Samples from infants that received routine vaccines only = orange. * denotes fdr < 0.05. Only p-values < 0.05 are shown.*

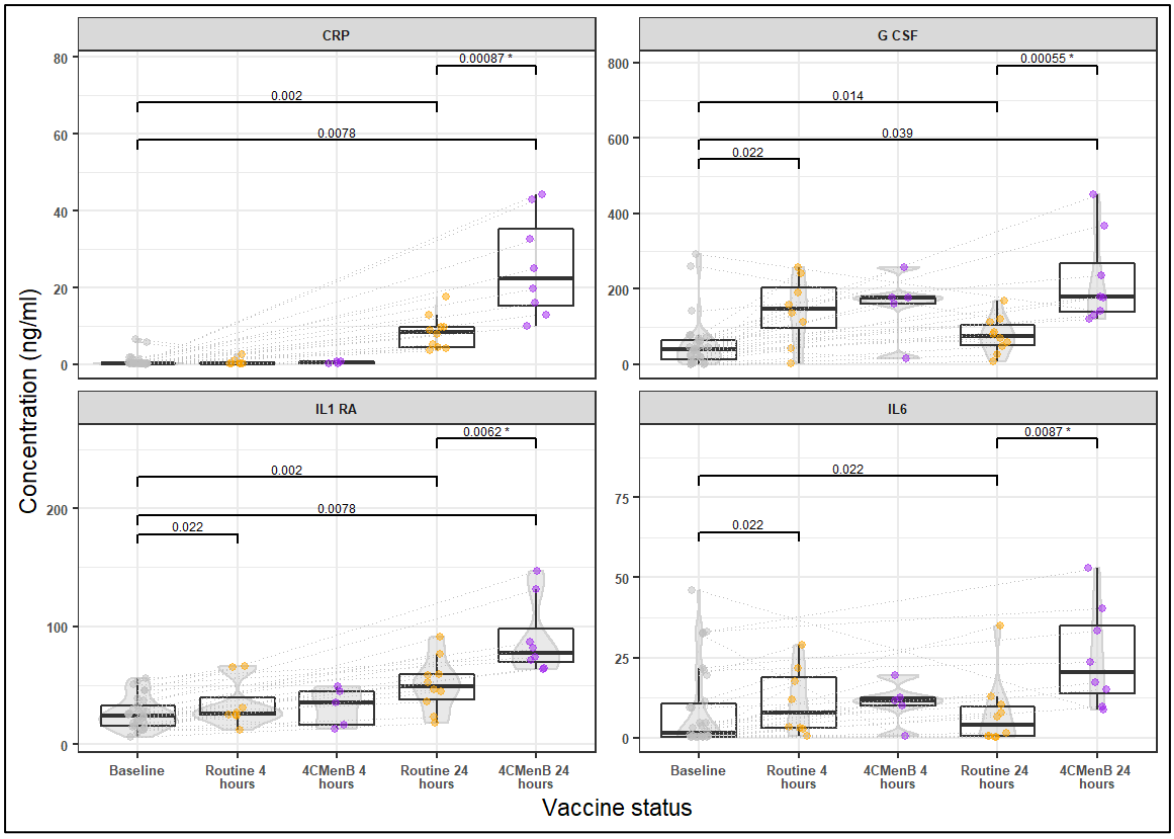


*Figure 48 Dot plots of DE proteins between vaccine groups in 4-month-old infants. Samples from infants that received concomitant 4CMenB = purple. Samples from infants that received routine vaccine = orange. Numbers represent raw p-values. Only p-values < 0.05 are shown. * denotes fdr < 0.05*

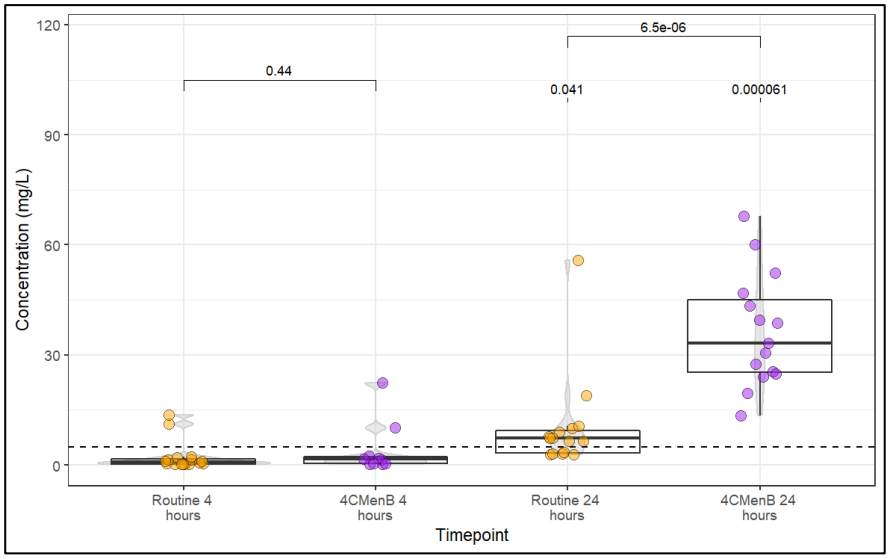


*Figure 49 12 Dot plots of lab-measured plasma CRP levels between vaccine groups in 12-month-old infants. Samples from infants that received concomitant 4CMenB = purple. Samples from infants that received routine vaccines only = orange. The normal level for CRP is represented by the dotted horizontal line. Numbers represent raw p-values. Only p-values < 0.05 are shown. * denotes fdr < 0.05*

4.4.3.1.2 Plasma CRP and IL1RA correlate with neutrophil abundance

Cytokines DE 24 hours after vaccination were investigated for association with markers of reactogenicity. Neutrophil abundance was associated with CRP and IL1RA in the 4-month but not 12-month cohort. No proteins correlated with fever in either cohort.

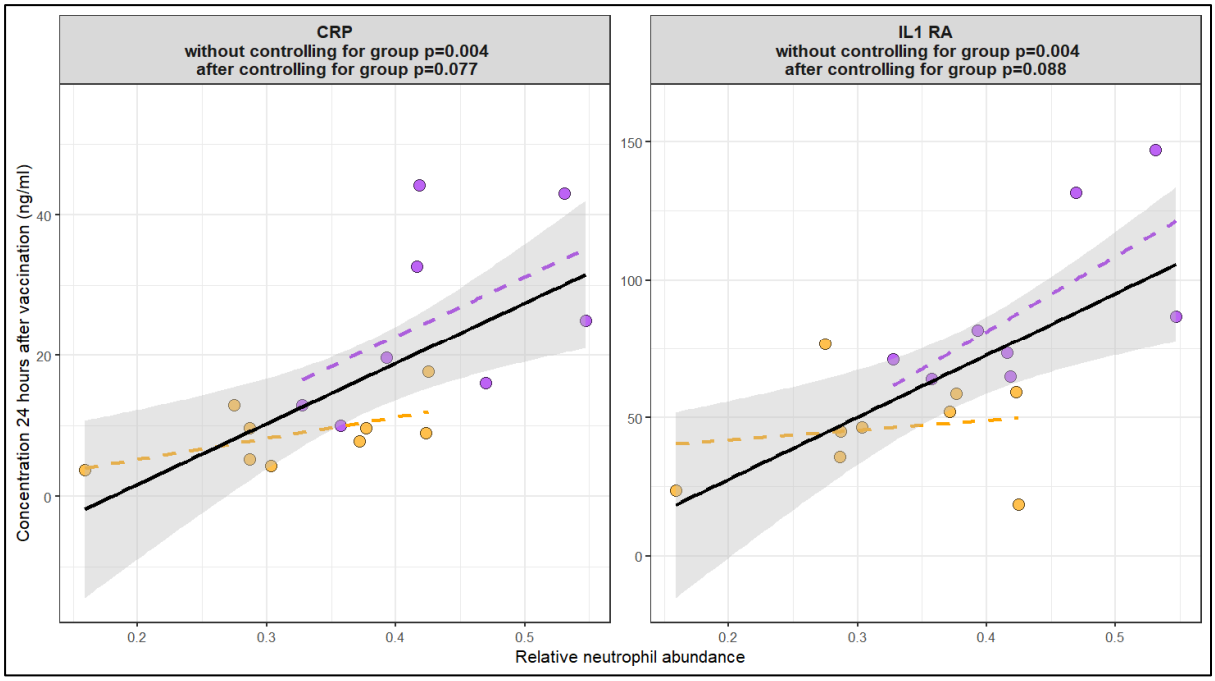


Figure 50 Correlation between plasma protein levels 24 hours post-vaccination and neutrophil abundance in 4-month-old infants. Neutrophil count derived by CIBERSORTx algorithm applied to the whole blood RNA sequencing data. The p-values represent the statistical significance of the relationships between post-vaccination protein abundance and marker of reactogenicity with and without controlling for vaccine group determined via a linear model.

4.4.4 Functions of differentially expressed proteins after vaccination

Details on the class, functions and cellular producers of DE proteins are listed in Supplementary Table 5. Classes of DE proteins include serum amyloid proteins, apoproteins, pentraxins, serine proteases, lipocalins, and complement.

Most DE proteins in the mass spectrometry analysis are acute-phase proteins made by the liver (Supplementary Table 5). Many DE proteins interact with one another (Figure 51) and collectively regulate or participate in the complement cascade, coagulation, inhibition of proteolytic enzymes, iron homeostasis, transport of small molecules, binding to pathogen-associated molecular patterns,

vitamin A metabolism and lipid transport and metabolism (Table 18); this was true in both cohorts except for homeostasis and vitamin A metaboliswhich was limited to the 4-month cohort (Table 18).

Over-representation analysis (ORA) of proteins DE at 4 months using the STRING database revealed significant enrichment of the pathways involving the aforementioned processes except for iron homeostasis. Additional enriched terms relate to the acute-phase response, antibody responses, cytokine production and signalling, vesicle transport, and the immune response.

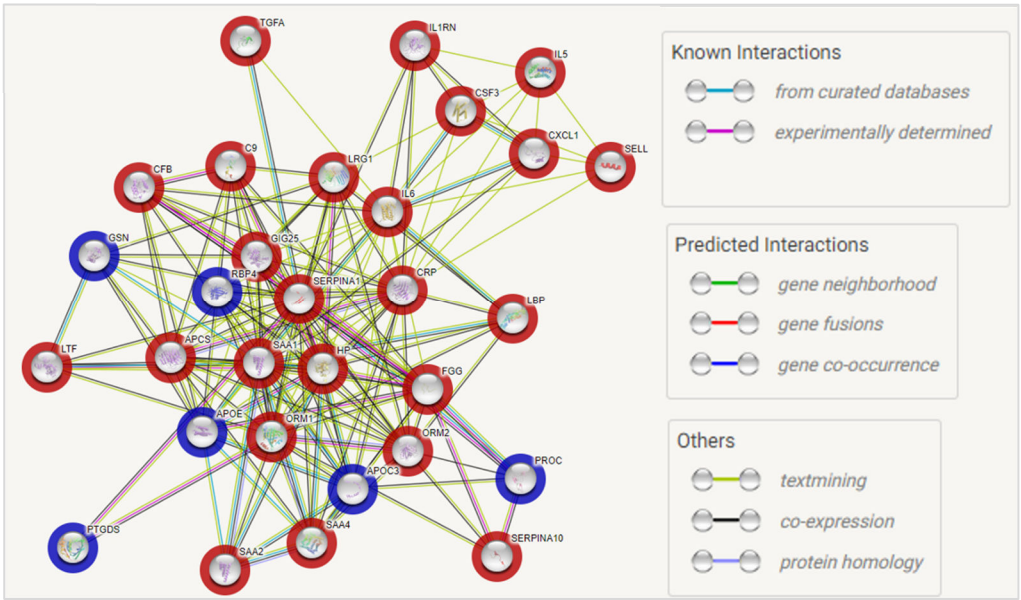
Table 18 The key processes that DE proteins play a role in. For more detailed information, see Supplementary Table 5. Red = proteins upregulated, blue = proteins downregulated

Process	Proteins (* = validated as DE in the 12-month cohort)	Terms relating to this process enriched in the over-representation analysis?
complement cascade	C9* Complement factor B	Yes
coagulation	Protein Z-dependent protease inhibitor*, Fibrinogen gamma chain, Vitamin K-dependent protein C* Prostaglandin-H2 D-isomerase	Yes
Protection from proteolytic enzymes released by leukocytes	Alpha-1-antichymotrypsin*, Alpha-1-antitrypsin	No
Iron homeostasis	lactotransferrin haptoglobin	No
Transport of small molecules	Alpha-1-acid glycoprotein 1 Alpha-1-acid glycoprotein 2	Yes

Table 16 (continued)

Lipid transport and metabolism	Serum amyloid A-2* Serum amyloid A-1* Apolipoprotein A-I Apolipoprotein A-II Apolipoprotein C-I apolipoprotein C-III Apolipoprotein E	Yes
Binding to pathogen-associated molecular patterns	CRP* Serum amyloid P-component lipopolysaccharide-binding protein* Serum amyloid A-2* Serum amyloid A-1* Serum amyloid A-4 Lactotransferrin Gelsolin	Yes
Vitamin A metabolism	Retinol-binding protein 4	Yes
Leukocyte differentiation, maturation and degranulation	Granulocyte colony-stimulating factor * GRO-alpha (neutrophils)* IL5* IL6* Leucine-rich alpha-2-glycoprotein* Haptoglobin (lymphocytes) Prostaglandin-H2 D-isomerase (mast cells)	Yes
Leukocyte chemotaxis and recruitment	GRO-alpha IL5* IL6*	Yes
Induction of the acute-phase response of the liver	IL6*	Yes
Inhibition of inflammatory response	IL1RA*	Yes

A


TGFA
IL1RN
IL5
CSF3
CXCL1
SELL
C9
LRG1
CFB
IL6
GSN
GIG25
RBP4
SERPINA1
CRP
LBP
LTF
APCS
SAA1
HP
FGG
APOE
ORM1
ORM2
PROC
APOC3
PTGDS
SAA4
SAA2
SERPINA10
Known Interactions
from curated databases
experimentally determined
Predicted Interactions
gene neighborhood
gene fusions
gene co-occurrence
Others
textmining
co-expression
protein homology

B

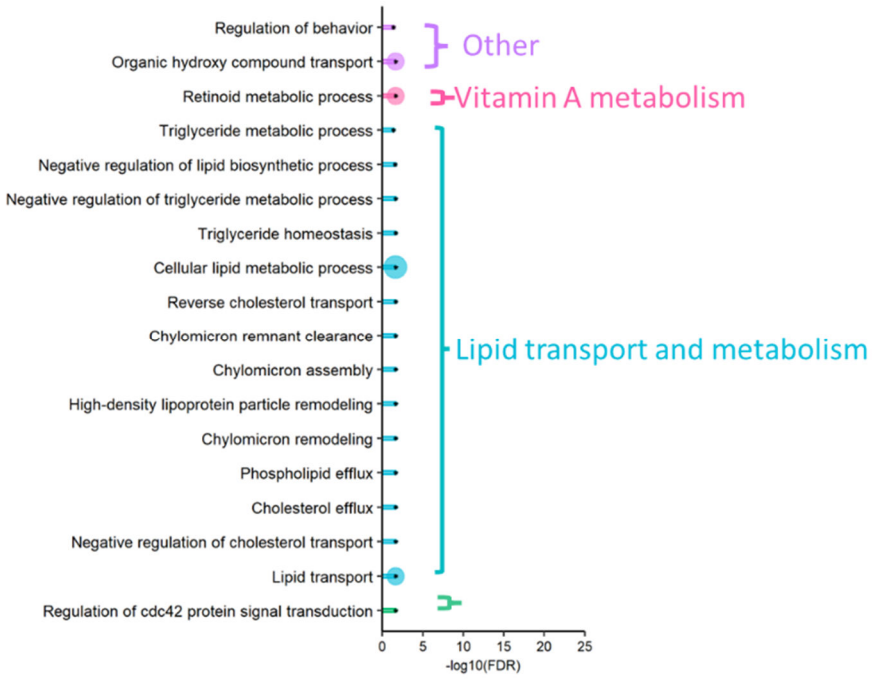

Regulation of behavior
Organic hydroxy compound transport
Retinoid metabolic process
Triglyceride metabolic process
Negative regulation of lipid biosynthetic process
Negative regulation of triglyceride metabolic process
Triglyceride homeostasis
Cellular lipid metabolic process
Reverse cholesterol transport
Chylomicron remnant clearance
Chylomicron assembly
High-density lipoprotein particle remodeling
Chylomicron remodeling
Phospholipid efflux
Cholesterol efflux
Negative regulation of cholesterol transport
Lipid transport
Regulation of cdc42 protein signal transduction
0
5
10
15
20
25
-log10(FDR)
Other
Vitamin A metabolism
Lipid transport and metabolism

Figure 51 (continued)

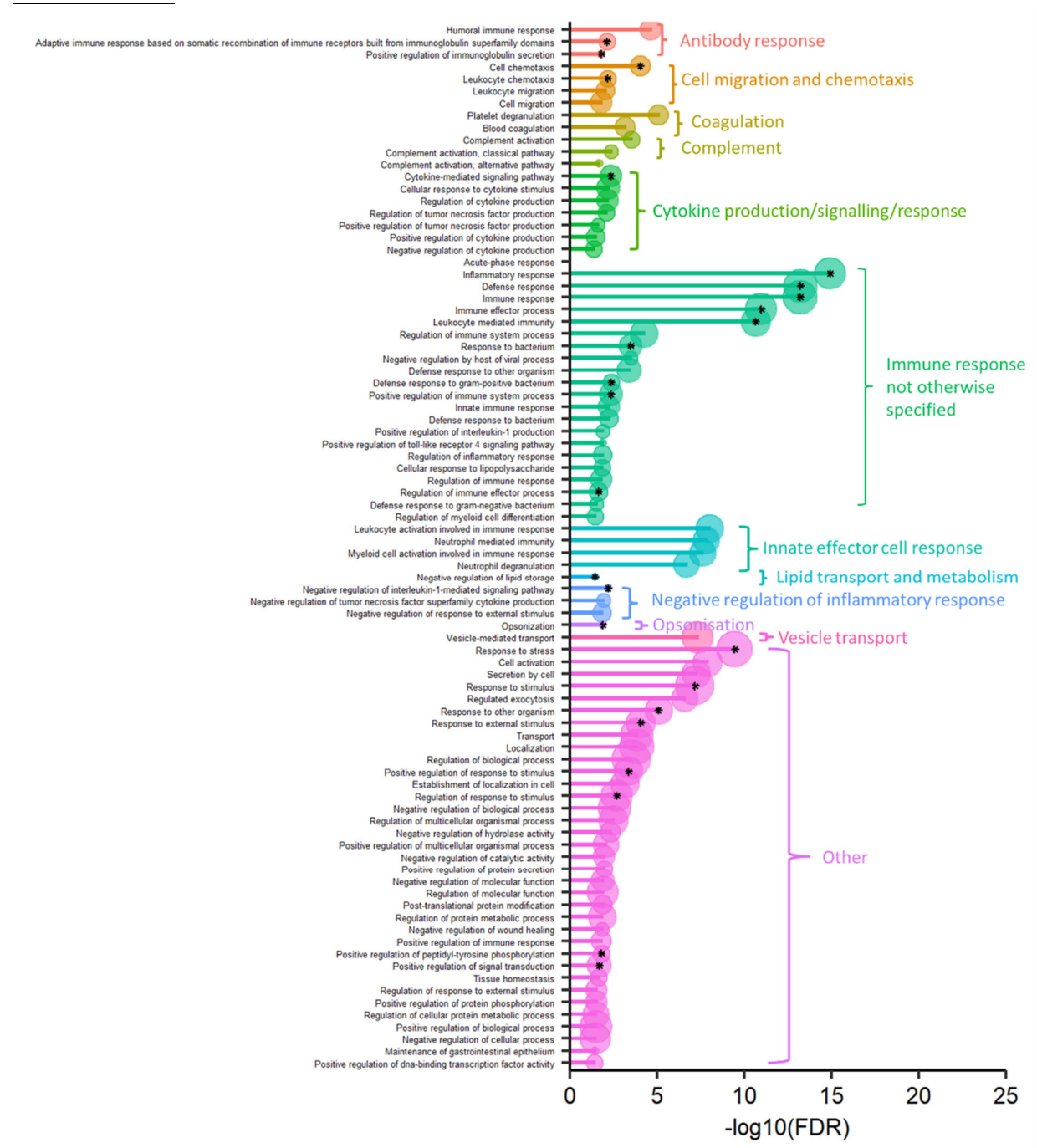


Figure 51 Protein expression enrichment results

A: Protein-protein interaction network for DE proteins between baseline and 4 or 24 hours in the 4-month-old infants derived from the STRING database. Nodes = proteins. Edges = associations. Upregulated proteins are in red, and downregulated are in blue.

*B: Results of over-representation analysis based on proteins downregulated between baseline and 24 hours in the 4-month-old infants. Results derived from the STRING database. Gene ontology biological process pathways with fdr< 0.05 are shown. **

Figure 51 caption continued
starred pathways also enriched in the 12-month-old cohort. The size of the dot reflects the number of proteins present in that pathway. 5 10 15 20

*C: Results of over-representation analysis based on proteins upregulated between baseline and 24 hours in the 4-month-old infants. Results derived from the STRING database. Gene ontology biological process pathways with fdr< 0.05 are shown. *starred pathways also enriched in the 12-month-old cohort. The size of the dot reflects the number of proteins present in that pathway.* 5 10 15 20

4.5 Discussion

4.5.1 Vaccination stimulates an acute-phase response

The mass spectrometry study confirmed that vaccination induces the hepatic acute-phase response. This is not surprising given cytokines that induce fever also induce the acute phase, e.g., IL-1B and IL-6. Previous studies have identified vaccine-associated increases in individual acute-phase proteins, e.g., CRP, but not the wider acute-phase response as captured here.

4.5.1.1 Functions of acute-phase proteins induced and repressed after vaccination

Some of the processes that vaccine-associated proteins participate in, act directly against pathogens. For example, the complement system kills and opsonises pathogens and activates and recruits leukocytes[392]. Pentraxins like CRP and Serum amyloid P-component opsonise and neutralise bacteria by binding PAMPs[393,394]. Some processes protect the host from the deleterious effects of the inflammatory process. For example, alpha-1-antichymotrypsin and alpha-1-antitrypsin inhibit proteolytic enzymes like elastase that can destroy tissue. Gelsolin scavenges free actin released by dead cells that would otherwise create a hyper-viscous environment at infection sites.

The utility of coagulation, iron homeostasis, vitamin A metabolism and lipid transport and metabolism to the immune response is less obvious and, therefore, briefly discussed below.

4.5.1.1.1 Perturbation in proteins involved in coagulation post-vaccination

Protein Z-dependent protease inhibitor (an anticoagulant) and fibrinogen gamma chain (pro-coagulant) were upregulated after vaccination, whilst vitamin K-dependent protein C (anti-coagulant) and prostaglandin-H2 D-isomerase (an inhibitor of platelet function) were downregulated. The net impact of these changes on clotting is hard to infer as three changes are pro-coagulant (higher fibrinogen gamma chain, lower vitamin K-dependent protein C and prostaglandin-H2 D-isomerase) and one change is anticoagulant (upregulation of Protein Z-dependent protease inhibitor). Several factors influencing clotting were not measured, e.g., mediators of platelet function and endothelial activation, making it hard to infer how vaccination influences coagulation in these infants overall. Clinically significant changes in clotting propensity seem unlikely as acquired clotting abnormalities are rare in children unless there is severe illness (e.g., disseminated intravascular coagulation in sepsis) or an acquired platelet disorder. Inflammation induces a prothrombotic state; therefore, perturbation in coagulation pathways post-vaccination in these infants may reflect this state and attempts to counteract it.

4.5.1.1.2 Perturbation in proteins involved in iron homeostasis post-vaccination

Lactoferrin and haptoglobin increased after vaccination in the 4-month-old cohort. Haptoglobin binds to free haemoglobin (a strong oxidiser) that would otherwise harm host cells. Lactotransferrin binds to free iron. Scavenge of free haemoglobin and iron inhibits bacterial growth by preventing bacterial access to iron.

4.5.1.1.3 Perturbation in proteins involved in retinol metabolism transport and metabolism post-vaccination

Retinol binding protein (RBP) is a negative acute-phase response protein—i.e., its expression reduces during inflammation. It transports retinol (vitamin A) from the liver to peripheral tissues. The drop in RBP is in line with a drop in vitamin A levels during acute inflammation and infection[395]. Vitamin A plays an essential role in immune homeostasis by binding to the nuclear retinoic acid receptor (RAR) in leukocytes. RAR is a transcription factor crucial to cell homing and differentiation[396]. In addition, vitamin A has anti-inflammatory properties, inducing monocytes towards the anti-inflammatory M2 phenotype and inhibiting the differentiation of naïve T-cells into pro-inflammatory Th17 cells[396397]. A reduction in serum vitamin A during infection seems to promote an inflammatory response; however, its role in the immune system is complex, contradictory at times and context-specific[396].

4.5.1.1.4 Perturbation in the regulation of lipid transport and metabolism post-vaccination

Apolipoproteins are found within lipoprotein complexes, e.g., high-density lipoprotein (HDL) and very low-density lipoprotein (VLDL), which transport triglycerides and cholesterol through the circulation. In the 4-month cohort, apolipoprotein E (ligand for VLDL uptake by the liver) and apolipoprotein C-III (inhibits hydrolysis, clearance and hepatic uptake of VLDL and LDL) were downregulated. Apolipoprotein C-I (inhibits hepatic uptake of VLDL and LDL and prevents cholesterol loading into lipoprotein particles) and serum amyloid A (apolipoprotein of the HDL complex, which associates with HDL during inflammation to accelerate HDL clearance) were upregulated. Changes in apolipoproteins were less pronounced at 12 months. Only serum amyloid was significantly upregulated again. Some apolipoproteins were uniquely DE at 12 months, e.g., apolipoprotein B100 (FDR=0.01). The benefit of changing the composition of serum lipoproteins in infection is not fully understood; however, there is evidence that lipoproteins and apolipoproteins bind LPS and lipoteichoic acid[398]. Lipids were not measured in this study, but the inferred outcome of the changes outlined above is an increase in VLDL and a decrease in HDL—something previously described during the acute-phase response[399].

A study of healthy adults receiving a quadrivalent split virion influenza vaccine (fluzone®) showed no change in HDL, LDL or cholesterol. The reactogenicity of flu vaccines is often milder, and they may be weaker inducers of the acute-phase response compared with the vaccines given in EUCLIDS; therefore, results

from the adult flu vaccine study cannot be extrapolated to EUCLIDS[400,401]. It would be useful to measure lipid levels in infants to provide a complete picture of lipid metabolism after vaccination, potentially providing mechanistic insight into circulating lipid perturbation in acute-phase responses[398]

4.5.1.1.5 Other possible functions of differentially expressed proteins

The immune functions of DE proteins could also lie outside their classical roles. For example, haptoglobin binds to lymphocytes and myeloid cells to modulate their function, and apolipoprotein E inhibits neutrophil and macrophage activation and facilitates CD1 lipid antigen presentation to natural killer T-cells.

4.5.1.2 Clinical implications of the acute-phase response post-vaccination.

CRP is frequently measured in febrile children presenting to hospital. It is an imperfect yet valued biomarker for assessing the possibility of serious bacterial infection. In this study, 4-month-old infants who received 4CMenB had a median clinical laboratory-measured CRP 24 hours after vaccination of 45mg/L (IQR 27-54mg/L) with a maximal measurement of 81 mg/L. At 12 months, levels were slightly lower, with a median of 22 mg/L (IQR 25-45mg/L) and a maximal measurement of 67 mg/L. Though CRP levels should always be interpreted in the context of clinical history and examination, a CRP of more than 50 to 60mg/L increases the clinical index of suspicion of serious bacterial infection; indeed, studies looking for novel infection biomarkers often define children as having "probable bacterial infection" at CRP >60 mg/L[402]. This study suggests CRP may

have limited utility for clinically evaluating febrile infants after vaccination; insufficient awareness of this phenomenon could increase the incidence of unnecessary invasive infection screens and antibiotics in infants presenting to the hospital with vaccine-related fever.

4.5.2 Changes in plasma cytokines and chemokines post-vaccination

Changes in circulating cytokines in this study may partially reflect local cytokine production at vaccination sites. There was a rise in pro-inflammatory cytokines after vaccination. These included chemokines like GRO (CXCL1), which is released by macrophages, neutrophils and endothelial cells and recruits neutrophils. The rise in plasma GRO 4 hours after vaccination may reflect a rapid "call to arms" by cells at the vaccination site.

IL5 is a Th2 cytokine that increased 24 hours after vaccination at 4 and 12 months. It is made by mast cells and Th2 lymphocytes. No other changes in Th2-associated cytokines occurred (IL-4 and Il-13 were unchanged, and IL-9 was not measured). Th2 responses generally favour production of antibody responses, so are relevant for the diseases (all bacterial) covered by the vaccines given in this study. There was no change in IL12 IFN-α and IFN-γ (which skews lymphocytes towards a Th1 phenotype), nor was there any change in classical Th1 cytokines IL2 and interferon-γ (TNF-beta was not measured). The paradigm of Th1 and Th2 immunity is often discussed in vaccine response but may be less relevant for the initial 24 hours after vaccination because lymphocyte responses trail the immediate responses of innate cells.

Increased plasma IL1-RA presumably reflects attempts to balance pro-inflammatory cytokine secretion. IL1-RA inhibits the pro-inflammatory cytokine IL-1B by competitively binding to the IL1R1 receptor.

4.5.2.1 Induction of acute-phase response by IL6

Vaccine induction of the acute-phase response presumably results from the rise in plasma IL6 identified in the Luminex data. Macrophages predominantly express IL6. IL6 was not detected in the whole-blood gene expression data suggesting IL6 may be released by macrophages at the injection site or associated draining lymph nodes. This hypothesis is supported in a mouse model by McKay et al., in, which LPS immunisation showed IL6 upregulation in serum at the protein level. At the transcript level, IL6 increased at the injection site but not in lymph nodes PBMCs or whole blood.

In contrast to IL6, two other cytokines, which stimulate the hepatic acute-phase response—IL1B and TNFa—were not DE after vaccination, despite being DE in the whole-blood transcriptome at 4 hours and 24 hours (IL1B only). Such discrepancies will be addressed in the next chapter when the multiple omics modalities are integrated. Technical reasons for the lack of differential expression in IL1B and TNFa protein levels could reflect timing or assay sensitivity. IL1B is potent; therefore, its production and release are tightly controlled. It was barely detectable above the background, so small yet biologically significant rises in IL1B may be difficult to detect in this system. Changes in IL1B post-vaccination have been reported in one study using the

Quanterix system[403]. In the EUCLIDS study, marginal changes cannot be definitively ruled out, especially as multiple testing adjustment raises the bar for detection of differential abundance—note TNFa was upregulated at 4 hours in 4-month-old infants at a raw p-value of 0.042, but FDR was 0.16.

4.5.3 Kinetics of protein expression

Changes in systemic circulating cytokines at 4 hours may be expected, given fever typically occurred within 6 hours; however, only GRO was DE between baseline and 4 hours. Peak time for differential abundance was 24 hours, but some cytokines may be starting to rise earlier as four of the cytokines upregulated at 24 hours (IL-1RA, L-selectin, GCSF and IL6) were upregulated at 4 hours at $p<0.05$ but did not pass multiple testing correction.

4.5.4 Validation of results across assays

Strong correlations between the clinical laboratory CRP measurements and the Luminex and mass spectrometry readouts validated the approach used to process each dataset (Supplementary figure 3). CRP, haptoglobin and L-selectin were measured by mass-spectrometry and Luminex. CRP and haptoglobin were upregulated 24 hours after vaccination in the Luminex and mass-spectrometry data sets. L-selectin was upregulated in the Luminex dataset only; this could be because only four peptides of L-selectin were identified in the mass spectrometry data, potentially limiting the quantification accuracy.

4.5.5 Comparing findings with the broader literature

No studies have used mass spectrometry to survey the acute-phase response to vaccines. Several have looked at CRP and serum amyloid post-vaccination. Serum amyloid increases after split virion influenza vaccine and BNT162b2 mRNA in adults but not to the level seen in this study[400,404]. A study of 13 infants aged 6 weeks to 10 months presenting to hospital within 48 hours of vaccination with adverse reactions (e.g., fever, seizures, excessive crying, vomiting) revealed serum CRP values ranging from 10 mg/L to 67 mg/L, which is in keeping with the results of this study[405]. The infants had received various vaccines, and no breakdown was given[405]. Similarly, post-vaccination CRP levels around 40 mg/L are common in premature infants[406].

Multiple studies show IL6 and IL-1RA rise post-vaccination[404,407,408]. GCSF, GRO-alpha, IL5, L-selectin, and TGFa have not previously been shown to rise post-vaccination, possibly because these cytokines were not tested or were under the limit of detection[104,404,407–410]. No studies have evaluated cytokine responses four hours post-vaccination.

The lack of post-vaccination associated changes in several cytokines in this study contradicts the literature, e.g., IP-10, IL-8, and IFN-γ are upregulated, and IL7 is downregulated 24 hours after BNT162b2 mRNA vaccine[404]. TNF-alpha is upregulated after adjuvanted H1N1 vaccination[104]. This discrepancy may be due to technical differences such as the platform used (most studies use mesoscale discovery, and one study used a cytometric bead array kit) and sample size (e.g., the BNT162b2 mRNA study linking reactogenicity and plasma cytokines

had 92 participants). Different vaccines and participant ages may also explain discrepant findings.

4.5.6 Changes in plasma protein levels are greater when 4CMenB is co-administered

More proteins were DE when 4CMenB was co-administered at 4 and 12 months, suggesting that 4CMenB enhances the acute-phase response. Despite there being differences between the vaccine groups, no robust correlates of reactogenicity were found at 4 and 12 months. This contrasts with the whole blood transcriptome data, which showed L-selectin expression four hours post-vaccination was associated with vaccine-related fever. This discrepancy is not necessarily surprising because upregulation in L-selectin mRNA in whole-blood may represent changes in circulating cell populations, whereas upregulation of L-selectin at the protein level represents increased shedding of L-selectin by activated leukocytes as they migrate across the endothelium.

It is perplexing that the post-vaccination levels of IL1B and IL6 and TNFα does not correlate with temperature after vaccination, given these cytokines are key pyrogenic drivers. This lack of correlation could be technical or biological. Firstly, the window for measuring these cytokines may have been missed as the median time of fever onset was 6 to 9 hours[103]. Secondly, the sensitivity and precision of Luminex may be insufficient to detect correlations above technical noise. It is also possible that fever is partly mediated by one of the other previously postulated mechanisms—e.g., from neurones transmitting messages to the vaccination site (see 1.2.1.3) or by binding of blood-borne vaccine PAMPS to

TLRs on the vessels at the blood-brain barrier. The lack of a robust correlate of reactogenicity is unsurprising as only one study has previously found serum cytokines (IP-10) to correlate with reactogenicity (as measured by a composite vaccine-related adverse event score)[411].

4.6 Conclusion

This study describes vaccine-induced changes in cytokines and the acute phase response. Changes were more marked in the 4CMenB group, potentially linking vaccine reactogenicity to plasma protein levels; however, no correlates of reactogenicity could be confirmed. Induction of the acute phase is notable, given that several miRNAs DE after vaccination are hepatically expressed. The next chapter asks how changes in circulating miRNAs, proteins and whole-blood gene expression relate to each other by integrating these datasets.

5 Multi-omics Integration of Plasma Protein, Plasma sRNA and whole-blood RNA Sequencing Data from an Infant Vaccine Study

5.1 Introduction

EUCLIDS data presented in Chapters 3 and 4 generate further questions. Why are predominantly liver-derived miRNAs downregulated in plasma after vaccination? Do cytokines released after vaccination play a role in this? Does downregulation of (probably hepatically derived) plasma miRNAs relate to the acute phase? Given the action of sRNA-dfs are unknown—can connections be drawn between their expression and the transcripts they derive from? This chapter integrates the EUCLIDS plasma sRNA, protein and whole-blood RNA-seq dataset to address these questions.

5.2 Chapter aims

Chapter aims are:

1) Explore the relationship between plasma miRNA, protein and blood RNA and develop an integrated model of vaccine-induced changes that may identify correlates of vaccine reactogenicity.
2) Mine the above outputs to form new hypotheses surrounding the infant immune response to vaccination, which can be tested *in-vitro.*

5.3 Chapter methods

5.3.1 Samples and datasets

The sRNA-seq and protein datasets come from the EUCLIDS study analysed in Chapters 3 and 4. Whole blood gene expression was measured by RNA-sequencing—generated using a ribo-depleted library preparation method and presented in O'Connor et al.,[76]. Only data from baseline and 24-hour post-vaccinations samples were included in the integrative analysis as this coincides with maximal differential expression in each dataset.

5.3.2 Assessing the overlap between differentially expressed features

The counterparts of sRNA-dfs were the full-length gene or protein to, which they map. miRNA counterparts were their experimentally validated targets (see methods 2.6)[339]. DE features with a DE counterpart were included in a chord diagram created using the circlize package in R[412].

5.3.3 Pathway enrichment and overlap

GO:BP terms were interrogated for enrichment in each dataset. MIEA2 was used for the miRNA data using a GSEA-style approach (see 2.7.2 and 3.4.6)[413]. Fgsea was used for mRNA data GSEA[414]. The STRING database was used for the ORA of DE proteins[415].

Despite some overlap in identical term names across the omics datasets, it was clear that synonymous but non-identical terms also existed, e.g. "coagulation",

"haemostasis", and "blood coagulation". REVIGO provides a method of showing common themes in a union of GO terms. However, terms *between* lists of GO terms sets must be manually parsed to look for similarities. An automated method was developed that links common but non-identical terms for each dataset in a bipartite network. This method exploits the hierarchical organisation of GO:BP in the GO database in, which specific (child) terms are linked to less specific (parent) terms. The method highlights instances where different child terms of the same parent term are identified across all three datasets—making it easier to spot common themes.

5.3.3.1 Algorithm

The GO:BP terms in each omics dataset were mapped to its parent and child terms. An iterative algorithm used these terms to map each enriched GO:BP term to the first common ancestor shared across all omics datasets ("first multi-omics common ancestor"). For example, "positive regulation of blood coagulation", enriched in the miRNA results, and "blood coagulation" in the mRNA results are child terms of "coagulation", which was present in the proteomics data. Networks were created that linked each original term to its first multi-omics common ancestor. This network was visualised in Cytoscape and manually searched to identify communities with a common theme.

5.3.4 Integration of omics datasets based on feature expression profiles

An issue with integrating multiple omics datasets through simple correlation (i.e., correlating each feature in each omics dataset with all other features across all datasets) is that it fails to consider variability between datasets due to technical noise and omics-specific bias. For example, differences in library preparation between the miRNA and mRNA datasets create omics-specific noise—essentially omics-specific "batch effects". This means features tend to correlate better within a dataset than across datasets, obscuring cross-omics correlations. To overcome this, a regularised generalised canonical correlation analysis (CCA) implemented in the mixOmics R package was used to integrate datasets[416]. CCA (implemented by the wrapper.rgcca function) is an unsupervised approach that aligns multiple omics datasets by maximising correlation across the datasets.

5.3.4.1 Pre-processing of data prior to CCA

Normalised, logged expression values were input into CCA. The effect of the individual was regressed out to account for pairing in the data using removeBactEffect in the limma package. PCA showed cohort age strongly affected feature expression; therefore, cohort effect was removed using the removeBactEffect() function. wrapper.rgcca() automatically scales each dataset.

5.3.4.2 Interpreting CCA

The CCA algorithm produces linear combinations of the original variables multiplied by variable-specific weights to maximise correlation between datasets. This involves multiplying the scaled and normalised expression of each feature ("variable") in a dataset by a number (known as a weight or loading) and adding

up the results to give a "score"—effectively performing a weighted sum of feature expression. CCA finds the optimal loadings for each feature in each dataset that results in similar sample scores across each dataset. Consequently, highly correlating mRNAs, miRNAs and proteins across datasets receive higher weights.

The CCA algorithm creates one or more components. Each linear combination represents one component of the CCA. Variable loadings can be visualised for each dataset in each component. Features with the highest loadings in a component are the top correlating features across the omics datasets in that component, i.e., they "drive that component". Two components are usually visualised simultaneously in a graph, with each axis relating to a component score.

Rather than plotting the CCA results for each omics type for each sample individually, the average ("centroid") component score across omics datasets can be plotted. Samples clustered together in this graph have similar expression patterns of highly correlating features.

5.3.4.3 Relevance network

A relevance network was created on the CCA output using the mixOmics' network() function. This function creates a bi-partite network of nodes with edges connecting pairs of features correlating across datasets. The network was built from features with a correlation of at least 60% in components 1 or 2 of the CCA. Subgraphs within the network were identified by edge-based clustering

using the community clustering function, cluster_fast_greedy(), from the igraph package[417].

5.3.4.4 Finding a signature that can differentiate vaccine groups using PLS discriminant analysis

A partial least squares discriminant analysis (PLS-DA) approach was employed to identify features that differentiate vaccine groups at the post-vaccination timepoint—this was done using the MixOmics' DIABLO method[416]. To find a model that would perform well for both ages, the 4 and 12-month cohorts were amalgamated and split into a training and test dataset at a 7:3 ratio, respectively. The data split was stratified to ensure a balanced number of samples from each cohort and vaccine group. Feature selection was employed to reduce over-fitting, minimise redundancy and create a signature amenable to testing using non-omics technologies.

5.3.4.5 Interaction network centred on hepatocytes

An *in-silico* network was created to investigate what may drive downregulation of hepatically derived miRNAs in plasma and how this may relate to the liver-derived acute phase response. The network highlights interactions between validated DE features that have some relationship to the liver, i.e. hepatically derived plasma proteins and miRNAs, and DE mRNAs that encode secreted proteins with hepatic receptors. Figure 52 contains the algorithm behind this network[339,418–422].

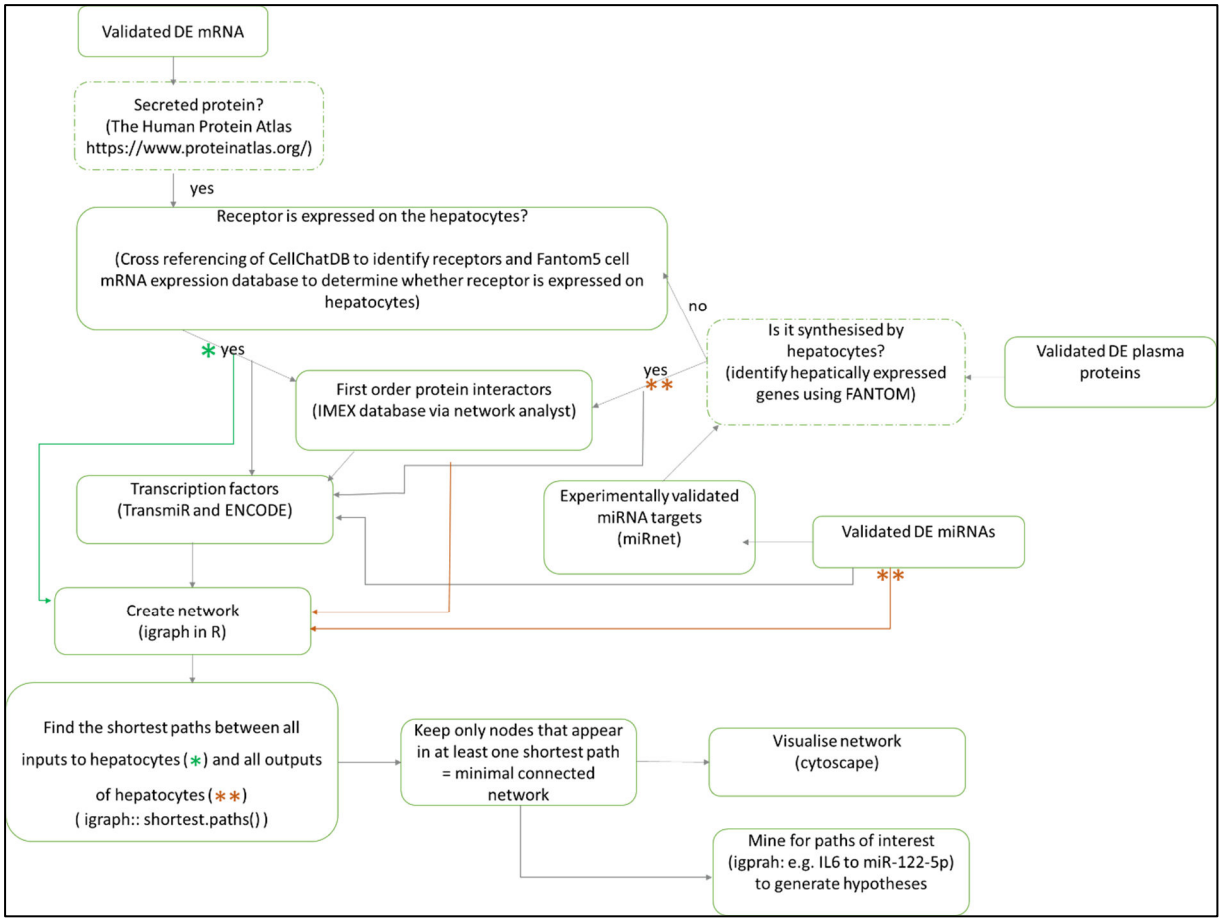


Figure 52 Algorithm for building the interaction network.
IMEX is a protein-protein interaction database. Cell-chat database contains ligand-receptor pairings. TransmiR and ENCODE contain information on transcription factors target genes. Information on hepatocyte gene expression was downloaded from the FANTOM5 project. miRnet contains experimentally validated miRNA—mRNA target pairs. The network was built from: validated DE mRNAs that are secreted proteins and have liver-expressed hepatic receptors; validated DE plasma proteins that either have hepatic receptors or are synthesised by the liver; hepatic receptors of DE features; DE miRNAs and their targets; hepatically expressed first order interactors of the above proteins and genes; and liver expressed transcription factors of all the above. To simplify the network, the shortest paths were calculated between all potential hepatocyte inputs and outputs. Only features present in at least one shortest path were retained in the network to prevent a "hairball" effect from an overly dense network.

5.3.5 In-vitro model

An *in-vitro* model was set up to recapitulate downregulation of miR-122-5p release 24 hours after vaccination. The huh-7 hepatoma cell line was used as its

expression of hsa-miR-122-5p is closest to primary hepatocytes compared with other commonly used liver cell-lines[423,424].

Based on a pilot experiment, huh-7 cells were seeded at 300,000 cells per well of a 6-well plate in 500μl of media supplemented with streptomycin and 5% FCS. After being left to recover for 48 hours, they were washed three times with serum-free media and then suspended in 100 pg/ml of IL6 in serum-free media (IL6 stimulated) or serum-free media only (control). Supernatant and cells were harvested after 20 hours. This timepoint was used because the Luminex data showed IL6 was not increased at 4 hours; therefore, the maximum time hepatocytes had been exposed to elevated IL6 in infants was 20 hours before blood draw. Differential expression of hsa-miR-122-5p in control versus IL6 supernatant was tested using a one-sided Wilcox test, H_0 $\log_2$ foldchange equal or greater than 0. Differential expression of intracellular hsa-miR-122-5p between control and IL6 stimulated cells was tested using a two-sided Wilcox test.

5.4 Results

5.4.1 Samples

Only samples with paired data collected across all omics datasets were used for integration at the sample-wise level (
Table 19). Thirteen participants had mRNA, protein and sRNA datasets at both 4 and 12 months.

Table 19 Number of samples with paired data collected across all omics datasets.

	Number of samples with all three omics at their pre and post-vaccination time points	
Vaccine group	4 Month Cohort	12 Month Cohort
Routine	10	10
Routine + 4CMenB	11	10
Total	21	20

5.4.2 The overlap between detected and differentially expressed sRNAs, mRNA and protein data sets at the feature level

sRNA-df, mRNA and protein that map to the same gene are called "counterparts". Counterparts identified in two or more datasets define an overlap. The overlap in detected and DE features across datasets is summarised in Figure 53. Figure 54 depicts the overlap between DE features in the sRNA, mRNA and protein datasets, plus DE miRNAs that target them (at the transcriptomic level).

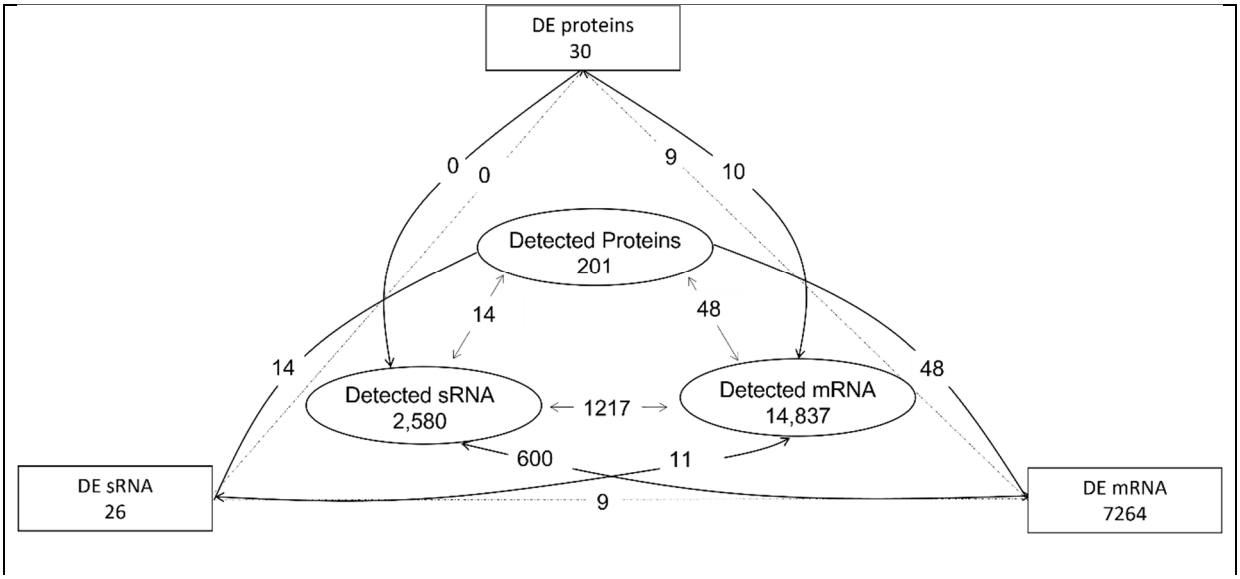


Figure 53 Overlaps between features detected in the mRNA, sRNA and protein datasets. The number of features within each dataset is contained within the ellipses and rectangles. The number of features with counterparts in two datasets is in the arrow joining those datasets. DE means differentially expressed 24 hours post-vaccination. Detected means present in the dataset after filtering out low abundance features.

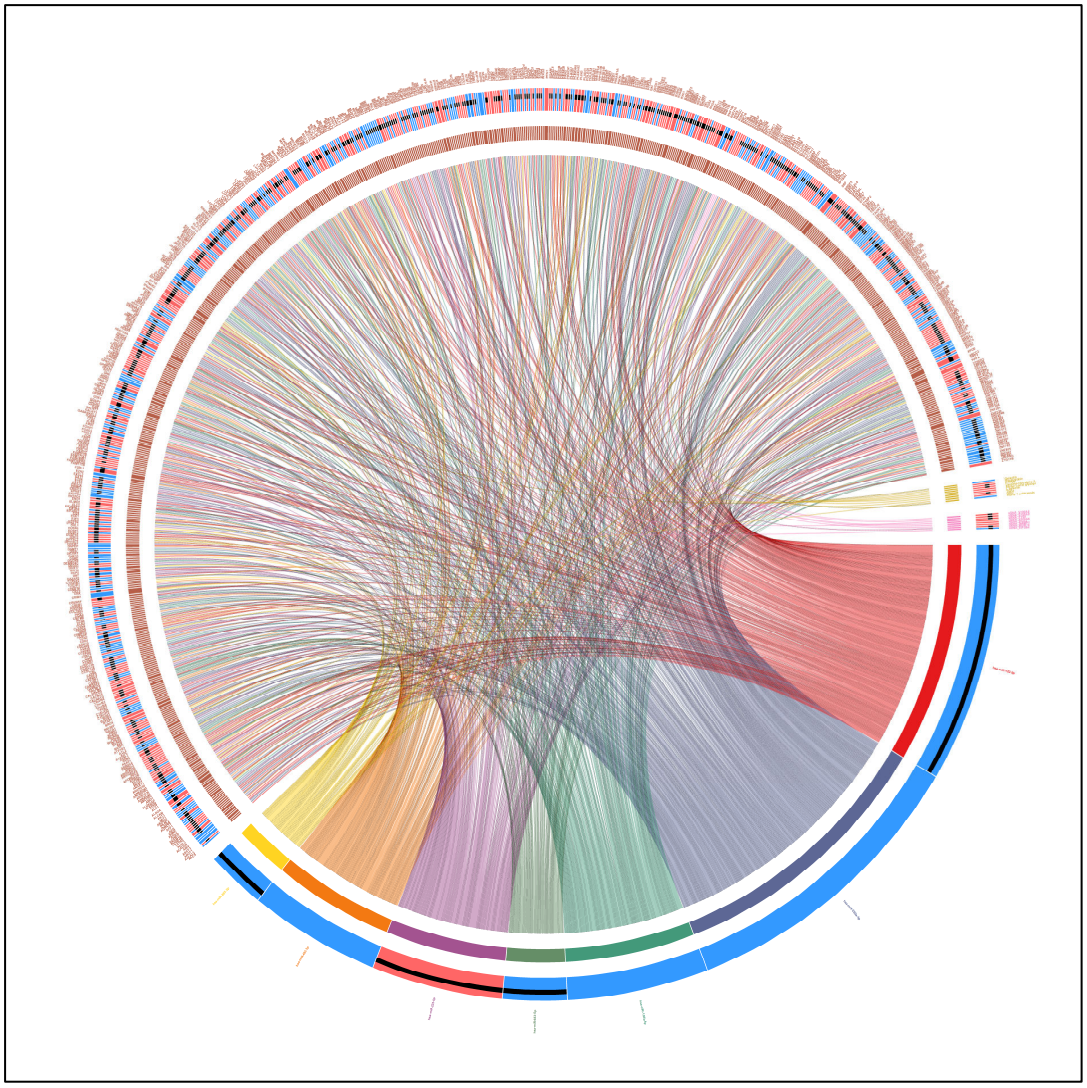


Figure 54 Chord plot showing the relationship between features (miRNAs, sRNA-dfs, proteins and mRNAs) DE at baseline versus 24 hours post-vaccination. Download the high-resolution link to zoom:
https://www.dropbox.com/s/3nxdjdktf0hst2y/Euclids%20chordplot.png?dl=0

***Track 1** (Outermost track) = feature name*
***Track 2** = whether that feature was up (red) or downregulated (blue) 24 hours after vaccination. The black bar indicates whether this feature was validated as DE in the 12-month-old cohort.*
***Track 3** = colour bar for that feature. The length of the feature bar corresponds to the number of links the feature has with other features.*
***Linking lines** ("chords") represent a connection between features. mRNA/protein/sRNA and their counterparts are connected. E.g. the protein Alpha-1-antitrypsin is connected to its mRNA counterpart SERPINA1. sRNA-dfs are connected to their full mRNA counterparts. miRNAs are connected to their targets. Links with dotted edges connect features with opposite directions of foldchange. Links without dotted edges connect features with the same direction of foldchange. Features with at least one link are shown.*

5.4.2.1 Overlap between mRNAs and proteins

Ten out of thirty DE proteins were detected at the mRNA level in the whole blood RNA-seq data, of which 9 were DE, demonstrating consistency between the RNA and protein data sets at the DE level (Figure 53). Despite agreement in DE, there was no correlation between protein and mRNA level $\log_2$ fold changes (Appendix Figure 23), even when restricting to DE features. This lack of correlation could be due to some proteins being produced outside of blood, e.g. in the liver; though this cannot explain upregulation of lactotransferrin in plasma ($\log_2$ foldchange = 3.22, p=6.7x10^{-9}, FDR=2.3x10^{-7}) yet no change at the gene expression level ($\log_2$ foldchange = 0.37, p=0.2) given this protein is enriched in neutrophil secondary granules (Figure 53).

5.4.2.2 *Overlap between differentially expressed miRNAs and their targets*

Of the 7,294 DE proteins and genes, 853 were targets of DE. Gene set enrichment is required to understand whether this figure represents a statistically significant overlap. GSEA showed that at 4 months, miR-122-5p gene targets are positively enriched in the results of the mRNA sequencing analysis when genes are ranked by p-value. This indicates targets of miR-122-5p generally change in expression irrespective of direction (Figure 55A upper panel). This finding was validated at 12 months (Figure 55B upper panel). In addition, miR-885-3p targets were enriched in upregulated genes at 4 and 12 months (Figure 55 lower panels).

To determine whether the collective targets of DE miRNAs predominantly expressed in the liver are enriched in the mRNA data, the combined targets of these miRNAs were used as a gene set. This gene set was significantly enriched in the mRNA expression data at 4 and 12 months (Figure 56 upper panels). Since miRNAs can act co-operatively, the analysis was repeated, but the gene set was restricted to targets of at least two DE miRNAs; the enrichment was even more significant (Figure 56 lower panels). These enrichment results do not prove miRNAs are intercellular regulators of gene expression, but they add plausibility to that possibility.

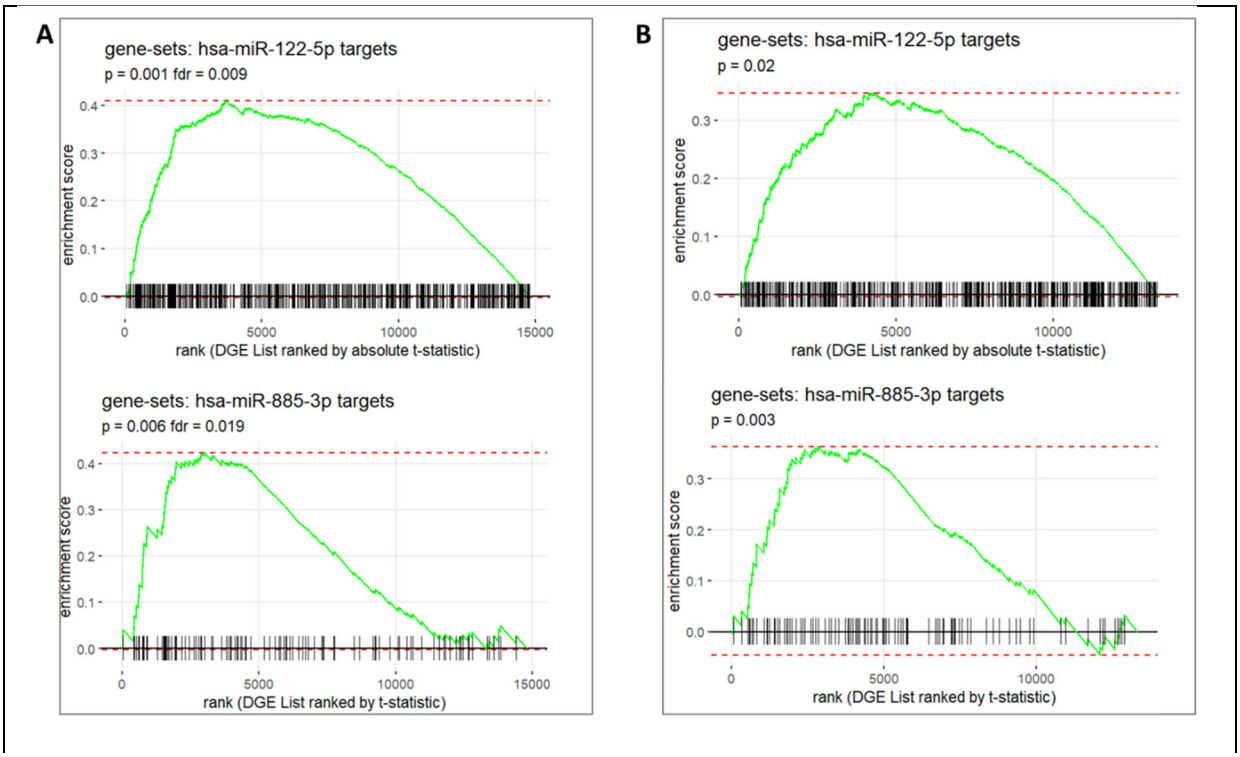


Figure 55 GSEA using miRNA targets as gene sets in the mRNA data analysing baseline versus 24 hours post-vaccination. The plots show target enrichment of hsa-miR-122-5p and hsa-miR-885-3p amongst RNAs whose expression changes the most in whole blood after vaccination.
A. results in the 4-month-old cohort.
B. results in the 12-month-old cohort.
Upper panels = GSEA for hsa-miR-122-5p targets when the mRNA differential gene expression list is ranked by absolute t-statistic. Lower panels = GSEA for hsa-miR-885-3p when the mRNA differential gene expression list is ranked by t-statistic

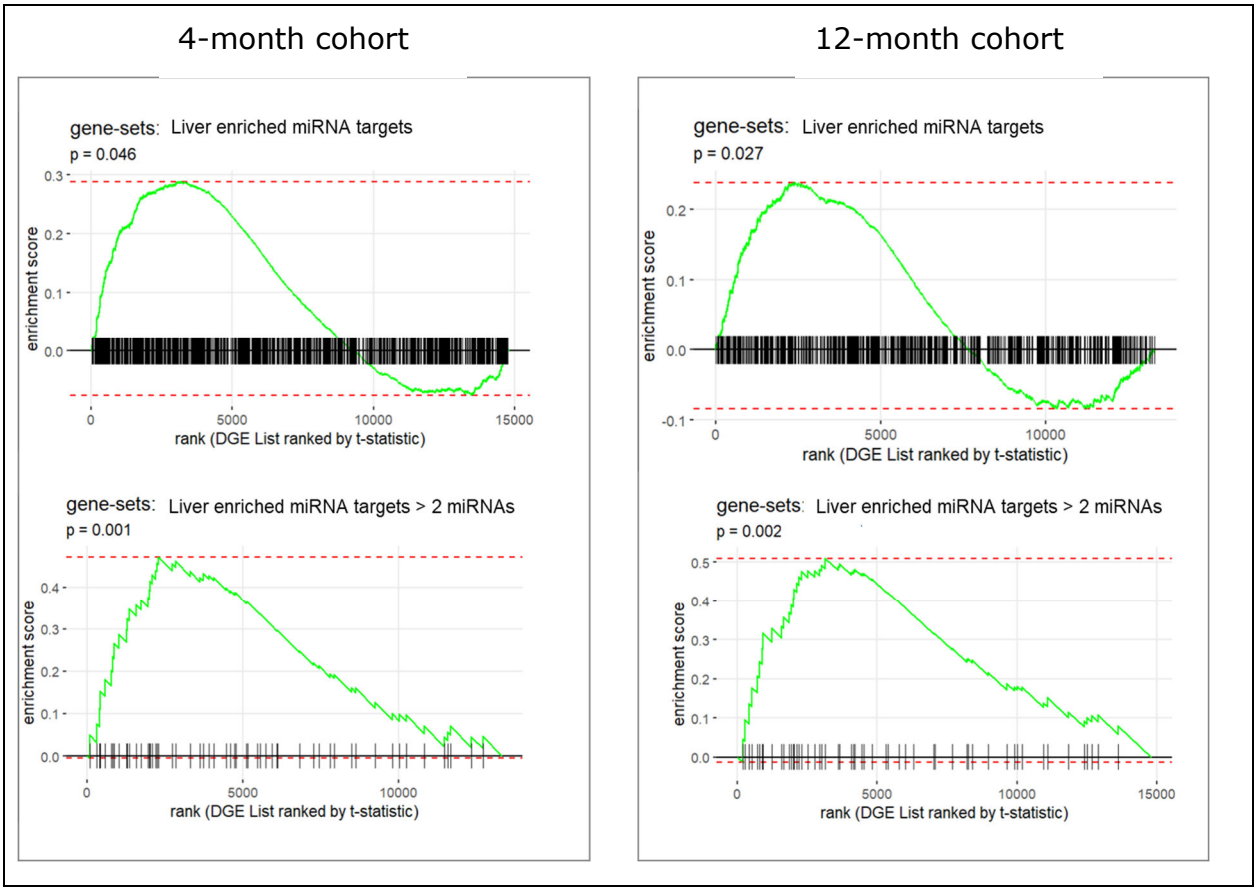


Figure 56 GSEA of the mRNA data (baseline versus 24 hours post-vaccination) using targets of DE liver-enriched miRNAs. The plots show target enrichment of miRNAs downregulated in plasma amongst RNAs whose expression changes the most in whole blood after vaccination.
A. results in the 4-month-old cohort.
B. results in the 12-month-old cohort.
Upper panels = GSEA for all targets of the DE liver-enriched miRNAs—hsa-miR-122-5p, hsa-miR-483-5p, hsa-miR-483-3p, hsa-miR-885-3p, when the mRNA differential gene expression list is ranked by t-statistic
Lower panels = GSEA for genes that are targeted by at least two DE liver-enriched miRNAs when the mRNA differential gene expression list is ranked by t-statistic

5.4.2.3 Overlap between mRNAs and sRNA-dfs

At 4 months, 26 sRNA-dfs were DE at baseline versus 24 hours post-vaccination in at least one or both vaccine groups. Eleven of these had counterparts identified in the whole blood RNA-seq data, of which nine were differentially expressed. These were S100A8, S100A9, FYB1, LUCAT1, S100A6, RGS2, NAMPT, WDR74, and EIF2B4 (Figure 57). Though the expression of DE sRNA-dfs and

their mRNA counterparts positively correlated, there were three examples where differential expression occurred in the opposite direction (Figure 57).

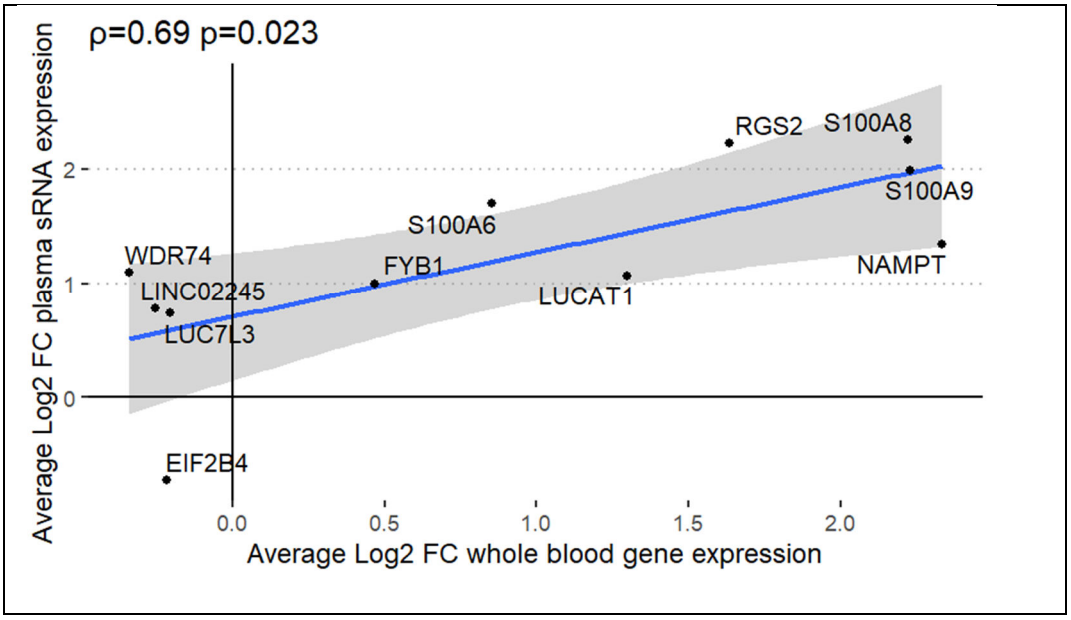


Figure 57 Correlation between DE sRNA-dfs and mRNA counterparts. Correlation tested with Spearman correlation. Features in the upper left quadrant have opposing direction in log_2 foldchanges.

5.4.3 Overlap between miRNA and mRNA and protein data sets at baseline vs 24-hour post-vaccination at the pathway level.

GO:BP terms enriched after vaccination were compared between datasets. This revealed cross-omics similarities; 46% of enriched terms in the proteomics analysis were enriched in the mRNA or miRNA analyses (Figure 58). Only "Defense response to gram-positive bacterium" was enriched in all datasets (Figure 58).

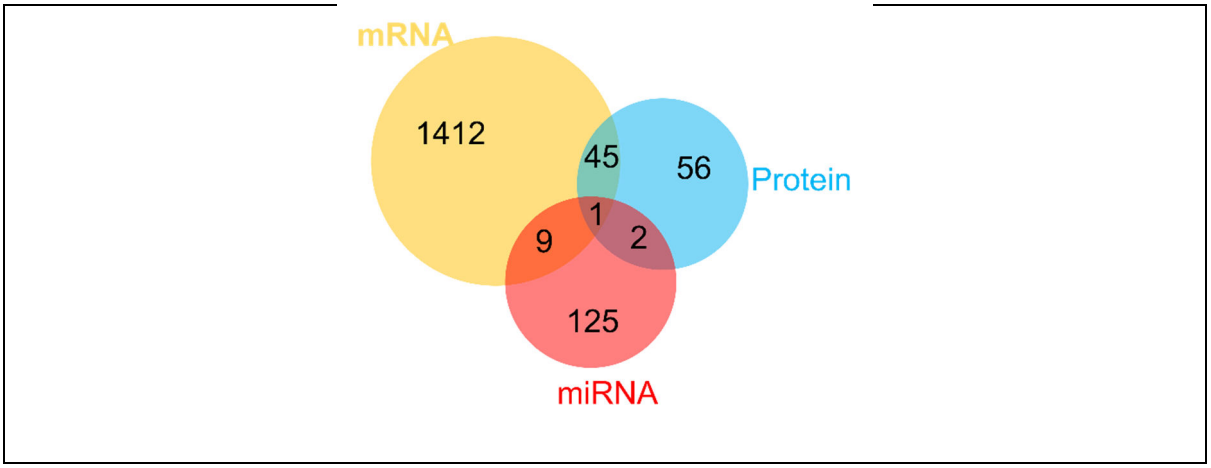


Figure 58 Venn diagram of overlapping terms derived in gene set analyses of baseline vs 24 hours for the mRNA, miRNA and protein datasets in the 4-month cohort.

Some similarities across the dataset were missed by looking for identical term names. For example, "regulation of coagulation", "blood coagulation", and "positive regulation of blood coagulation" were present in the protein, mRNA and miRNA datasets, respectively and relate to the theme of haemostasis, yet there was no overlap at the term level. Thematic commonalities across the data sets were better revealed by identifying common ancestor terms between enriched GO:BP terms in each dataset (Figure 59). Shared themes included: generic immune system terms, humoral immune responses, protein transport and localisation, chemotaxis and cell migration, TLR signalling, cytokine production and signalling, lipid metabolism, coagulation and vesicle transport (Figure 59).

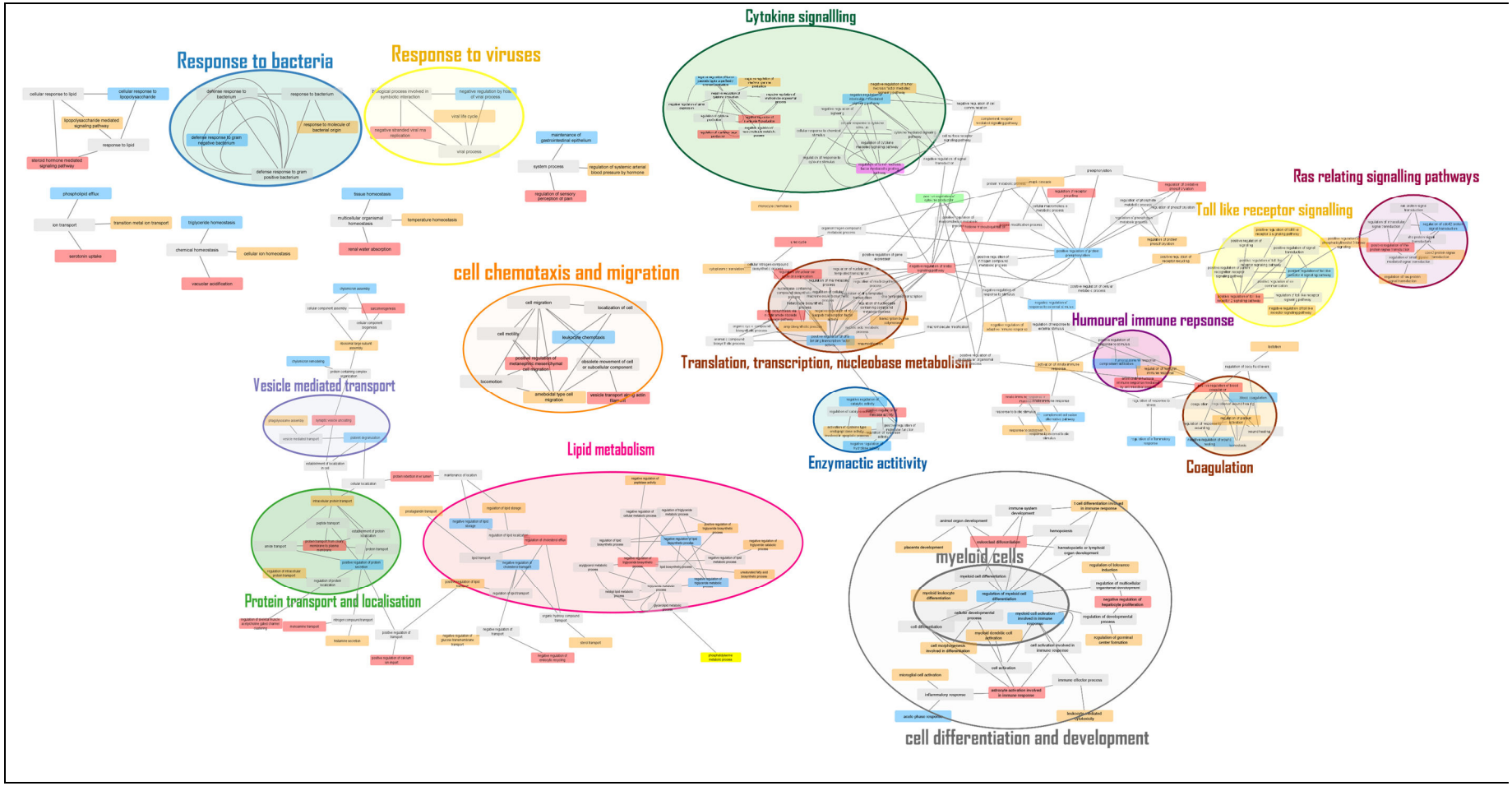


Figure 59 Consolidation of miRNA, mRNA and protein gene ontology networks. Each omics dataset's significant terms (FDR<0.05) were mapped to the first common ancestor. Coloured terms show child terms; grey terms show the first common ancestor. Note that the first common ancestors that were differentially enriched/depleted in one or more datasets are still coloured grey. Terms that coalesce around a theme, e.g. lipid metabolism, are encircled. Red nodes=miRNA, blue nodes=protein, orange nodes=mRNA, grey nodes = first common ancestor shared by GO:BP terms in all the datasets. High-resolution image download link: [High-resolution network image]

5.4.4 Integration of omics datasets based on feature expression profiles

CCA identified correlation patterns between datasets. Pre- and post-vaccination samples clearly separated in the CCA1 and 2 (Figure 60) at 4 and 12 months, with 4CMenB and routine group post-vaccination samples also separating. These findings suggest vaccination (and 4CMenB coadministration in particular) drives perturbation of networks of co-correlating features (Figure 60). These co-correlating features were extracted from the CCA and presented as a bi-partite relevance network (Figure 60B). Community clustering identified two linked subnetworks of this bi-partite relevance network (Figure 60B). The blue subgraph contains features, which drove CCA 1. ORA of features in this network revealed generic terms regarding the immune response and the acute phase, chemotaxis, cytokine production and response to lipids (Appendix Figure 25). The red subgraph mostly contained features that drove the CCA 2 and were associated with differences between vaccine groups after vaccination, particularly at 4 months. ORA revealed terms relating to mitochondrial action, e.g. oxidative burst (Appendix Figure 26).

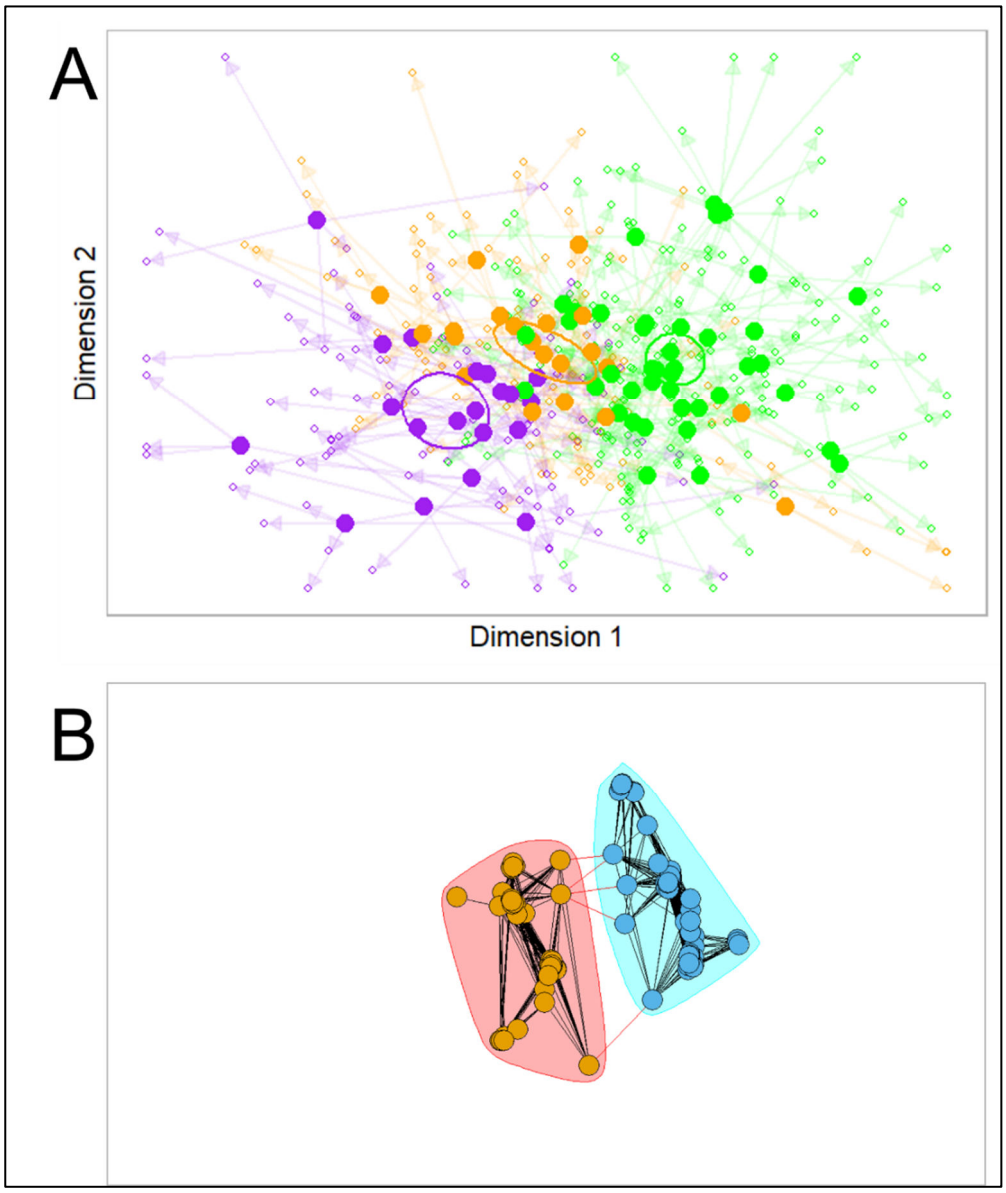


Figure 60 Multi-omics CCA of 4 and 12-month cohort pre and post-vaccine samples
A. Sample plot on the first two CCA components. Arrows show where a sample plots in the individual datasets in the first two CCA components; dots represent the centroid of this location (essentially the average location of the sample across all 5 datasets after maximisation of the correlation between the datasets). Green = pre-vaccination samples, orange = post-vaccination samples in the routine group, and purple = post-vaccination samples in the 4CMenB group. The ellipses represent 95% confidence intervals around the centroids of samples taken pre-vaccination, post-vaccination in the routine group and post-vaccination in the 4CMenB group.
B. Bipartite relevance network constructed using feature pairs with a correlation of >0.6. The two sub-networks identified by community are demarcated by colour.

Given that CCA (an unsupervised analysis) suggested differences between vaccine groups, PLS-DA (a supervised analysis) was used to further tease out these differences by identifying features that could differentiate vaccine groups at the post-vaccination timepoint. The model perfectly assigned the training samples and 11/12 test samples, equating to a balanced error rate of 8%. The loadings for this model are contained in Figure 61A. Vaccine groups had significantly different sample centroid scores (Figure 61B). Sample centroid scores significantly correlated with fever (Figure 61 C). ORA on the selected features showed enrichment of terms relating to the acute phase response, immune response, cell migration and motility (Figure 62).

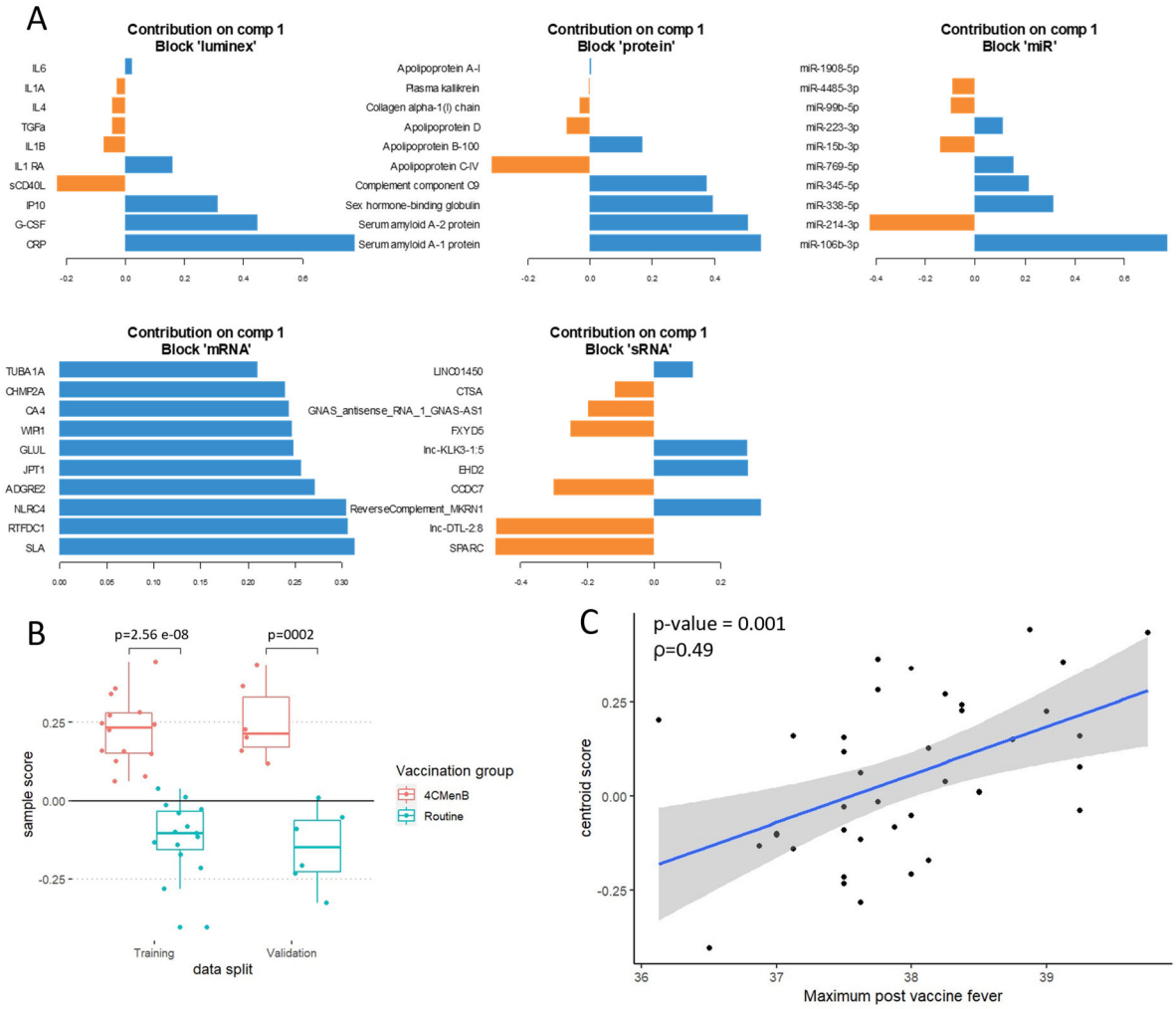


Figure 61 DIABLO output
A. Top 10 features in the loadings for each omics dataset
B. Box plot of sample centroid scores in the training and test (Validation) sets calculated from the DIABLO model
C. Correlation plot between maximum fever post-vaccination and centroid score calculated from the DIABLO
Block is Mixomics' notation for dataset

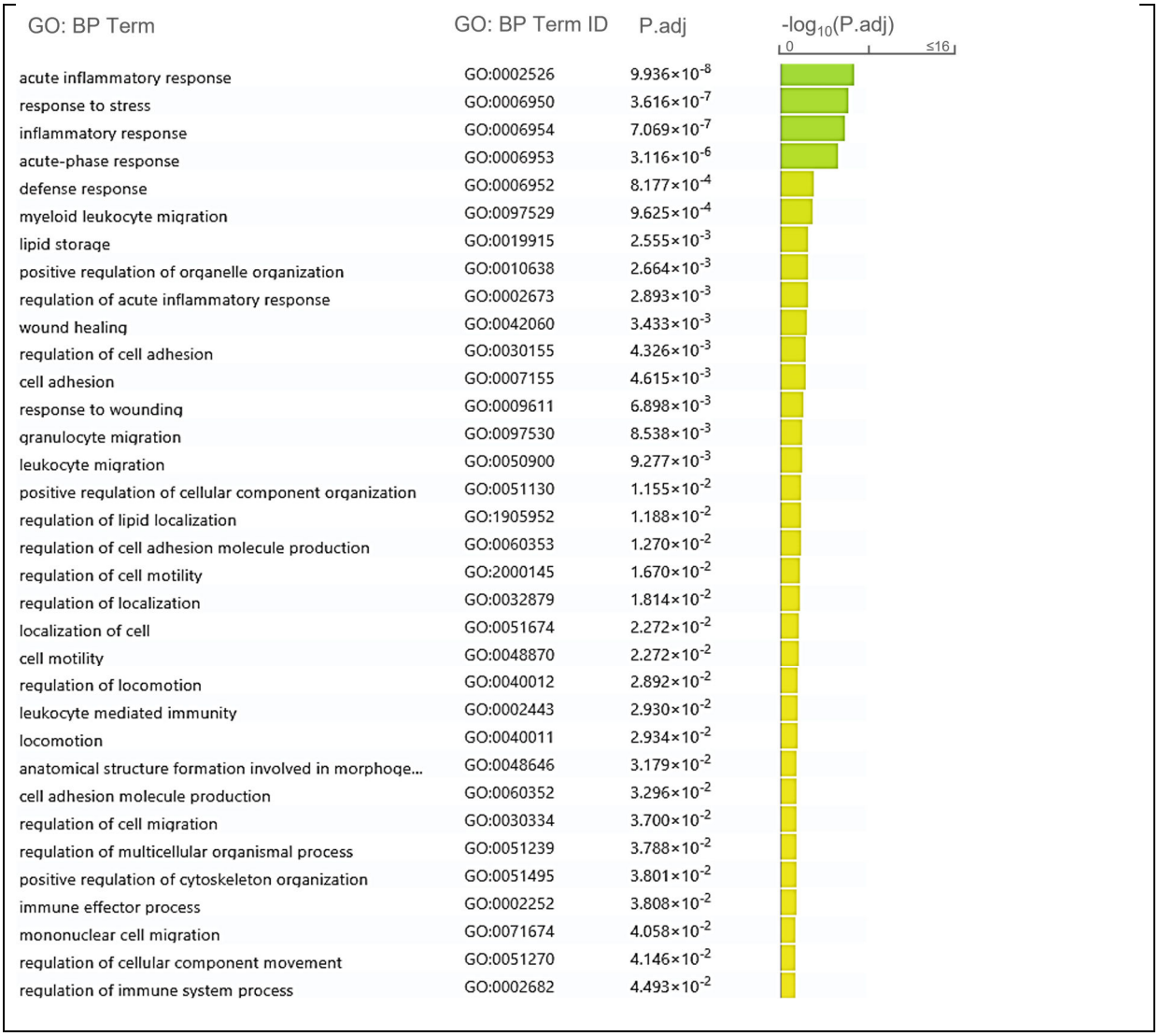


Figure 62 Gene set enrichment on features selected by the DIABLO model built to differentiate the vaccine groups

5.4.5 Multi-omics integration provides potential insights into the regulation of the acute phase response by the liver

Plasma downregulation of liver-enriched miRNAs and induction of the hepatic acute-phase response raises the possibility that miRNAs regulate pathways involved in the acute phase. To study this, an interaction network was built from validated DE features and their interactors, filtered for relevance to hepatocytes. This method identified proteins expressed within hepatocytes that might be

intermediaries in pathways linking stimulation of hepatocytes to the acute phase response (Figure 63). This network revealed that none of the validated DE miRNAs target DE acute phase proteins; however, they do target transcription factors and proteins within pathways that induce acute phase proteins. Moreover, there are links between known inducers of the acute phase and proteins that influence the sorting of miRNA into exosomes: e.g. IL6 interacts with the transcription factor IRF1, which induces YBX1 that influences miRNA packaging into exosomes.

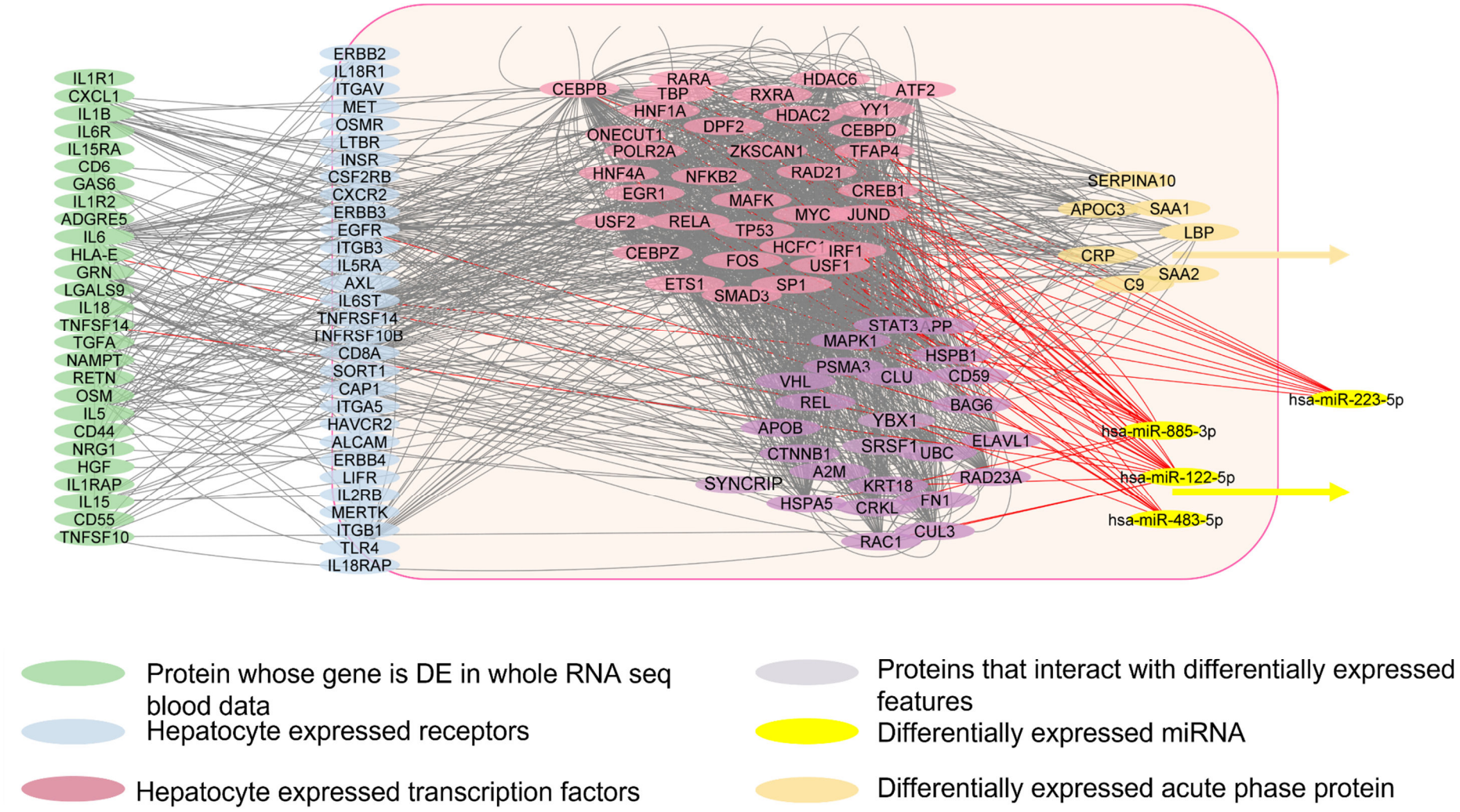


Figure 63 Interaction network centred on the hepatic response. An interaction network was built using validated DE features as seeds. This network aims to generate hypotheses for how DE features may be related.

5.4.6 IL6 inhibits miR-122-5p release from hepatocytes

An *in-vitro* model was set up to study the possible relationship between miR-122-5p, IL6 and the acute phase response using huh-7 cells. The potential of IL-6 to influence miR-122-5p release from huh-7 cells was investigated because IL-6: is a key inducer of the acute phase response; was validated as upregulated after vaccination in the Luminex data, and links existed between it, acute phase proteins and hsa-miR-122-5p in the interaction network. If a model of IL6 downregulation of miR-122-5p release could be established, it could be used to explore the mechanism behind vaccine-induced changes in miR-122-5p and determine the miRNA's role (if any) in the acute phase response.

IL6 stimulation of huh-7 cells led to a significant drop in hsa-miR-122-5p in the supernatant 20 hours after stimulation compared with control cells. The magnitude of the drop was similar to that seen in the EUCLIDS infants 24 hours after vaccination. Preliminary results do not show an intracellular change in hsa-miR-122-5p levels in huh-7 cells at 20 hours post-stimulation, but further repeats are needed to confirm this.

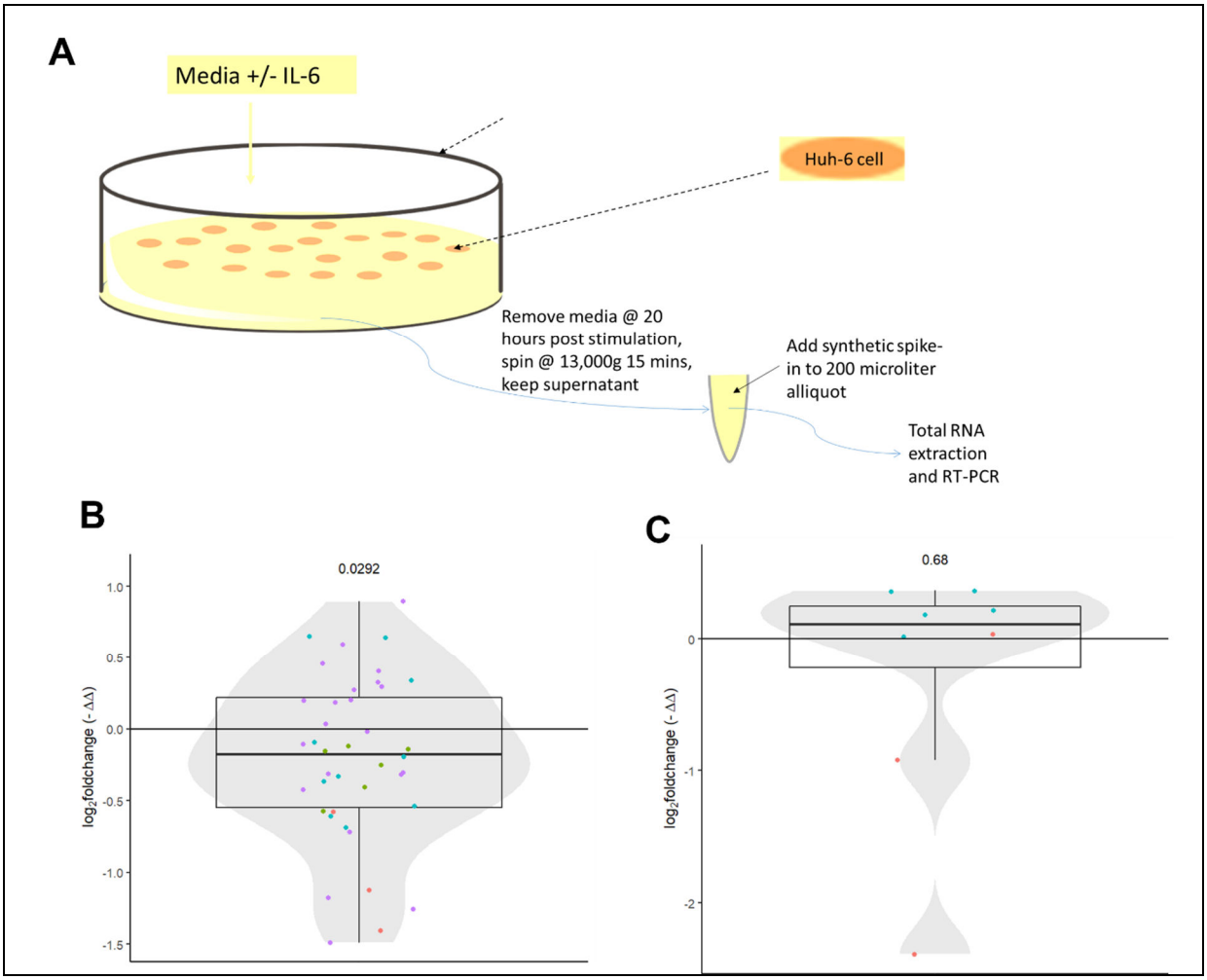


Figure 64 In-vitro model
A. Overview of the experimental method
B. Boxplot of log_2 foldchange of miR-122-5p expression in supernatant for IL6 stimulated vs control stimulated cells. Average of multiple replicates (replicate indicated by colour)
C Boxplot of log_2 foldchange of miR-122-5p expression in cells for IL6 stimulated vs control stimulated cells. Average of multiple replicates (replicate indicated by colour)

5.5 Chapter Discussion

This discussion is split into two; the first part discusses the data integration. The second part brings together the results of Chapters 3, 4 and 5, putting them into a broader perspective.

5.5.1 General comparability of the omics datasets

The simplest explanation for the overlap between DE counterparts in the sRNA-df and RNA-seq data is that DE sRNA-dfs are degradation products of DE full-

length RNAs. The positive correlation between the average log2 foldchange of DE sRNA-dfs and their full-length RNA counterparts is in line with this. Countering the degradation hypothesis, sRNA-df expression profiles along the genes they arise from are not random uniform (e.g., Appendix Figure 29), i.e. they are not characteristic of degradation products. Furthermore, direction of foldchange of three differentially sRNA-dfs was opposite to their DE mRNA counterparts. Disconcordance between sRNA-dfs and full-length RNAs could arise due to non-blood production of sRNA-dfs and technical differences between sRNA and standard RNA-seq. An alternative possibility is that sRNA-dfs may be endo-siRNAs (see introduction 1.2.2.7) with biological functionality, e.g. a regulatory role—a concept well established for piRNAs but not endo-siRNAs. There is very little literature on endo-siRNA since they were first postulated to exist in human cells in 2014[425].

Average protein and mRNA $\log_2$ fold changes did not correlate. This may be because of protein production outside of blood, unbalanced protein transcription and turnover as proteins are used up (e.g. scavenging proteins like gelsolin and alpha-1-antitrypsin), lag-times between mRNA transcription and protein secretion, and post-transcriptional regulation; all of these explanations may occur concurrently.

The miRNA-enrichment analysis provided circumstantial evidence that plasma miRNAs derived from liver may be linked to post-transcriptional regulation of mRNAs in blood after vaccination. Positive enrichment of the targets of DE liver-enriched miRNAs in the mRNA DE data was surprising given the limitations of

comparing two different sample types. The overlap in pathway enrichment results across the omics datasets at the GO:BP term and thematic level provide corroborative evidence that findings within individual datasets are robust, and again hints at the possibility that DE miRNAs are part of the acute phase response either within their cognate cells or as intercellular post-transcriptional regulators. It is even possible that miR-122-5p drops in serum after vaccination due to leukocyte uptake. This study does not prove DE miRNAs are intercellular regulators of gene expression, but the concept is in line with findings from other studies[164–169]. If extracellular miRNAs produced by the liver are intercellular post-transcriptional regulators, they represent a novel component of the hepatic acute phase response.

5.5.2 A network of post-vaccination immune perturbation

CCA identified two networks of correlating features, which separated baseline and post-vaccine samples at the group level. Interestingly, miRNAs within the relevance network do not connect with their targets. This may be because nodes are indirect targets of miRNAs (with correlations arising via intermediaries) or correlations between miRNAs and proteins/mRNAs are driven by separate facets of the immune response that run in tandem.

The CCA2 network drove separation between the vaccine groups in the 4-month cohort. ORA identified mitochondrial functions as enriched within this network, reflecting DE mitochondrial sRNA-dfs after vaccination in the 4CMenB group.

5.5.1 Finding a PLS signature that captures reactogenicity could facilitate vaccine design

The EUCLIDS study (using all study participants, not just those with samples in the multi-omics study) and other clinical studies show fever is higher when 4CMenB is included in vaccine regimes[427]. However, maximum fever was not significantly higher in the 4CMenB group in the 12-month cohort of the EUCLIDS multi-omics sub-study, likely due to sample size. This illustrates the difficulties in capturing and comparing reactogenicity in vaccine studies. Noise and high person-to-person variability in reactogenicity measurements (arising from a subjective recording of symptoms and technical issues with fever measurement) necessitate moderately large sample sizes to get an accurate representation. Such sample sizes are not conducive to phase I studies and means "best guesses" must be made about which dose, formulation and administration site to use rather than systematically testing each variable. A signature that can predict vaccine reactogenicity in a small number of participants could enable multiple small subgroups in phase I studies, enabling empirical identification of the best vaccine variables, reducing guesswork and rationalising vaccine development. A PLS-DA signature differentiated the more reactogenic 4CMenB group from the routine group. Future work is needed to demonstrate whether the PLS-DA signature can predict the relative reactogenic potential of other vaccines/vaccine schedules more accurately than fever measurement. To translate this signature to a clinical setting, it would be prudent to shrink the number of features (e.g. using a LASO approach) and convert feature measurement onto a simple, easy-to-run, economical platform.

If the PLS-DA signature is confirmed to correspond to vaccine reactogenicity, it may be possible to develop it into an *in-vitro* assay to assess the reactogenic potential of vaccine batches. There is a precedent for this in the monocyte activation test (see section 1.2.1.4.2). This would require further work to characterise the cells that secrete features in the PLS-DA signature in response to vaccine exposure.

5.5.2 The Role of the Liver in Vaccine Responses

Coincident induction of the acute phase response and plasma downregulation of liver-expressed miRNAs raises the possibility that hepatocyte miRNAs regulate pathways involved in the acute phase response. Given the impossibility of liver biopsy in vaccine studies, the interaction network inferred from DE features and their known interactors generated a theoretical window into hepatocytes. The interaction network identified paths linking IL-6 to miR-122-5p and CRP production. The network also showed IL6 interacts with the transcription factor IRF1, which induces YBX1 and hnRNPA2B1, both of which influence miRNA packaging into exosomes; this may explain why liver-enriched miRNAs are downregulated in plasma. Future *in-vitro* work can explore these hypotheses.

5.5.3 Limitation of integration

A limitation of the integration work is its oversimplification of complex processes. Though single-cell studies show miRNAs correlate with their targets, detecting such correlations in bulk sequencing is complicated[428]. It may be unrealistic to

expect miRNAs to correlate with their targets given omics datasets were derived from different sample types, miRNA binding can lead to mRNA storage rather than degradation, and mRNAs and protein levels depend on many factors influencing their transcription, translation and turnover rates. These factors could obliterate correlations, even in the presence of miRNA targeting.

Another limitation of the integration work, particularly the interaction network, is the assumptions that must be made; for example, there can be false positives in experimentally validated targets of miRNAs. The interaction network is also at risk of over-simplifying complex interactions. Firstly, only first-order interactors were included in the network. Secondly, only features identified in one or more shortest paths between hepatocyte inputs and outputs were retained. Thirdly, protein-protein interactions are undirected, so a path from A to C through B (i.e. A→B→C) is not biologically valid if B influences A or C influences B. Fourthly, interactions are modelled as static when really they are dynamic, comprising positive and negative regulatory loops, with groups of features acting in an additive, synergistic or antagonistic manner. Putting aside these caveats, the network offers a heuristic approach to infer what may be happening within hepatocytes: its purpose is to generate hypotheses that can then be tested *in-vitro* or *in-vivo* in mice models. In support of this approach is the *in-vitro* work, which showed IL-6 stimulation of huh-7 cells leads to a drop in miR-122-5p release. The next step is to mine features from the network that link IL6 to miR-122-5p. These features can then be assessed in the *in-vitro* model to investigate how and why vaccination leads to a drop in hsa-miR-122-5p and whether this is linked to the release of acute phase proteins.

The huh-7 model requires further work to reduce assay variability. Media was not supplemented with FCS during stimulation as it can contain hsa-miR-122-5p, and many huh-7 studies omit it during stimulation. Omission of FCS however, may stress cells leading to death and miR-122-5p release. The effect of this should be explored in future experiments.

5.5.1 Why does IL-6 lead to a drop in miR-122-5p release?

miR-122-5p plasma levels could be downregulated due to: 1) reduced transcription within hepatocytes so less secretes out, 2) reduced secretion to maintain or increase hsa-miR-122-5p within hepatocytes, or 3) increased uptake by other cells, thereby depleting plasma. The first two possibilities could be tested in the huh-7 model. Hsa-miR-122-5p levels intracellularly did not change 20 hours post-stimulation. Further repeats are needed to confirm this lack of change; however, this mirrors the lack of change seen in mice liver post-vaccination (section 3.3.5), which was discussed in Chapter 3, section 3.5.4.3. As with the mouse experiment, earlier time points should be investigated as intracellular levels may change before 20 hours. Future work should focus on proteins known to control miRNA sorting into exosomes, e.g. YBX1 and hnRNPA2B1, in case this influences miR-122-5p efflux with little effect on intracellular hsa-miR-122-5p levels.

5.5.1.1 Limitation of the in-vitro model

Cell lines are useful but limited models of *in-vivo* phenomena. Repeating the experiment in another liver cell line, or better still, cells more genetically similar to normal liver, e.g. hepatocytes induced from pluripotent stem cells or primary cells, would be helpful. Once the mechanism behind reduced hsa-miR-122-5p plasma levels after vaccination is established, inducible knock-out mouse models could be used to manipulate hsa-miR-122-5p to establish its effect on the acute phase and vaccine responses.

5.6 Overall discussion of the EUCLIDS multi-omics study

5.6.1 Insights into the biology of the innate immune response in infants generated from the EUCLIDS study.

The first year of life is pertinent to vaccinology as it is when the majority of vaccines are given. Most studies investigating immune ontogeny and immunocompetence in infancy focus on the neonatal period; only a few have looked at early childhood[429–434]. Though there is a peak of morbidity and mortality in the neonatal period, deaths from infections remain elevated in the first year of life[435]. The immune response in the first year of life must be studied to reduce this death rate. Capturing the onset of the innate immune response is difficult in a clinical setting where children present with a multitude of aetiologies at a variety of times, rarely within 24 hours after infection onset. The EUCLIDS study is a unique model of innate immune response to immunological challenge in infants. Given that infants are responding to bacterial antigens, it may, to some extent, model innate immune responses to infection.

5.6.2 Age-related effects

Across all datasets, the number of DE features pre versus post-vaccination, and between vaccine groups were higher in the 4-month-old cohort compared with the 12-month cohort. There were also more correlations between features and vaccine reactogenicity markers and more differences between vaccine groups in the 4-month cohort. Differences between vaccine groups after vaccination were limited at 12 months (only CRP was different); however, integration of the datasets rescued this, with vaccine groups separated in both cohorts in the CCA analysis, even when proteins were excluded (Appendix Figure 1).

Age-related differences in effect sizes exist despite similar cohort sizes. The amelioration of effects may be due to a chance finding, technical differences between the cohorts (e.g., batch effects), or it could be biological (e.g., age-related, slightly different vaccine combinations, or the effect of multiple exposures to vaccine antigen). All these factors may be at play. Age-related effects could only be isolated by re-running the study with the same vaccine regimes given at both time points and simultaneous sample processing. The possibility of age-related differences is reasonable, given that the immune response evolves over childhood and includes a falling white cell count, reduced granulocyte and monocyte chemotaxis and complement activity, a decline in IL-6 and IL-10 production, increased TNF-alpha and IL-1B production and a switch from a Th2 to a Th1-skewed response[429–434].

Amelioration of the innate immune response to vaccination in the 12-month cohort compared with the 4-month cohort, fits into the paradigm proposed by

Kollman et al. in their review of immune ontogeny in newborns and young infants. In their synthesis of the evidence, Kollman et al. concluded infants have a higher threshold for responding to inflammatory stimuli, which, once breached, unleashes "an immune response of substantial magnitude", leading to higher immune pathology that "partly explain(s) the profound morbidity observed during infections incurred early vs. later life"[436]. A vaccine study designed to chart the evolution of the immune response throughout infancy could explore this theory further, potentially providing insights, which could rationalise therapeutic approaches to serious infection, sepsis, and vaccine development in infancy.

5.6.3 Drivers of vaccine-related fever

5.6.3.1 Age is not a factor in the magnitude of post-vaccination fever

Interestingly, despite the reduced perturbation in the immune response at 12 months, maximum post-vaccine temperatures were similar at 4 and 12 months. This suggests that either the element of the immune response, which drives reactogenicity is equally induced at both ages or the sensitivity of the hypothalamus to pyrogenic signals differs between the two cohorts. It would be useful to study the correlates of fever in another cohort of 4-month-old infants in case these are age-specific correlates.

5.6.3.2 Correlates of reactogenicity and differences between the vaccine groups

At 4 months of age, features across the omics dataset correlated with fever (5 miRNAs and 1 protein), white cell counts (7 miRNAs and 1 protein) and/or were DE between vaccine groups (1 miRNA, 4 proteins and 1 mRNA). None of these associations persisted at 12 months; however, the CCA analysis and PLS-DA signature discriminated between vaccine groups at 4 and 12 months and correlated with fever. The tendency for greater perturbation in the immune response in the 4CMenB group compared with the routine group 24 hours after vaccination mirrors the increased reactogenicity of concomitant 4CMenB. The potential clinical application of these findings is discussed in section 5.6.4. There is a debate about whether reactogenicity and immunogenicity are linked and, if so, whether they could be uncoupled. Unfortunately, sample size precluded testing for associations with vaccine immunogenicity. A larger study could address this.

5.6.3.3 What property of 4CMenB drives its reactogenicity?

The link between co-administration of 4CMenB and increased stimulation of the immune response and reactogenicity may be due to a dose effect as the 4CMenB group received 3 vaccines, whereas the routine group only received 2; or it may be due to a property of 4CMenB. Note in mice, administration of 4CMenB alone led to increased temperature and downregulation of plasma miR-122-5p. LPS in the outer membrane vesicles of 4CMenB is thought to explain its pyrogenicity; however, the RNA-seq results and mouse work in the published EUCLIDS study indicate peptidoglycan in 4CMenB OMVs and the PCV vaccine may also be

responsible[76,77]. This latter point ties in with the positive enrichment of "positive regulation of toll-like receptor 2 pathway" in the miRNAs DE at 4 months.

5.6.4 Clinical application of study results

One application already discussed is using the PLS-DA signature in vaccine development (see section 5.5.1). Another potential application is infection screening. A clinical test that can distinguish vaccine-related fever from serious bacterial infections may abrogate infant hospital admissions and septic screens ascribed to 4CMenB reactions. Having characterised what is expected post-vaccination, it may be possible to identify features—e.g. plasma miRNAs/proteins/mRNAs that are distinctly altered during infection but not vaccination. This could be achieved by mining publicly available data from paediatric infection biomarker studies and using machine learning algorithms to find immune responses unique to infection.

Alternatively, given children are exposed to bacterial antigens in this study (all vaccines were against bacteria), DE features after vaccination could be candidate biomarkers of bacterial infection in situations where clinicians are trying to distinguish viral from bacterial aetiology. This idea is supported by the fact that the validated DE miRNAs in this study are already being investigated as biomarkers of bacterial infection and sepsis, albeit with conflicting results[437]. No published studies looking to differentiate viral and bacterial infections have used a serum/plasma mass spec approach[421]. The mass spec and Luminex results offer a set of proteins that rise in children after exposure to bacterial antigens

and thus could be explored as potential biomarkers of viral and bacterial infections.

5.6.5 Limitations of the EUCLIDS multi-omics study

Some of the limitations of this study have been highlighted above. General limitations are outlined here. The sample size in this study is relatively small; a larger sample size would have provided more power and increased the likelihood of a more representative population. Batch effects are another issue, which means age-specific differences in immune responses between the 4 and 12-month cohorts are confounded by inter-cohort differences in sample processing, assay conduct and data processing (e.g. normalisation). Finally, an important limitation is the lack of a large, genuinely independent cohort to validate the findings of the 4-month cohort. Many participants had samples assayed at both 4 and 12 months. To overcome this, validated pre versus post-vaccination results were assessed in the subset of children with samples only in the 12-month cohort. This genuinely independent cohort was very small and thus was not exclusively used as the validation dataset. Nevertheless, many features were validated in the genuinely independent cohort.

Other factors that could influence results have not been explored; for example, some but not all children received one or more doses of paracetamol. Although paracetamol was not associated with maximum post-vaccination fever, its potential effect on feature expression, especially hepatic miRNAs, should be explored.

Fever was used to measure vaccine reactogenicity because of its objectivity and clinical relevance. This is appropriate because fever alone can drive healthcare attendance and lead clinicians to admit, investigate and treat children for infection, particularly young infants. An alternative approach to assessing reactogenicity would be to use a composite score created using the symptoms scores recorded by participants' guardians in the solicited adverse events diary. If this score captured reactogenicity in a way that is relevant to vaccine licensing (e.g. number of adverse events graded as severe) and/or drives carer or clinical concern, then it would be a clinically relevant readout in the reactogenicity analyses.

Another limitation of EUCLIDS is its conduct in a Caucasian population consisting of participants of similar ages. Biological insights in such a homogenous population cannot be assumed generalisable to other populations. Furthermore, the study set-up differs from that of a clinical setting in, which the timing of presentation and population demographics (e.g. patients' ages, population ancestry) vary, and baseline samples are lacking, which may limit the effectiveness of this study at identifying features that can act as good biomarkers. Nevertheless, despite these limitations, the current study offers a starting point for generating hypotheses, which can then be tested for generalisability to a clinical setting.

5.7 Conclusion

The EUCLIDS multi-omics study adds to the literature on vaccine responses, particularly in infants, who are an understudied population. As the maxim

favoured by paediatricians goes, "children are not small adults", and, arguably, babies are not small children. It is a mistake to assume the development of the immune system in the first year is a linear process that can be interpolated from comparing results in neonates and children over one and adults. Given that by 12 months (inclusive), all UK infants are offered 12 vaccine doses covering 6 different vaccines against 15 diseases, it is imperative that studies are done on this age group. This study has given insights into infants' innate response to vaccination and attempted to link it to vaccine reactogenicity. This appears to be the first dataset to profile infants' immune responses to vaccines in such depth. This is the first study to integrate plasma sRNA-seq, proteomics (including mass spec) and whole blood RNA-seq data to provide a holistic overview of the immune response to vaccination. The challenges and, therefore, scarcity of vaccine studies in infants make this a rare and, thus, valuable dataset.

A challenge in this study was comparing omics data sets generated on different sample types. The following chapters will explore miRNA and mRNA responses in overlapping sample types in two other vaccine studies and will explore whether miRNAs are potential correlates of vaccine immunogenicity and vaccine-attenuated disease severity.

6 sRNA insights into typhoid fever and typhoid vaccines in the context of a human challenge model

6.1 Introduction

Typhoid fever is a potentially life-threatening infection caused by *Salmonella enterica* subspecies *enterica* serovar Typhi (*S.* Typhi). This chapter uses samples from an RCT of typhoid vaccines set within a controlled human infection model of typhoid fever to explore the influence of vaccination and *S*. Typhi exposure on PBMC sRNA expression. A brief background on typhoid is presented below. Appendix 9.5.1 contains a more comprehensive background.

6.1.1 Typhoid

Typhoid and paratyphoid comprise the clinical syndrome enteric fever—an invasive, febrile illness caused by typhoidal *Salmonella*. Typhoid principally affects children and young adults within lower- and middle-income countries and is an important tropical disease in returning travellers[438,439]. In 2017 10.9 million estimated cases led to 116,800 deaths[439,440].

Unlike non-typhoidal salmonellae infections, gastroenteritis is not the predominant clinical problem; instead, the symptoms are often non-specific and flu-like. Typhoid is fatal in 10 to 20% of untreated cases, predominantly from systemic spread leading to encephalopathy, nephritis, hepatitis or intestinal perforation[441]. 2-5% of survivors become long-term carriers, maintaining infection within communities[442,443]. Naturally acquired immunity against *S*. Typhi

wanes, leading to reinfections. Treatment is with antibiotics, the choice depending on local resistance patterns. Third-generation cephalosporins (oral cefixime or parenteral cefotaxime/ceftriaxone) and oral azithromycin are often used; however, carbapenems like meropenem are used in some areas due to extensive multi-drugs resistant strains[444–446].

6.1.2 Typhoid pathogenesis

After ingestion, *S.* Typhi that survive the stomach's acidic environment, pass through the small bowel to the terminal ileum. *S.* Typhi crosses the gut wall through M cells, specialised antigen sampling epithelial cells overlying Peyer's patch in the terminal ileum[447]. Take-up may also occur via enterocytes. Once inside the cell, *S.* Typhi subverts host immune responses by expressing its Vi-polysaccharide capsule, which allows it to go undetected by PRRs and shields it from complement activation[448]. After passing through enterocytes and M cells, *S.* Typhi is taken up by monocytes, where it survives within phagosomes and is systemically disseminated through the reticuloendothelial system[447]. Human challenge models indicate primary bacteraemia can occur 12 hours post-ingestion of *S.* Typhi; however, it does not inevitably lead to typhoid, indicating immune responses can neutralise *S.* Typhi after systemic spread[449]. Dissemination of *S.* Typhi to the gall bladder enables re-entry into the gut and stool shedding. Failure to irradicate *S.* Typhi from the biliary system results in long-term carriage and shedding.

6.1.3 Preventing typhoid

Typhoid is spread by the faecal-oral route. Safe drinking water and sanitation are the mainstay of prevention but are not universally available. Vaccines offer another layer of protection. Until recently, two types of typhoid vaccines were in use; the oral live-attenuated typhoid vaccine Ty21a and the plain Vi polysaccharide vaccine (Vi-PS). These vaccines have limited efficacy and are unsuitable for young children (see sections 1.2.1.2.3.1-1.2.1.2.3.2). A typhoid conjugate vaccine (Vi-TCV) was therefore developed, extending typhoid vaccine coverage to infants. Conjugation of Vi-polysaccharide to tetanus toxoid protein in Vi-TCV elicits T-cell dependent responses, e.g., memory B-cells and high-affinity antibodies, making the vaccine immunogenic in children under 2. The location of vaccine-derived immunity is unknown. Vaccine-mediated immunity could protect at different stages of *S.* Typhi invasion and infection. For example, secreted IgA could prevent gut wall invasion, or antibodies could clear cell-free *S.* Typhi in interstitial tissues, the bloodstream or end organs. Vi-vaccines are unlikely to prevent entry into gut mucosa, given the Vi-capsule is not expressed within the gut lumen. The T-cell independent immune mechanisms of Vi-PS indicate these vaccines target an extracellular stage, possibly after *S*. Typhi traverses the gut lumen when S. Typhi expresses Vi and is accessible to antibodies, but more information is needed.

6.1.4 Challenges in typhoid

The non-specific presentation of typhoid and limited blood culture sensitivity make diagnosis difficult[450]. This can result in misdiagnosis, inappropriate antibiotic selection, treatment delay, and hinders epidemiological and interventional research. Host transcriptomics and metabolomics signatures can

identify *S.* Typhi infection, indicating evaluation of host response offers a novel diagnostic approach[451,452].

Antibiotic resistance is a serious concern in typhoid[453,454]. Novel treatments that switch host responses towards *S*. Typhi eradication would be useful, potentially limiting morbidity and mortality without conferring the selective pressure of antibiotics. Such an approach requires better understanding of host responses to S. Typhi.

Typhoid vaccines only offer partial protection against *S.* Typhi infection. Vaccine failure is poorly understood. Identifying correlates of protection and unpicking the mechanisms underlying vaccine failure could inform future vaccine design and possibly enable personalised or population-based vaccine strategies.

Overcoming the above challenges requires better characterisation and understanding of typhoid pathogenesis, heterogeneity in host responses, and the factors that promote or undermine vaccine effectiveness.

6.1.5 Vaccines Against *Salmonella* Typhi (VAST) study: an ideal model to dissect vaccine-mediated immunity and host response to *S.* Typhi.

Field studies of typhoid pathogenesis are marred by heterogeneity arising from differences in when patients present, *S.* Typhi strain, and previous *S.* Typhi infections. Uncertainty about exposure in controls who have never developed typhoid impedes efforts to identify immune factors protecting against disease.

Human challenge studies circumvent these issues, providing a controlled environment to chart the disease from exposure through to diagnosis and evaluate vaccine efficacy. The Vaccines Against Typhoid study (VAST) was a Phase IIb, observer-blind, randomised controlled trial to assess the immunogenicity and protective efficacy of Vi-TCV and Vi-PS in preventing typhoid compared to a control vaccine in a human *S*. Typhi challenge model (ClinicalTrials.gov Identifier: NCT02324751). In this study, participants were randomly assigned to receive menACWY (the control vaccine), Vi-PS or Vi-TCV 28 days before oral challenge with *S.* Typhi. Participants were followed up for 14 days after the challenge, and typhoid incidence compared between the groups (see methods section 6.3.1.4). See Figure 65 for trial schema. Multiple assays and datasets were run/generated in VAST after vaccination, challenge and during acute typhoid (Figure 65).

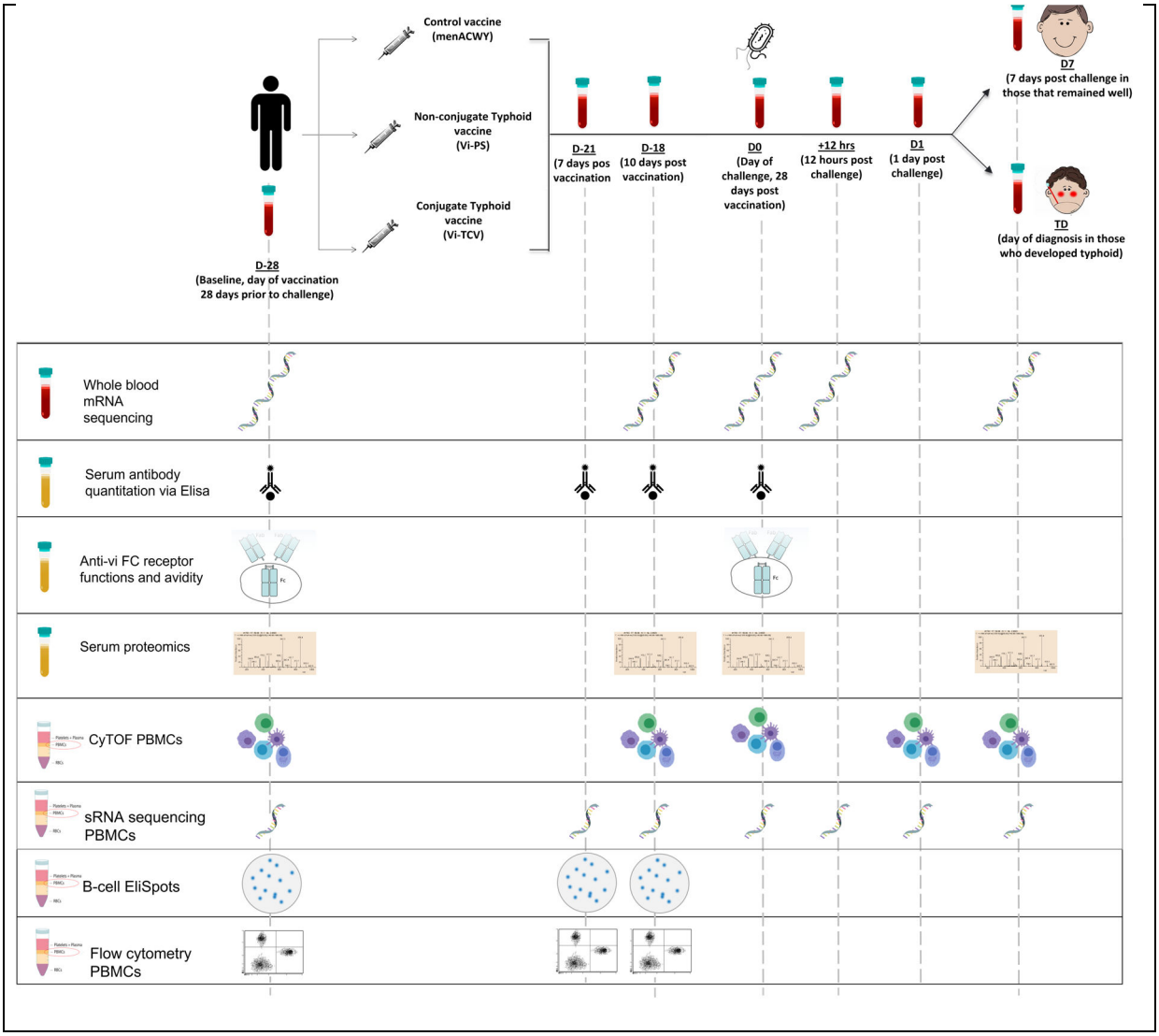


Figure 65 VAST study schema. Abbreviated sample visit names. Sampling time points indicated by blood tubes. Note time points are reported relative to day of challenge—denoted day 0 (D0), i.e. D-28, D-21, D-18, D0, D+12hrs, D1, D7, TD, respectively. Assays performed at each time point are indicated below the timeline.

6.1.6 Key findings from VAST

VAST found similar efficacies for Vi-TCV and Vi-PS (54·6%; 95% CI 26·8–71·8 and 52·0%; 95% CI 23·2-70·0, respectively)[40]. Key results included increased IgA and IgG Vi titres and Vi-specific plasma B-cells after vaccination; in all cases, increases were higher after Vi-TCV than Vi-PS. Though protection from typhoid was associated with IgA and IgG titres in Vi-vaccinees, a predictive

threshold for protection could not be identified[455]. A machine learning model identified slightly different correlates of protection for Vi-PS and Vi-TCV. Protection in Vi-PS vaccinees was positively associated with serum IgG, IgA, IgM, polyfunctional antibodies and antibody-dependent cell phagocytosis (ADCP) and negatively associated with antibody-dependent NK cell activity (ADNKA) and antibody binding to the immunoglobulin Fc receptor FCYR2B. Protection in Vi-TCV vaccinees positively correlated with IgA quantity and avidity and IgG1 avidity and was negatively associated with ADNKA, ADCP, and IgG1 titres[455]. *S.* Typhi challenge had a profound effect on PBMC cell subsets. A signature associated with protection identified higher baseline frequencies of naïve B-cell and CD16+ monocytes, whilst typhoid susceptibility was associated with higher baseline terminally differentiated T-cells and IgA+ and IgG+ B memory cells with homing markers for lymph nodes[456]. Gene expression studies identified B-cell receptor sequences in whole blood 7 days after vaccination, which correlated with Vi titres and protection (unpublished). Appendix Table 6 contains other VAST findings.

6.1.7 VAST: a unique opportunity to describe vaccine-mediated immunity and host response to *S.* Typhi at the sRNA level.

No studies have investigated miRNAs in typhoid. VAST represents an excellent opportunity to pursue the role of sRNAs in vaccine-mediated protection, bacterial exposure and infection. The wealth of paired datasets in VAST (Figure 65) offers a unique opportunity to pursue how miRNAs relate to cell counts, mRNA expression and immunological outcomes, e.g., antibody titres, to dissect the roles of miRNAs in immunity. Only PBMC samples from VAST were available for

sRNA-seq. See Figure 65 for sRNA-seq timepoints. The questions this study can answer are summarised in Figure 66.

6.1.1 VAST sRNA study set-up

sRNA-seq was done in two stages (Table 20). The first stage looked at post-vaccination and post-challenge samples from the Vi-vaccinated arms of VAST. This was because interest lay in how conjugation influences the immune response to the same antigen (Vi in this case) and identification of correlates of vaccine-mediated protection. Post-challenge samples were assessed in the Vi-vaccinees 24 hours after challenge (D1), at diagnosis (TD) in those that were diagnosed, and 7 days after challenge (D7) in those that did not develop typhoid. D7 was chosen as a comparison time point with TD because median day of diagnosis was 7 days post-challenge. A pilot run showed significant results were more pronounced post-challenge vs post-vaccination; therefore, fewer samples were sequenced.

Positive findings in stage I led to stage II of the sRNA-seq project, which looked at the effect of challenge 12 hours after vaccination and explored the effects of challenge and typhoid diagnosis in control vaccinated individuals. Therefore 12-hour samples from all vaccine groups were sequenced along with repeat D0 and D1 samples in Vi-vaccinees so 12-hour samples could be compared with D0 and D1 without batch effect. In addition, samples from control vaccinees at TD and D7 were also run. Control vaccinees sequenced in stage II provided an independent cohort to validate post-challenge results seen in stage I and allowed

exploration of whether early post-challenge differences between diagnosed and non-diagnosed participants were vaccine-related phenomena.

Table 20 Sampled time points broken down by vaccine arm in stage I and stage 2 of the sRNA-seq study

	Stage I		**Stage II**		
	Vi-PS	Vi-TCV	Control	Vi-PS	Vi-TCV
Day -28 (Baseline)	✓	✓	-	-	-
Day -21 (7 days post-vaccine)	✓	✓	-	-	-
Day -18 (10 days post-vaccine)	✓	✓	-	-	-
Day 0 (challenge) (28 days post-vaccine)	✓	✓	✓	✓	✓
D0 +12 hrs	-	-	✓	✓	✓
D1	✓	✓	✓	✓	✓
Typhoid diagnosis	✓	✓	-	-	✓
D7 (in people who did not develop Typhoid)	✓	✓	-	-	✓

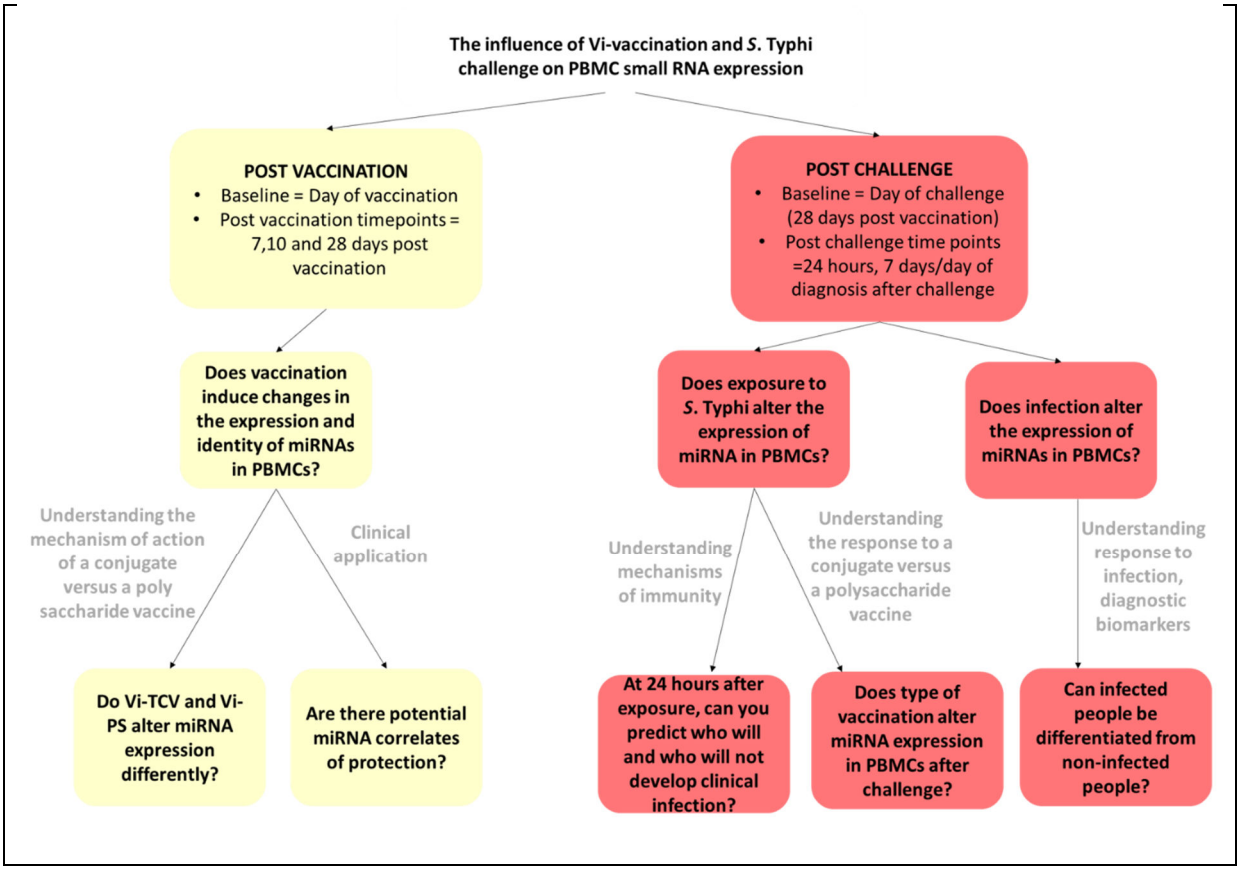


Figure 66 How study time points can be exploited to answer questions about vaccination and the host response to S. Typhi.

6.2 Chapter aims

Chapter aims are to establish whether:

1. Vaccination against Vi influences expression of miRNAs and sRNA-dfs in PBMCs
2. miRNA or sRNA-dfs correlate with vaccine-mediated protection
3. *S.* Typhi influences miRNA and sRNA-df expression in PBMCs shortly after exposure and during typhoid, and whether vaccination influences this

6.3 Methods

6.3.1 VAST study details

6.3.1.1 Study population

Study participants were healthy adults between 18 and 60 years old who were naïve to typhoid vaccination and enteric fever. Participants with gallstones and those on anti-reflux medications were excluded.

6.3.1.2 Vaccines

Participants were randomised to receive Vi-TCV (Typbar-TCV, Bharat Biotech, Hyderabad, India), Vi-PS (TYPHIM Vi, Sanofi Pasteur, Lyon, France), or the control vaccine, meningococcal ACWY-CRM conjugate vaccine (MENVEO; GlaxoSmithKline, Sovicille, Italy). Both Vi vaccines contain 25μg of purified Vi capsular polysaccharide (derived from *S.* Typhi Ty2 strain), and neither is adjuvanted.

6.3.1.3 S. Typhi challenge

Twenty-eight days after vaccination, participants consumed sodium bicarbonate solution to neutralise stomach acid and then drank a dose of *S. Typhi,* conferring a predicted attack rate of 60–75% in control vaccinees. A dose that causes typhoid in 100% of control participants could not be used because it risks severe illness in highly susceptible participants.

6.3.1.4 Post-challenge follow-up and definition of typhoid diagnosis

Participants were followed up with daily clinical review, full blood count, CRP and blood cultures. Typhoid diagnosis was defined as a fever of >38ºC for ≥12 hours or a positive blood culture ≥72 hours after challenge (or within 72 hours if there was an associated fever). Participants were commenced on antibiotics upon diagnosis or at day 14 post-challenge in non-diagnosed participants. The study was participant and observer-blinded.

6.3.1.5 PBMC samples

PBMCs were collected in sodium heparin tubes, isolated using lymphoprep, and stored in liquid nitrogen.

6.3.1.6 Sample selection

Samples were sequenced from participants with at least two vials of PBMCs at all or most time points.

6.3.2 sRNA sequencing

6.3.2.1 RNA extraction

PBMCs were thawed quickly in lukewarm water. 2.5 million cells were washed once with 3 volumes of phosphate-buffered saline. Total RNA was extracted using Norgen's Total RNA purification kit (catalogue number 17200), including a DNase treatment (RNase-Free DNase I Kit, Norgen, catalogue number 25710). RINe values measured on a subset of samples showed adequate values between 6.8 and 7.5.

6.3.3 Data processing

See section 2.5.2

6.4 Results

6.4.1 A note on terminology

For brevity, participants who remained well after challenge are called non-typhoid-diagnosed participants (nTD), and those who developed typhoid are called diagnosed participants (TD). Phrases like "samples taken 10 days after vaccination in TD participants" should be thus read as: "samples taken 10 days after vaccination in participants *who went on to develop typhoid after S. Typhi challenge"*.

6.4.2 Demographics

Samples from 101 people were sequenced (Table 21 and Figure 67), of which 49 (71% of Vi-vaccinees) had all pre-challenge samples sequenced, and 9 (10% of all participants sequenced) had every time point sequenced. Appendix Table 7 contains participant-level information.

Table 21 The numbers of participant samples subject to sRNA-seq broken down by vaccine group and challenge outcome. Numbers in brackets show how many Vi-vaccinees were re-sequenced in stage II.

Time point	Total number of samples	Outcome after challenge					
		Diagnosed			Not diagnosed		
		Vi-PS	Vi-TCV	Control	Vi-PS	Vi-TCV	control
D-28	67	12	12	-	21	22	-
D-21	66	11	13	-	21	21	-
D-18	58	10	10	-	19	19	-
D0	60	13 (6)	13 (3)	23	16 (7)	16 (9)	6
12 hrs		6	2	11	7	8	4
D1	48	11 (5)	9 (1)	19	13 (4)	13 (2)	5
D7	27	-	-	-	14	13	6
TD	26	14	12	12	-	-	-

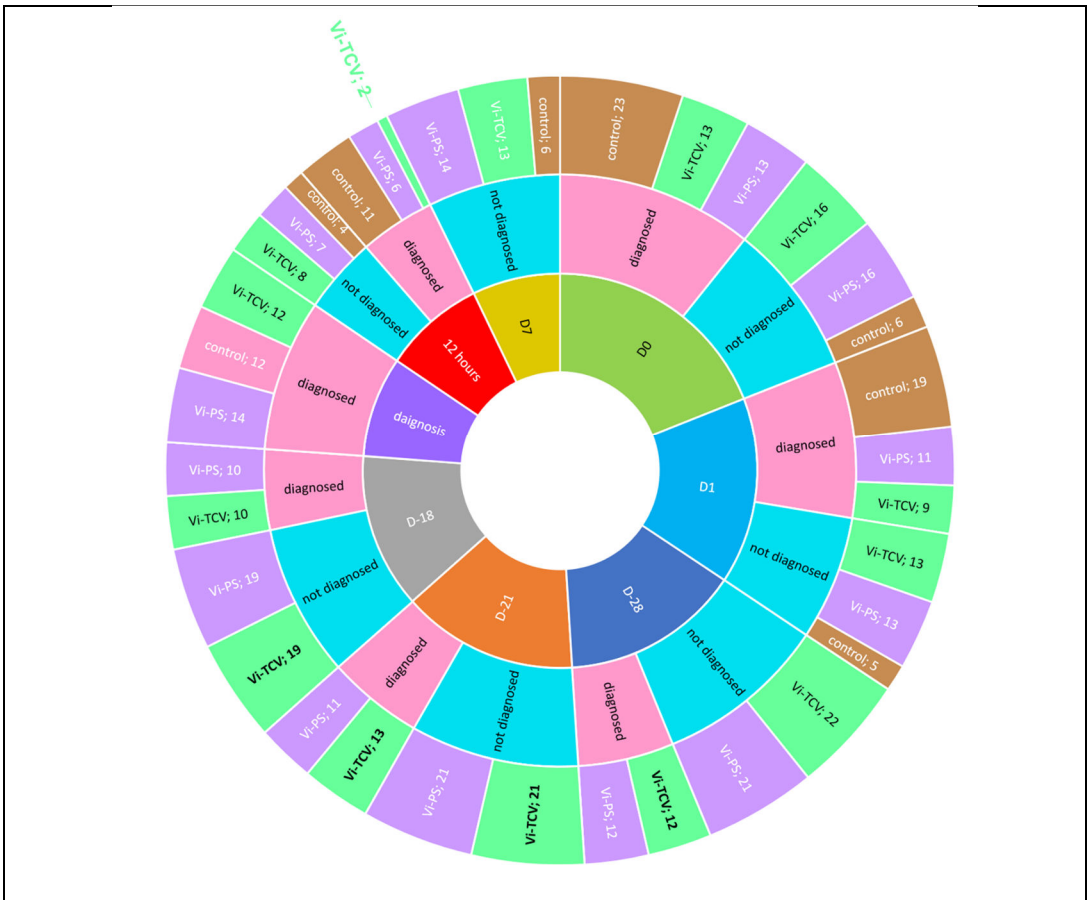


Figure 67 Starburst plot showing the number of samples broken down by vaccine group and outcome. This plot makes it visually easier to compare the relative sample sizes of subgroups.

6.4.3 sRNA composition of PBMC samples

The number of unique features analysed per RNA class (i.e. retained after noise removal) is contained in

Table 22

Table 22 Number of unique features analysed per RNA class (i.e. retained after noise removal)

RNA class	Number of features
mature miRNA	310
lncRNA	858
offset miRNA	66
protein-coding	723
rRNA	7
snoRNA	237
snRNA	19
tRNA	81
vault	3
Antisense to known features	358
yRNA	3
Total features	**2,665**

The QC plots and PCA for these samples can be found in Appendix Figure 30. One sample failed QC due to its small library size (1/1000th of the median) and was removed.

Figure 68 shows relative sRNA class-wise abundances at each time point. The most abundant sRNA-dfs were mature miRNAs, snoRNAs, yRNAs, and tRNA derivatives. There were significant differences in class-wise sRNA expression between time points (Appendix Figure 31). As miRNAs are the focus of this thesis, miRNAs were normalised and analysed separately from other sRNA classes to prevent class-wise differences influencing miRNA differential

expression analysis. Vaccine type did not influence global expression of sRNA classes.

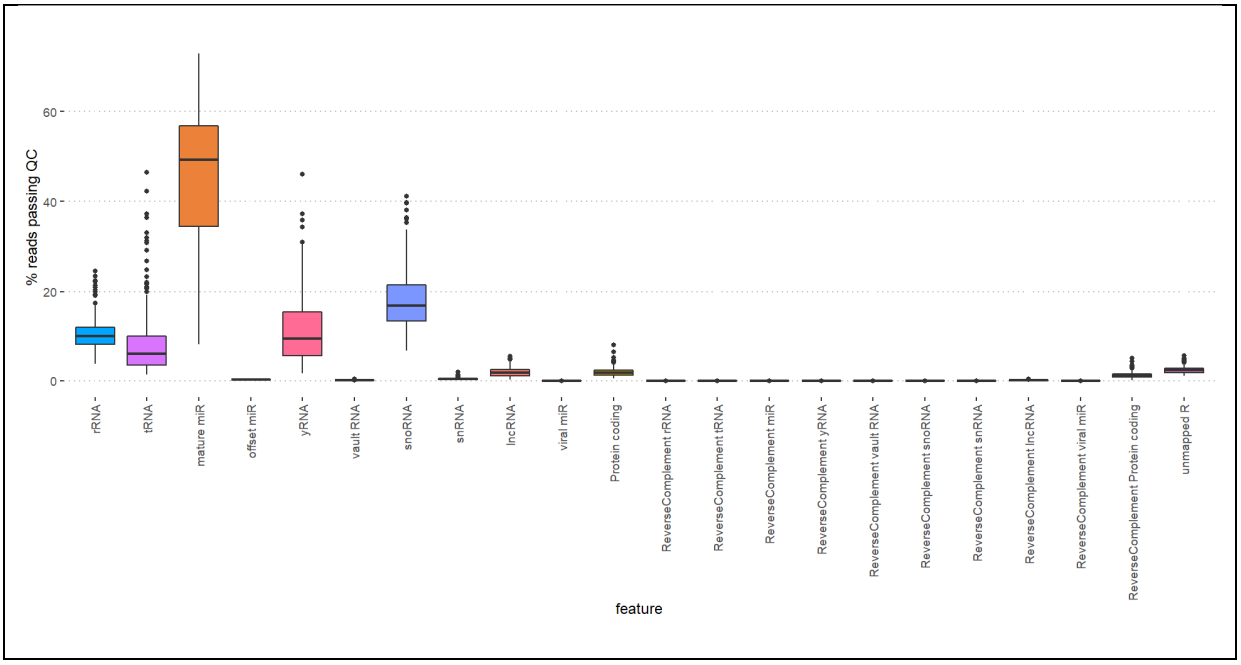


Figure 68 Relative abundance (by number of reads) of each class of sRNA identified in the sRNA-seq data.

Results are presented according to two themes:

- Theme 1: vaccine-induced changes in sRNA expression. This theme relates to pre-challenge samples.
- Theme 2: *S.* Typhi exposure and infection-induced changes in sRNA. This theme relates to post-challenge samples.

In both themes, samples were additionally analysed by vaccine group and challenge outcome in case vaccine type and response (i.e., protective/non-protective) influences sRNA expression.

6.4.4 Theme 1: Does typhoid vaccination influence PBMC sRNA expression?

6.4.4.1 Does vaccination influence sRNA expression over time?

A longitudinal analysis explored whether vaccination influences sRNA expression. Few (0 to 4) miRNAs were DE over time (Figure 69, Table 23, Supplementary Table 6). There is no independent validation cohort to re-test these associations. The only similar dataset is a published flu study. Cross comparison of the results of that study with this study revealed evidence of a weak generalisable effect of vaccination on miRNA expression, possibly more pronounced in those with an effective vaccine response but no DE miRNAs after Vi-vaccination were also DE in the week following flu vaccination (see Appendix 9.5.2).

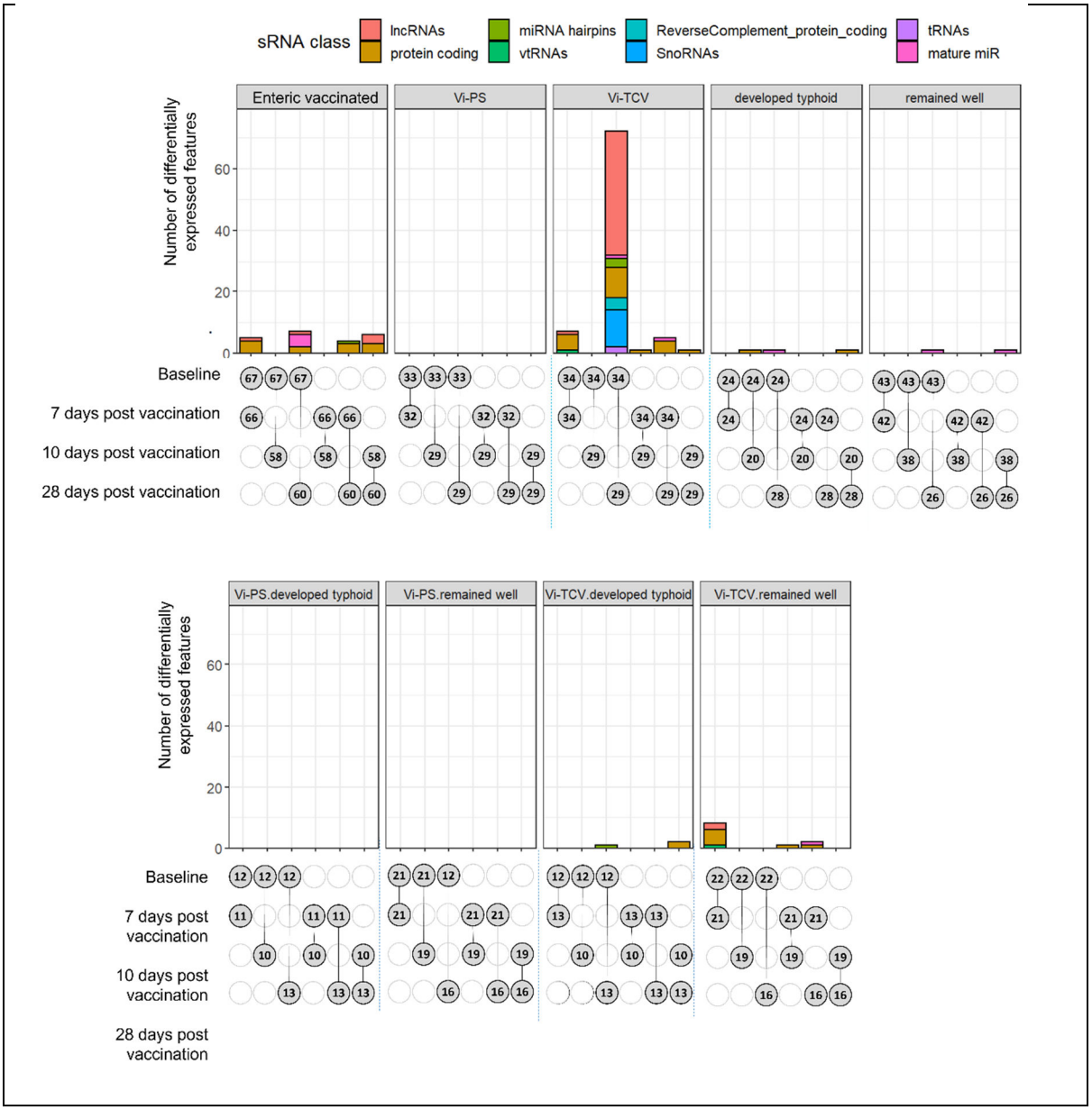


Figure 69 Overview of the number of DE sRNAs between time points. Number of samples in each analysis is indicated in the circles underneath each plot. Care should be taken when comparing the number of DE features due to differing group sizes. Supplementary Table 6 contains the number of DE features per sRNA class.

Table 23 DE miRNAs between time points

Timepoint 1	**vs**	**Timepoint 2**	**Subgroup for non-baseline samples (for baseline, all samples were used)**	**miRNA**	**Ave Exp (CPM)**	**Log_2 FC**	**p-value**	**FDR**
Baseline	**vs**	28-days post-vaccine	All participants	hsa-miR-92a-3p	14.3	0.2	0.00008	0.01
				hsa-miR-423-5p	11.2	0.3	0.00008	0.01
				hsa-miR-320a-3p	11.4	0.2	0.00008	0.01
				hsa-miR-1301-3p	7.1	0.3	0.00020	0.02
			Vi-TCV	hsa-miR-423-5p	11.2	0.5	0.00005	0.01
			Remained well	hsa-miR-92a-3p	14.3	0.2	0.00011	0.03
			Developed typhoid	hsa-miR-409-5p	3.5	0.5	0.00014	0.04
7-days post-vaccination	**vs**	28-days post-vaccine	Vi-TCV	hsa-miR-148a-5p	5.6	-0.5	0.00005	0.01
7-days post-vaccination	**vs**	28-days post-vaccine	Vi-TCV + remained well	hsa-miR-148a-5p	5.6	-0.6	0.00016	0.05
10-days post-vaccination	**vs**	28-days post-vaccine	Remained well	hsa-miR-7977	6.6	0.6	0.00002	0.01

6.4.4.1.1 sRNA-dfs are differentially expressed 28 days after VI-TCV.

The clearest signal of DE sRNA-dfs is baseline versus 28 days post-vaccination in Vi-TCV vaccinees (Figure 69, Supplementary Table 6). In this comparison, 72 sRNAs across multiple classes were differentially expressed, with most changes occurring in lncRNA-dfs (40/858 features DE), followed by mRNA-dfs (10/723 features DE) and snoRNA-dfs (12/237 features DE). Changes were independent of vaccine effectiveness as they were best seen when outcome groups were combined. ORA of these features did not identify enriched pathways.

6.4.4.1.2 sRNAs derived from immunoglobulin genes are upregulated 7 days after vaccination after Vi-TCV but not Vi-PS.

Interestingly, sRNA-dfs derived from genes encoding the constant region of immunoglobulin genes were upregulated 7 days after vaccination but only after Vi-TCV. These sRNA-dfs map to immunoglobulin kappa constant (IGKC), immunoglobulin heavy constant gamma 1 (IGHG1) and joining chain of multimeric IgA and IgM (JCHAIN) (Figure 70). These sRNA-dfs may relate to B-cell responses to tetanus toxoid or Vi (or both) contained within Vi-TCV. Evidence they come from Vi-specific plasma cells include 1) IGKC, JCHAIN, and IGHA1 levels 7 and 10 days after vaccination associated with protection after challenge using a relaxed threshold of $p<0.05$ (Appendix Figure 33), and 2) these sRNA-dfs correlate with Vi-specific B-cells 7 days after vaccination (Figure 70B). Despite correlating with B-cells, these sRNA-dfs did not correlate with Vi-antibody levels 28 days after vaccination. Baseline differences of these sRNAs were associated with susceptibility to typhoid in the Vi-PS group (and persisted 7

days after vaccination), whereas this was not the case in the Vi-TCV group, potentially indicating people who respond best to a Vi-TCV vaccine differ from those who respond best to a Vi-PS vaccine (Appendix Figure 33).

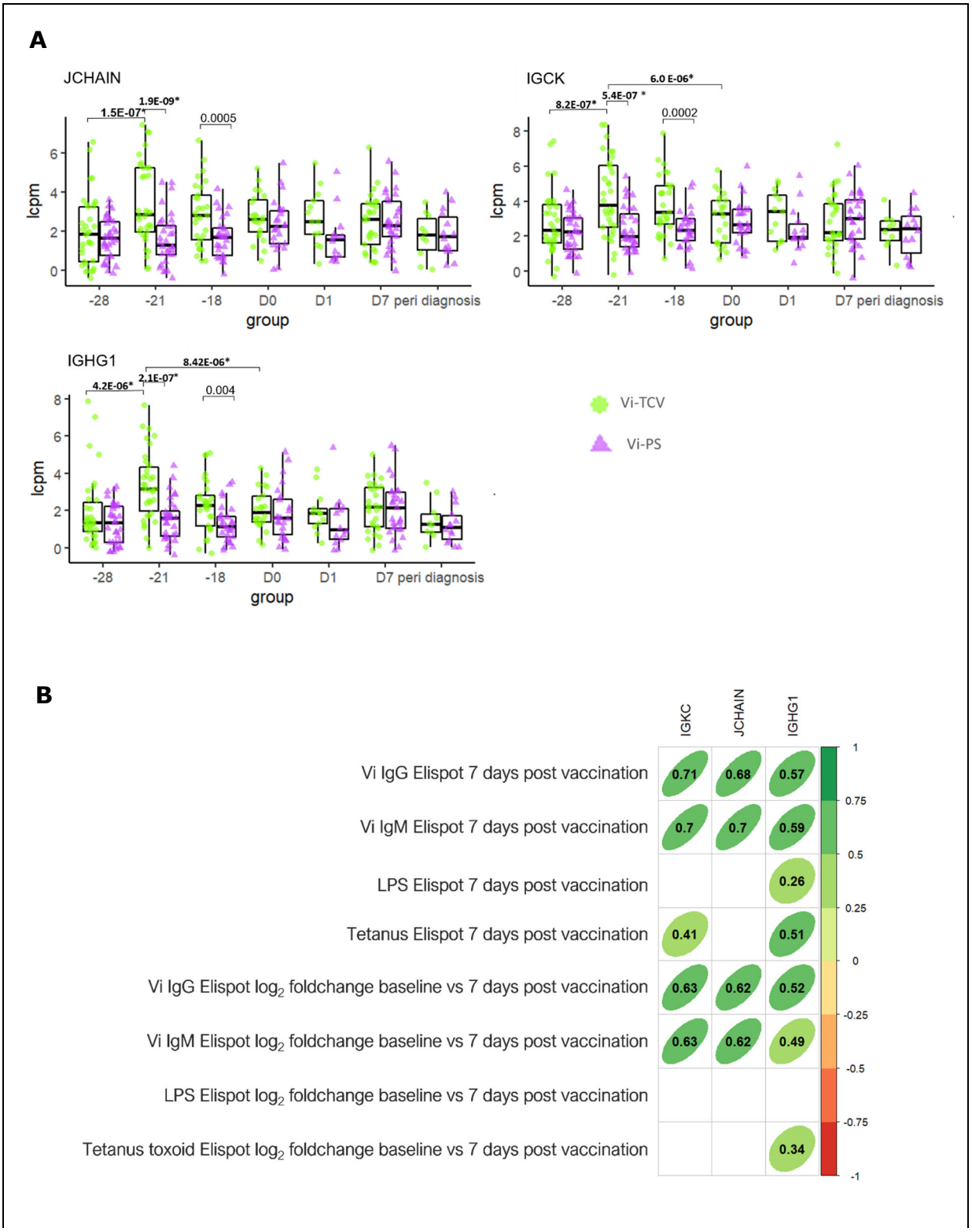


Figure 70 Immunoglobulin gene sRNA-dfs upregulated at 7 days after vaccination.
A: Dot plots showing the expression of immunoglobulin gene sRNA-dfs upregulated at 7 days after vaccination in the VI-TCV arm. Results with p<0.05 are highlighted. FDR significant p-values are in bold and starred.*
B: Correlogram showing correlations between immunoglobulin gene sRNA-dfs and plasma cell Elispots at 7 days after vaccination. Only significant correlations are reported.

6.4.4.2 Does vaccine type influence sRNA expression?

A cross-sectional analysis explored whether vaccine type influences sRNA expression (Figure 71, Supplementary Table 6). Reassuringly, given this was an RCT, there were no baseline differences between vaccine groups.

6.4.4.2.1 Vi-TCV induces changes in immunoglobulin genes compared with Vi-PS

Nine sRNAs were DE between Vi-PS and Vi-TCV 7 days after vaccination. Four of these sRNAs related to immunoglobulin genes (JCHAIN, IGCK, IGHG1 and IGHA1)—all upregulated after Vi-TCV compared with Vi-PS. Differences between Vi-PS and Vi-TCV were most prominent in protected groups. These differences arise from divergent associations in immunoglobulin sRNA-dfs and outcome between the vaccine groups—e.g., high JCHAIN-df expression 7 days after vaccination was associated with protection in Vi-TCV vaccinees but susceptibility in Vi-PS vaccinees (at a relaxed significance threshold of $p<0.05$) see Appendix Figure 33.

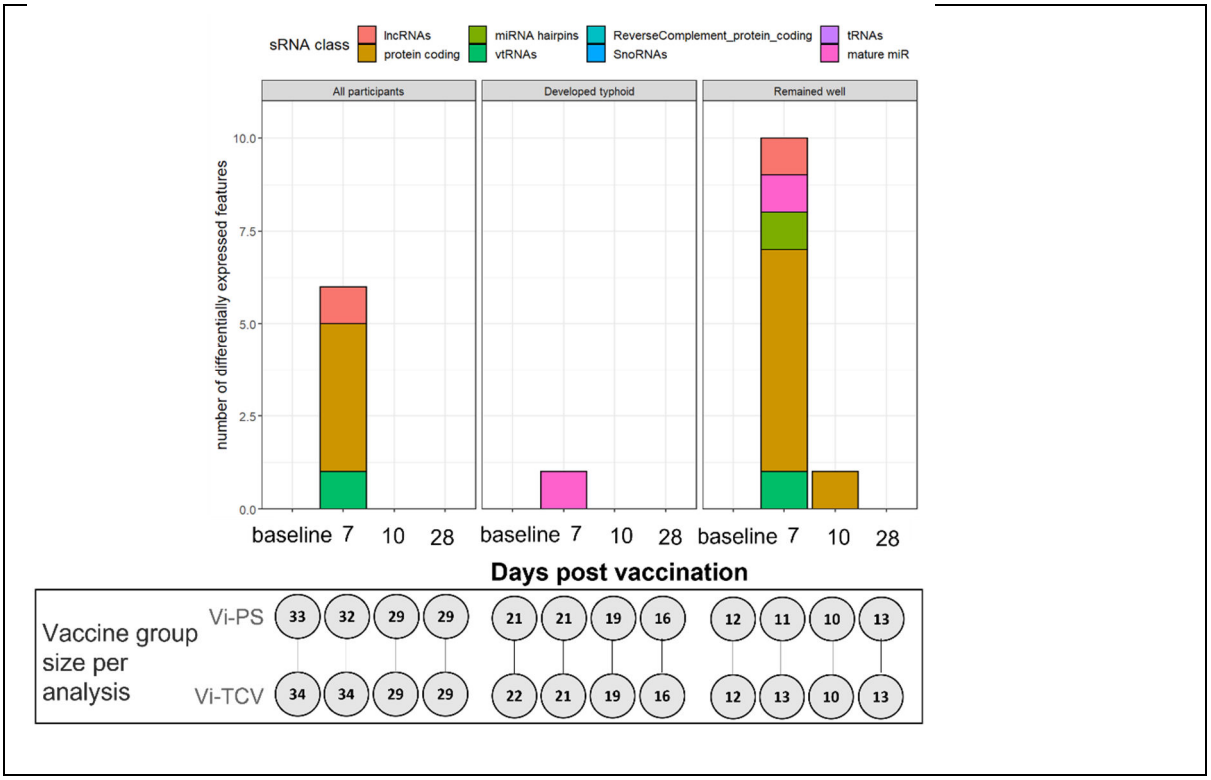


Figure 71 Bar plot showing the number of DE sRNAs between Vi-vaccine groups at each time point. Results broken down by challenge outcome. Number of samples in each analysis is indicated in circles underneath each bar. Care should be taken when comparing the number of DE features due to differing group sizes. Supplementary Table 6 contains the number of DE features per sRNA class.

6.4.4.3 Are there vaccine based correlates of protection?

Differential expression results between outcome groups are summarised in Figure 72.

6.4.4.3.1 Divergent miRNA expression 10 days after vaccination was associated with outcome of *S.* Typhi challenge.

The strongest differences between outcome groups occurred 10 days after vaccination and were mainly miRNAs (Figure 72C). PCA analysis of miRNA expression 10 days post-vaccine shows that centroids of each outcome group

cluster separately (Figure 72B). PC1 score was significantly associated with outcome (Wilcox p-value = 0.0003, Appendix Figure 36). This pattern was not seen at baseline (Appendix Figure 34). In the PCA, a group of participants who remained well overlapped with those that developed typhoid (Figure 72B). This pattern is remarkably similar to the pattern seen in Appendix Figure 35D, which predicts how a PCA would appear if vaccine-mediated correlates of protection drive sample separation. The pattern is independent of vaccine type and is not associated with other proposed correlates of vaccine responses related to challenge outcome in VAST, e.g., Vi-antibody titres taken 28 days post-vaccination. This suggests the miRNA signal 10 days post-vaccination may provide additional/orthogonal information on vaccine-mediated correlates of protection to cellular and serological ones previously identified in VAST (see Appendix Table 6).

miRNA differences between outcome groups were strongest when vaccine groups were combined, indicating outcome-related differences are independent of vaccine type. Furthermore, Vi-TCV and Vi-PS foldchanges positively correlate (Figure 72A, Figure 73), and this correlation is further strengthened amongst sRNAs with p-values lower than 0.05 in both analyses (Pearson correlation, r—0.96, p<2.7 e-18). An anti-correlation between outcome groups in terms of their $\log_2$ foldchanges at baseline vs 10 post-vaccination (Figure 74) indicates outcome groups have opposite miRNA responses to vaccination., i.e. miRNAs with a trend to *upregulation* in diagnosed individuals were *downregulated* in those who did not develop typhoid and vice versa (Appendix Figure 38).

A sparse PLS-DA model was used to predict outcome after challenge based on sRNA expression patterns 10 days post-vaccination. Appendix 9.5.1 and Appendix Figure 39 contain details of model development. The model showed moderately good performance in the training data with a balanced error rate (BER) of 0.35 and accuracy of 76% but performed less well on the test data with a BER of 0.31 and an accuracy of 67%. The inclusion of Vi-antibody titres in the model did not improve the result. Poor model performance is not surprising given some participants may have had a poor vaccine response yet were protected from typhoid due to intrinsic resistance to *S.* Typhi (see Appendix 9.5.3: The challenge of identifying correlates of protection in VAST because of participants who are inherently resistant to typhoid.).

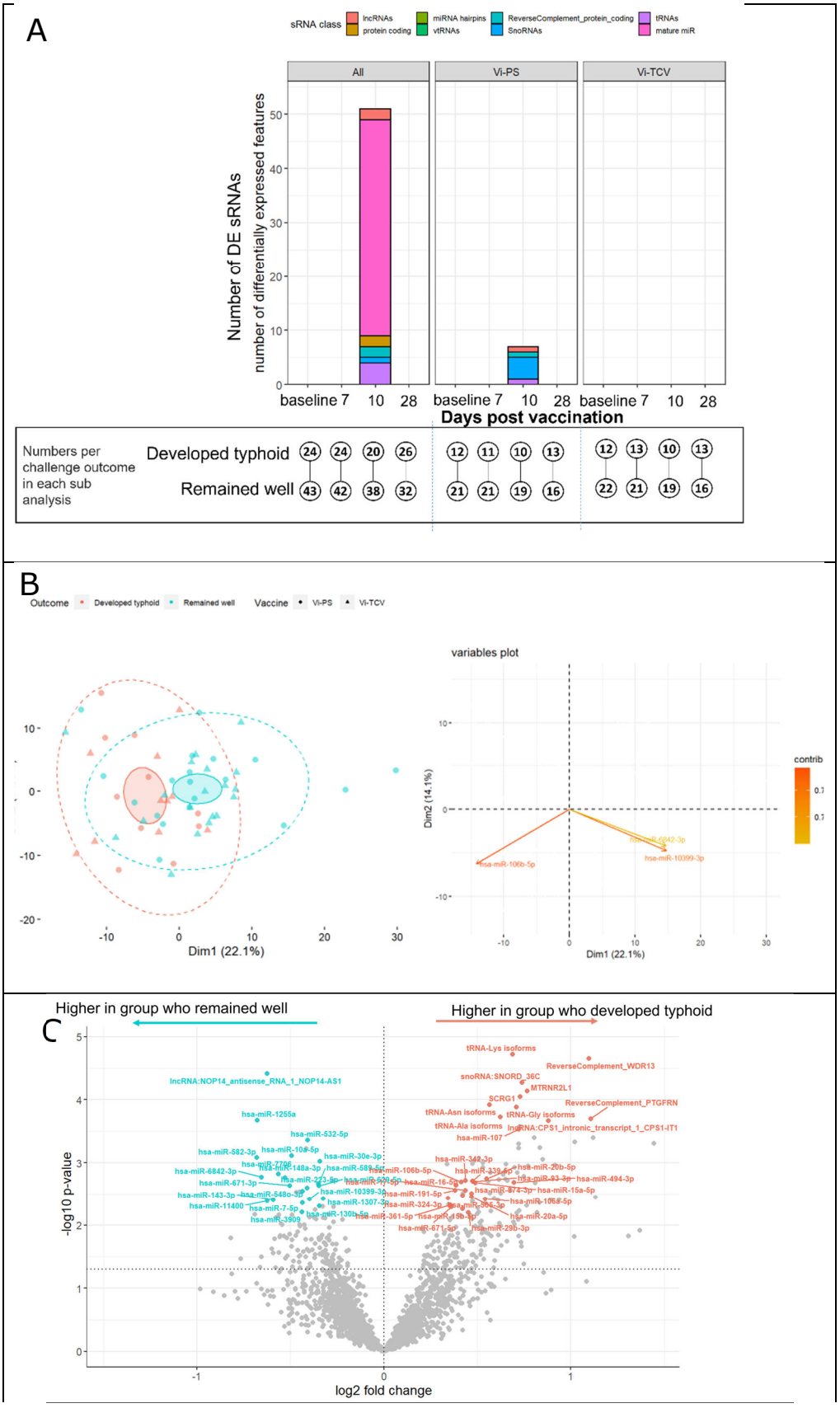


Figure 72 Identifying correlates of protection

A: Bar plot showing the number of DE sRNAs between outcome groups at each time point. Results presented broken down by challenge outcome. Number of samples in

Figure 72 caption (continued)
each analysis is indicated in circles underneath each bar. Care should be taken when comparing the number of DE features as group sizes vary. Supplementary Table 6 contains the number of DE features per sRNA class.
B: PCA of samples taken 10 days after vaccination, coloured by outcome after S. Typhi challenge. The outer dotted lines represent 95% confidence intervals of the data. The shaded ovals represent 95% confidence intervals of the centroids.
C: Volcano plot showing results of differential expression analysis comparing day 10 post-vaccination samples from those that did and did not develop typhoid after challenge. Coloured dots indicate FDR significant.

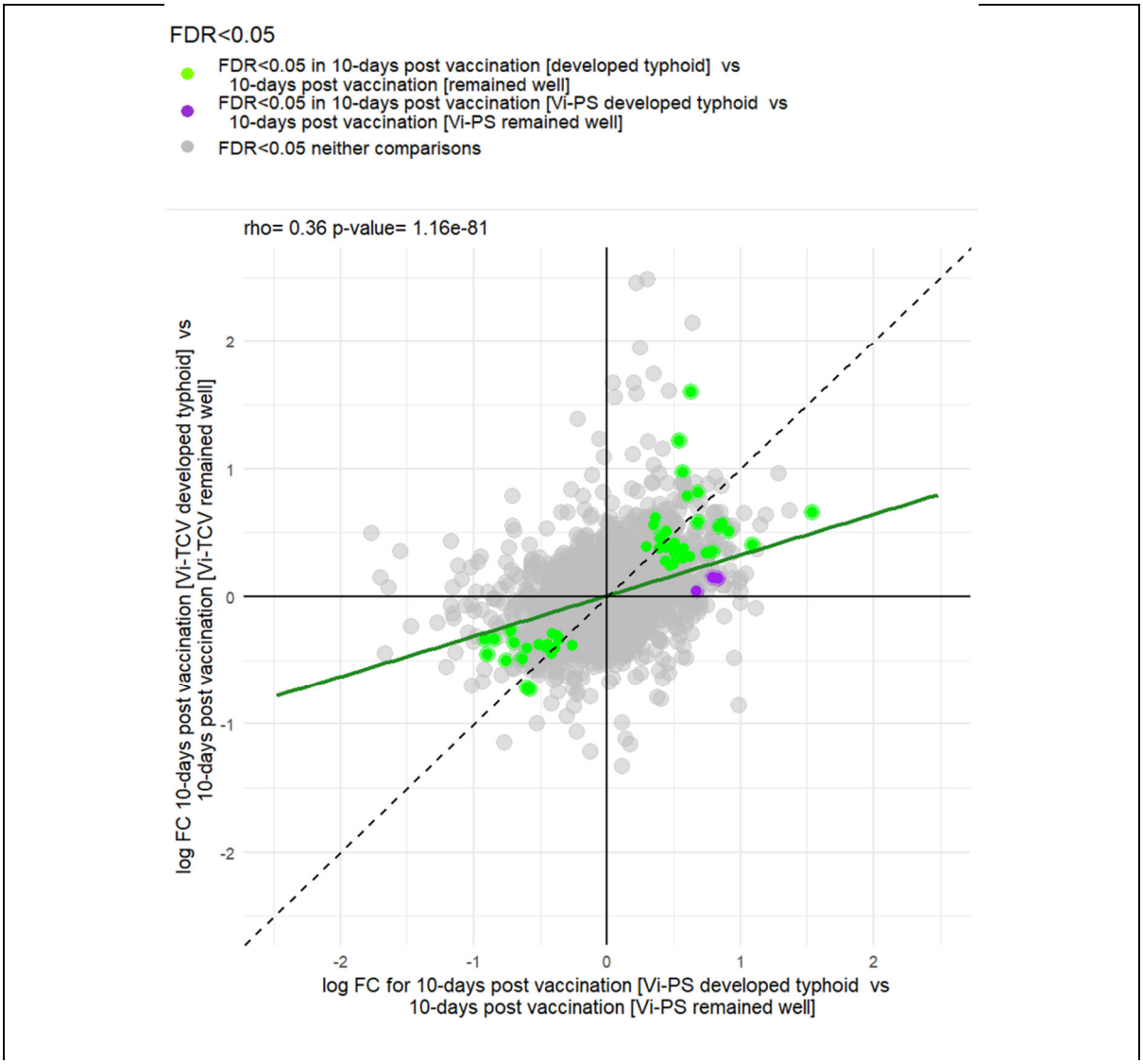


Figure 73 Agreement plot comparing log_2 foldchanges between outcome groups in Vi-PS vaccinees (x-axis) and Vi-TCV vaccinees (y-axis).

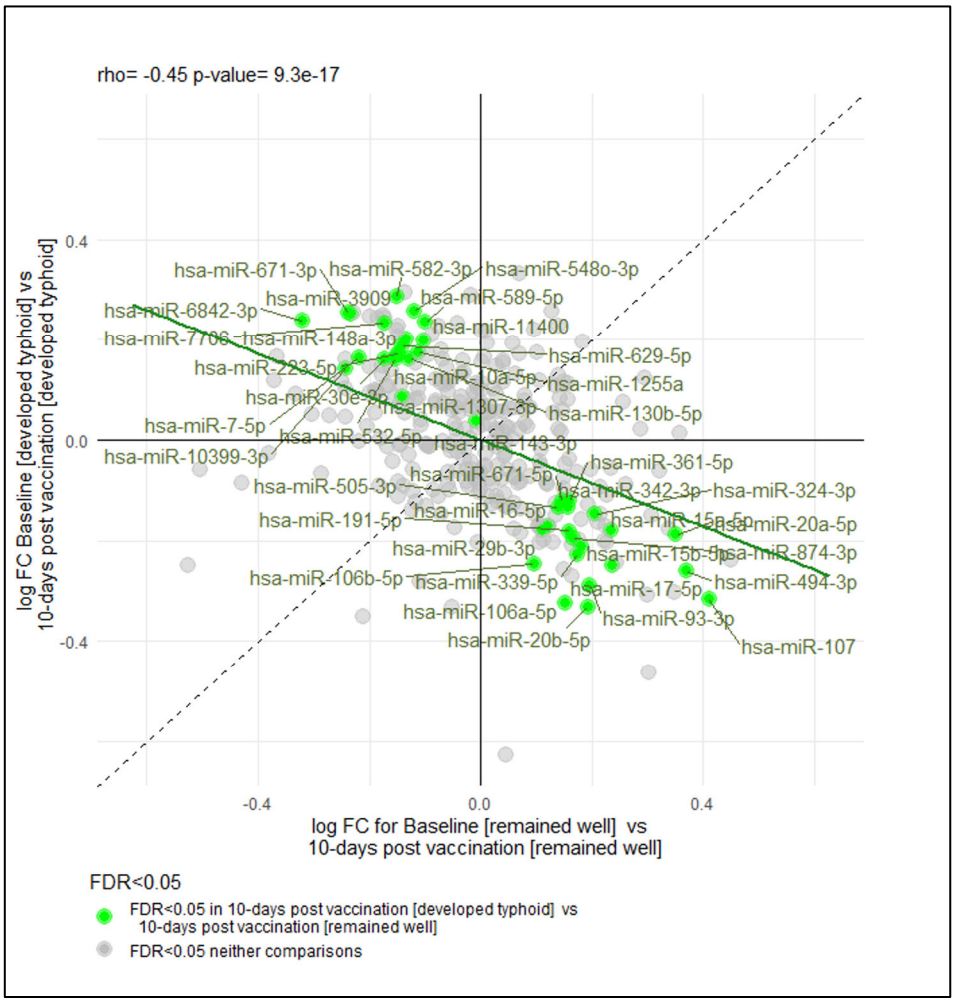


Figure 74 Agreement plot comparing baseline to 10 days post-vaccination log_2 foldchanges between outcome groups. X-axis: those that remained well after challenge. Y-axis: those that developed typhoid after challenge. Pearson r and p-value are shown.

6.4.4.3.2 MiRNA enrichment analysis reveals differences between outcome groups 10 days after vaccination

MIEAA2 miRNA enrichment analysis infers pathways targeted by miRNAs up and downregulated in diagnosed vs non-diagnosed participants (Figure 75A). At 10 days after vaccination, most pathways (3123/3134) were *positively enriched* in the diagnosed group compared with the non-diagnosed group—such pathways are predicted to be subject to increased miRNA suggesting greater miRNA

repression of pathways in people who developed typhoid after challenge. Common themes amongst enriched pathways related to the immune system (379 pathways) and RNA and DNA transcription/translation, modification and repair (317 pathways). Processes perturbed by immune responses, e.g., lipid metabolism and clotting, were also identified(78). Interestingly, processes associated with the nervous system (139), cardiac muscle and vasculature (119) were also identified (Figure 75A). All but 11 pathways were more enriched for targeting in the diagnosed group. In this analysis, enrichment is relative—positive enrichment in the diagnosed group can mean negative enrichment in the non-diagnosed group. Assessing enrichment with regards to baseline teases this apart (Appendix 9.5.2, Supplementary figure 5) and suggests the former interpretation is most correct, i.e. there is increased targeting of immune pathways in PBMCs 10 days after vaccination in those that developed typhoid.

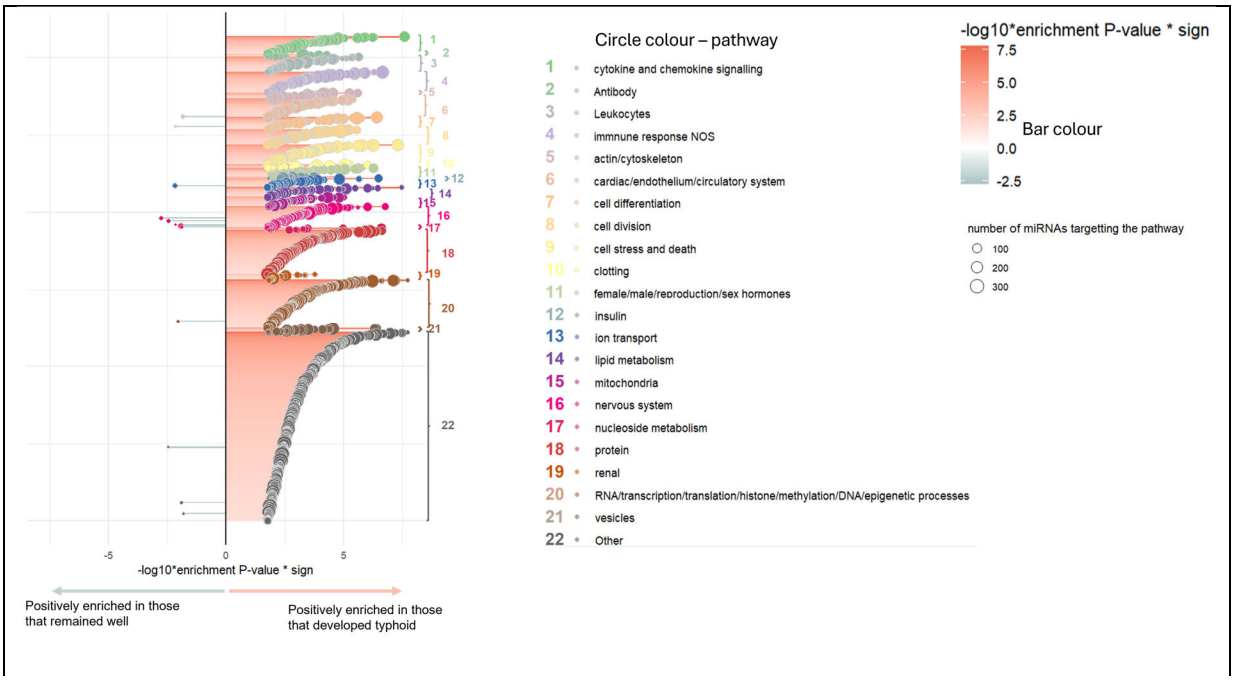


Figure 75 Lollipop plot summarising MIEAA2 miRNA pathway enrichment results comparing sRNA expression 10 days after vaccination in those that did and did not develop typhoid after challenge. Each dot represents the result for a different pathway. Only pathways enriched at an FDR<0.05 are shown. The MIEAA2 software does not

Figure 75 caption continued
return enrichment scores; therefore, the -log10 of the p-values are presented along with the direction of enrichment (positive values indicate positive enrichment in diagnosed participants, negative values indicate positive enrichment in non-diagnosed participants). The colour of the bar also reflects the -log10 of the p-value and direction of enrichment. Circle sizes represent the number of miRNAs targeting each pathway. Circle colour and number represent which theme each pathway belongs to. Results are ordered by theme and then significance. Positive enrichment means that a pathway will be subject to greater miRNA repression.

6.4.5 Theme 2: Does *S.* Typhi exposure and infection influence sRNA expression in PBMCs?

The challenge study permits exploration of host responses to a pathogen before disease onset. Where possible, the control vaccine group was used as a validation cohort for the Vi-vaccine cohort results. A summary of the differential expression results is presented in Figure 76 and Supplementary Table 7.

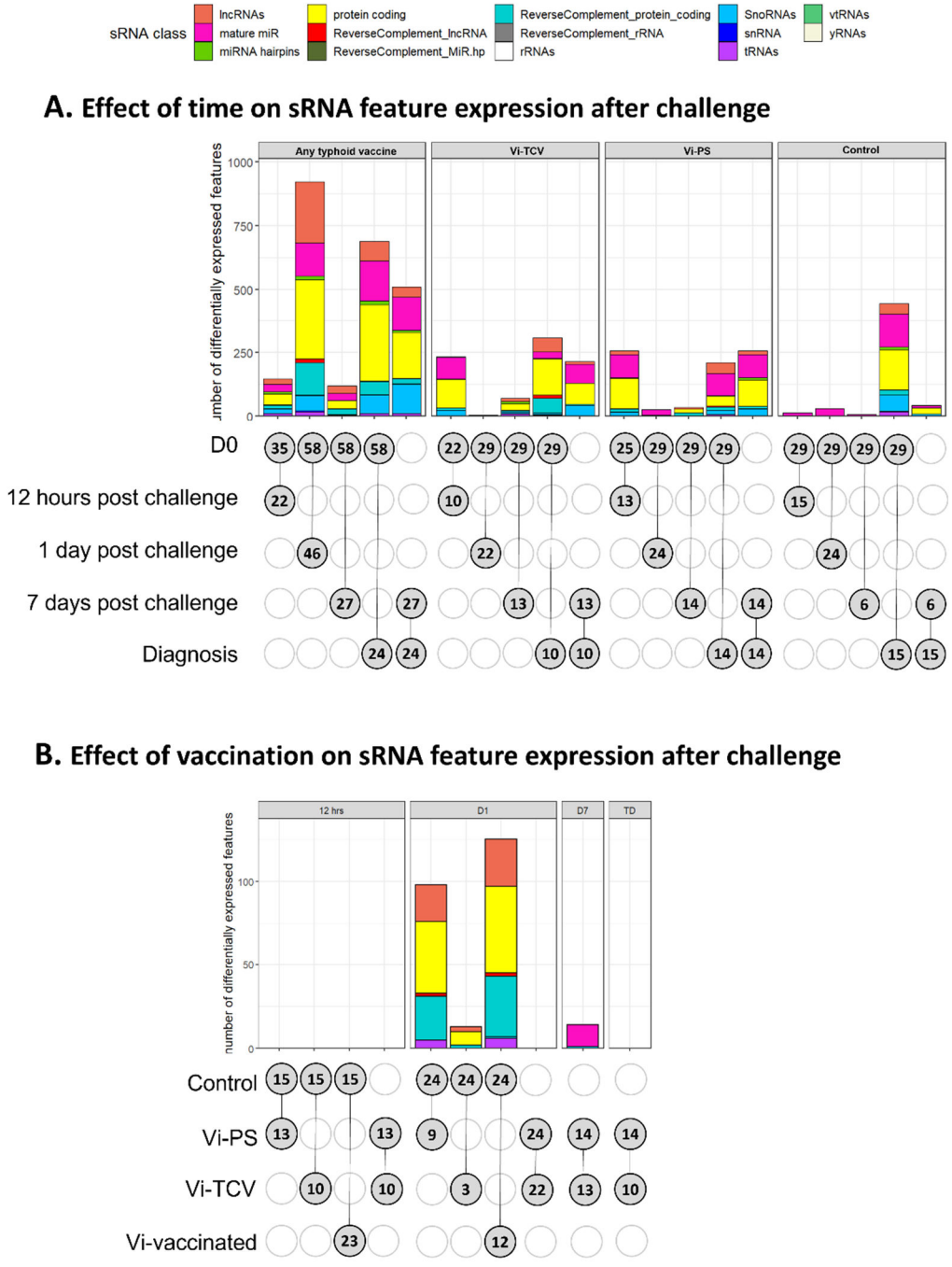


Figure 76 Bar plot showing the number of DE sRNAs between Vi-vaccine groups after challenge. Number of samples in each analysis is indicated in circles underneath each bar. Care should be taken when comparing the number of DE features between analyses in view of discrepancies in sample sizes.
A: longitudinal analysis of changes in sRNA feature expression over time
B: Cross-sectional analysis of changes in sRNA feature expression between vaccine groups at each time point.
Supplementary Table 7 contains the number of DE features per sRNA class.

6.4.5.1 Does S. Typhi exposure influence PBMC sRNA expression within 24 hours?

A longitudinal analysis explored whether *S.* Typhi exposure influenced sRNA expression. Figure 76 summarises the results.

6.4.5.1.1 Changes in sRNAs are present within 12 hours of *S.* Typhi exposure and strengthen at 24 hours.

S. Typhi exposure induces changes in PBMC sRNA expression within 12 hours, which further progresses at 24 hours (Figure *76*, Figure 77, and Figure 78). Separation between D0 and D1 is most pronounced when restricting to miRNA features (Figure 77B).

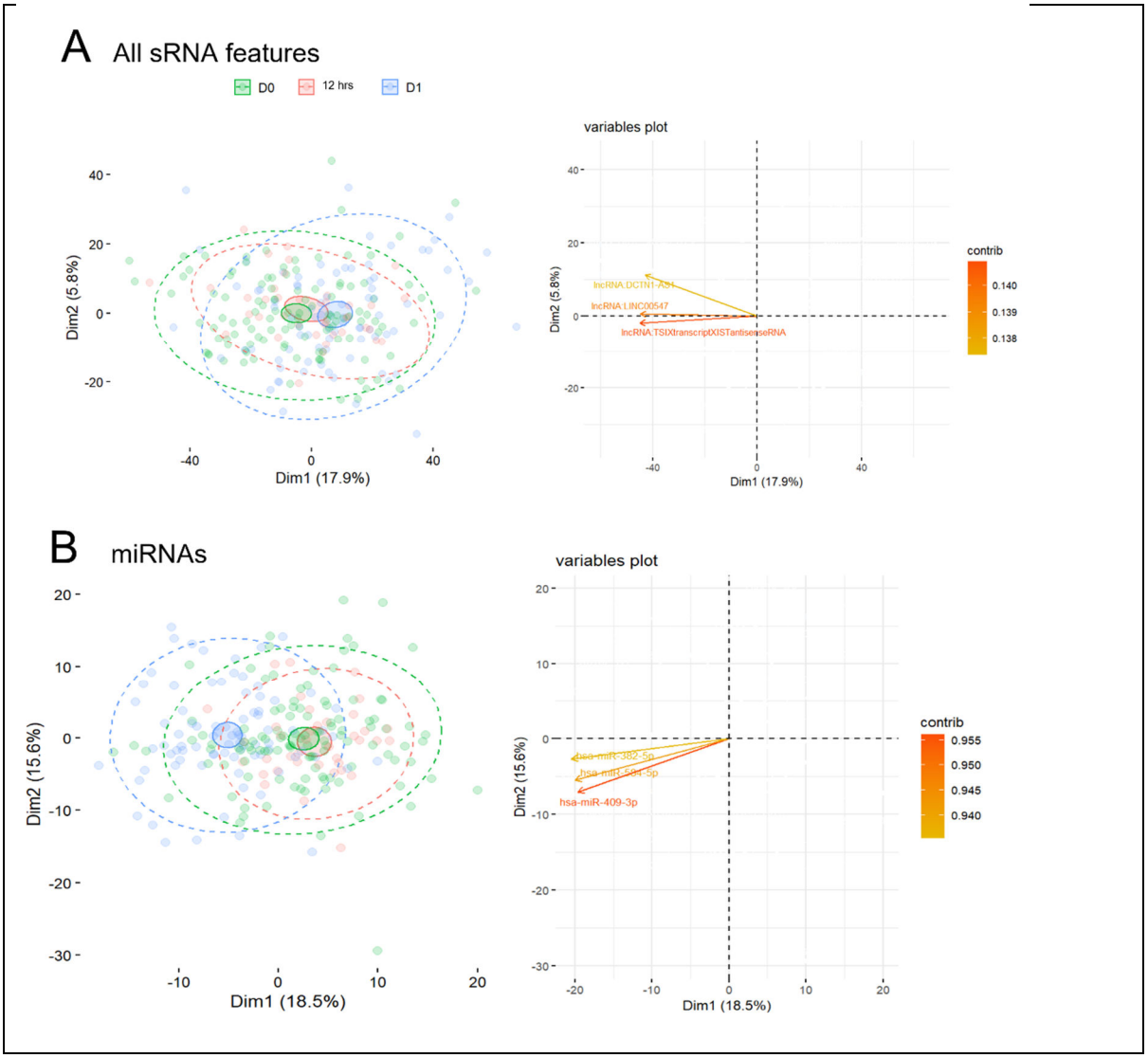


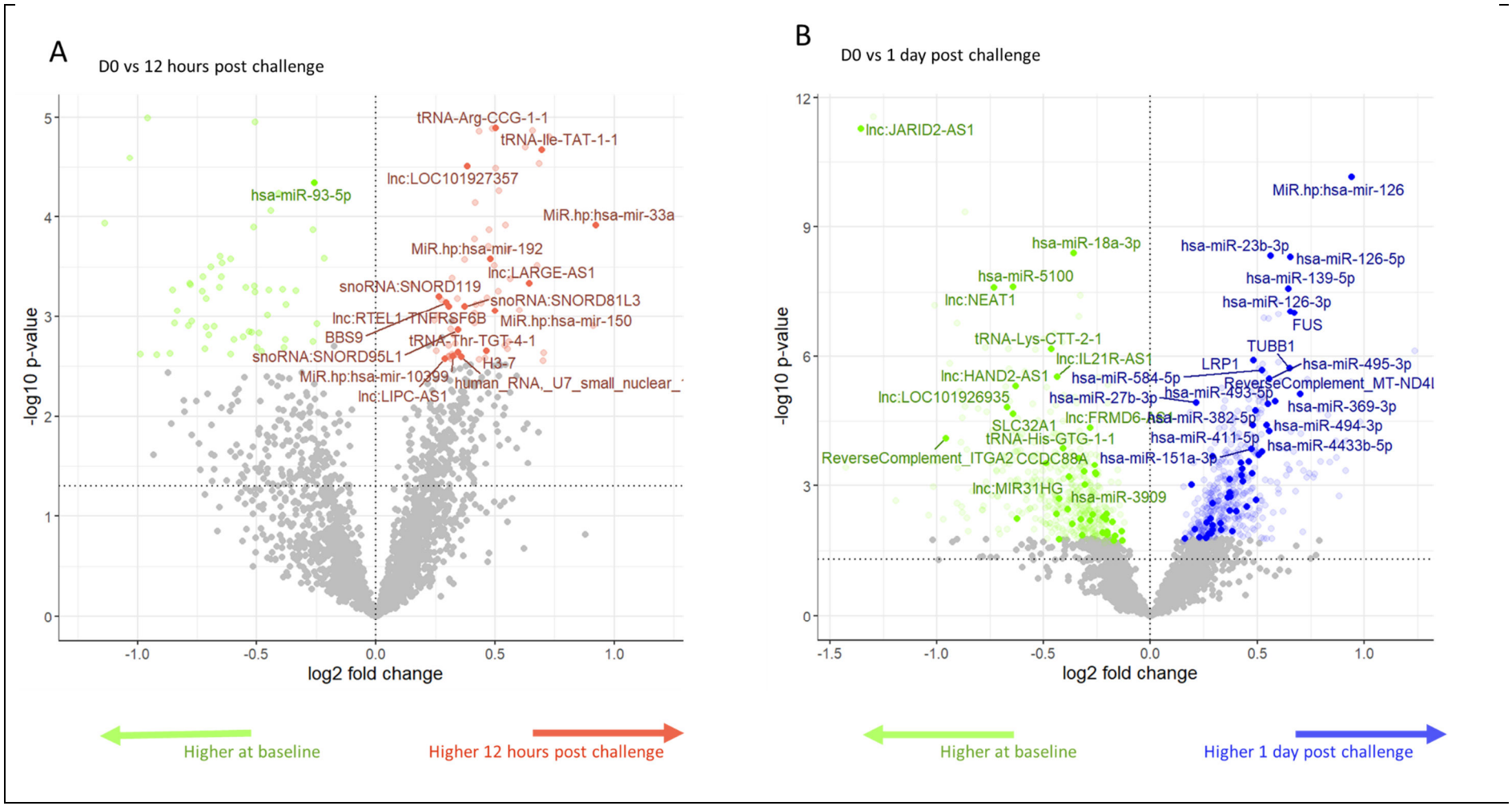
Figure 77 PCA of samples taken pre-challenge (day of challenge) and 12 hours and 1-day post-challenge. The outer dotted lines represent 95% confidence intervals of the data. The shaded ovals represent 95% confidence intervals of each timepoints' centroid.
A: All sRNAs used for PCA construction.
B: Only miRNA features used for PCA construction

Figure 78 A and B show which sRNAs were DE over the first 24 hours after challenge. sRNA changes at 12 hours are generally weaker than D1 but correlate nevertheless (Figure 78 C, r=0.38, $p<4.32 \times 10^{-93}$). Some sRNA changes contradict this agreement indicating time point-specific effects (orange dots in Figure 78C). sRNA-dfs uniquely DE at 12 hours were enriched in GO molecular

functions pertaining to ion transport, phosphorylation, nucleobase metabolism, and ATP metabolism.

6.4.5.1.2 miRNAs targeting pathways involved in post-transcription regulation are upregulated within 24 hours of *S.* Typhi ingestion

MIEAA2 miRNA pathway enrichment identified 0 and 93 pathways as positively enriched between D0 and 12 hr and D0 and D1, respectively. The most common theme amongst enriched pathways between D0 and D1 was RNA-induced silencing, which encompassed 17 pathways, and pertained to miRNA processing, RISC and siRNAs. Other themes were leukocytes, the nervous system and transcription (See Supplementary Table 8). The analysis predicts the downregulation of these pathways at the mRNA or protein level; however, the end effect on pathway expression cannot be assumed because miRNAs may be induced to buffer an increase in gene transcription rather than to downregulate protein expression.


A
D0 vs 12 hours post challenge
tRNA-Arg-CCG-1-1
tRNA-Ile-TAT-1-1
lnc:LOC101927357
hsa-miR-93-5p
MiR.hp:hsa-mir-33a
MiR.hp:hsa-mir-192
lnc:LARGE-AS1
snoRNA:SNORD119
snoRNA:SNORD81L3
lnc:RTEL1-TNFRSF6B
MiR.hp:hsa-mir-150
BBS9
tRNA-Thr-TGT-4-1
snoRNA:SNORD95L1
H3-7
MiR.hp:hsa-mir-10399
human_RNA._U7_small_nuclear_
lnc:LIPC-AS1
-log10 p-value
log2 fold change
Higher at baseline
Higher 12 hours post challenge
B
D0 vs 1 day post challenge
lnc:JARID2-AS1
MiR.hp:hsa-mir-126
hsa-miR-18a-3p
hsa-miR-23b-3p
hsa-miR-126-5p
hsa-miR-5100
hsa-miR-139-5p
lnc:NEAT1
hsa-miR-126-3p
FUS
tRNA-Lys-CTT-2-1
TUBB1
lnc:IL21R-AS1
LRP1
hsa-miR-495-3p
lnc:HAND2-AS1
hsa-miR-584-5p
ReverseComplement_MT-ND4l
lnc:LOC101926935
hsa-miR-27b-3p
hsa-miR-493-5p
hsa-miR-369-3p
lnc:FRMD6-AS1
hsa-miR-382-5p
SLC32A1
hsa-miR-494-3p
tRNA-His-GTG-1-1
hsa-miR-411-5p
hsa-miR-4433b-5p
ReverseComplement_ITGA2 CCDC88A
hsa-miR-151a-3p
lnc:MIR31HG
hsa-miR-3909
-log10 p-value
log2 fold change
Higher at baseline
Higher 1 day post challenge

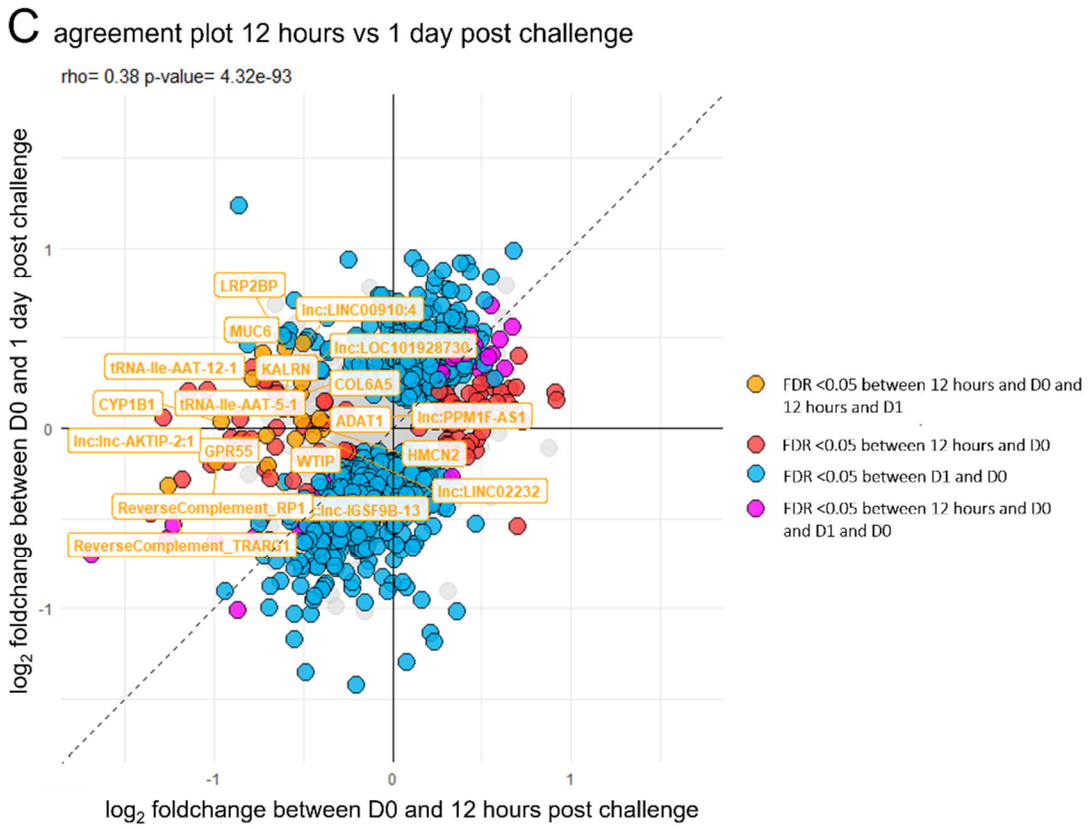


Figure 78 Changes in sRNA feature expression in the first 24 hours after challenge.
A: Volcano plot of log_2 foldchanges between D0 and 12 hours post-challenge in Vi-vaccinated participants. Coloured dots represent sRNAs with FDR<0.05. Labelled dots were validated in the control vaccinee group (i.e. the same direction of foldchange and p<0.05)
B: Volcano plot of log_2 foldchanges between D0 and 1-day post-challenge in Vi-vaccinated participants. Coloured dots represent sRNAs with FDR<0.05. Labelled dots were validated in the control vaccinee group (i.e. the same direction of foldchange and p<0.05)
C: Agreement plot of log_2 foldchanges between baseline and 12 hours (x-axis) and baseline and 24 hours (y-axis). Orange labelled dots are sRNAs that were unilaterally DE at 12 hours, i.e. compared with baseline and D1.

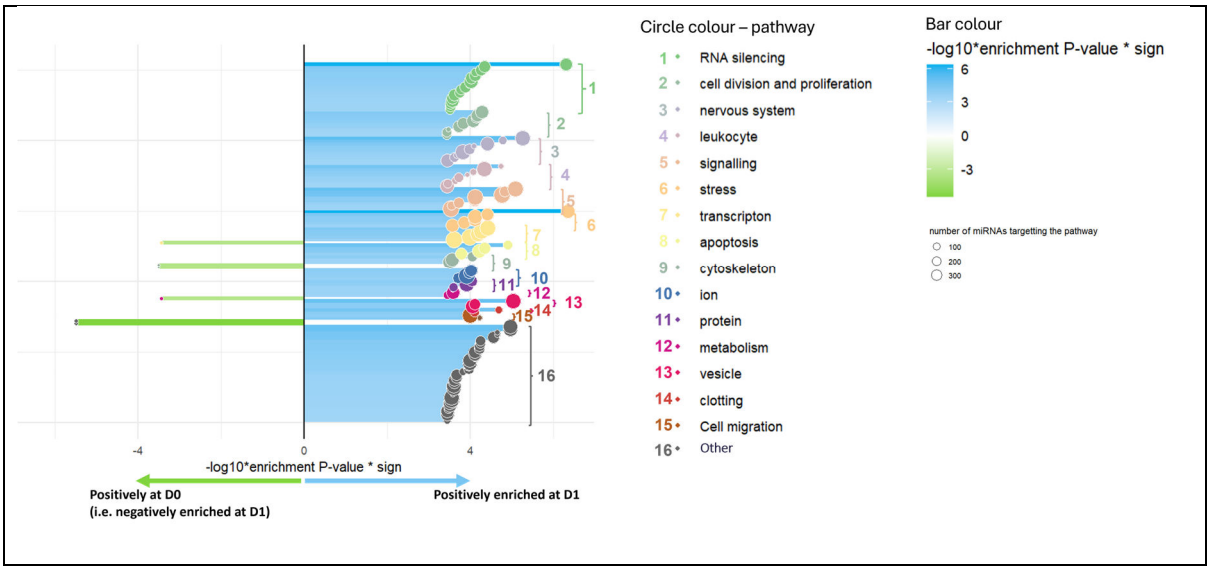


Figure 79 Lollipop plot summarising miRNA pathway enrichment results comparing sRNA expression at D0 versus D1. Each dot represents the result for a different pathway. Only pathways enriched at an FDR<0.05 are shown. The MIEAA2 software does not return enrichment scores; therefore, the -log$_{10}$ of the p-values are presented along with the direction of enrichment (positive values indicate positive enrichment in diagnosed participants, negative values indicate positive enrichment in non-diagnosed participants). The colour of the bar also reflects the -log$_{10}$ of the p-value and direction of enrichment. Circle sizes represent the number of miRNAs targeting each pathway. Circle colour and number represent which theme each pathway belongs to. Results are ordered by theme and then significance. Positive enrichment means that a pathway will be subject to greater miRNA repression and, therefore, may be downregulated (though this depends on other competing drivers of expression within cells).

Strikingly, MIEAA2 identified *Salmonella* infection as the top KEGG pathway (out of 326 possible pathways) for the D0 vs D1 comparison ($p=5.7 \times 10^{-6}$, FDR=9.7 $x10^{-4-}$). Eighty-four DE miRNAs target 166 genes within this pathway (Figure 80). This suggests changes to miRNA expression 24 hours after *S.* Typhi exposure is somewhat specific for *Salmonella* infection, compared with other infections listed in KEGG, including bacterial, viral and parasitic. This finding supports the concept of pathogen-specific host-pathogen interactions, with miRNAs potentially contributing to this specificity by curating cell responses at the post-transcriptional level.

Appendix Figure 43 and Appendix Table 9 contain details of pathways enriched amongst sRNA-dfs DE at 12 and 24 hours post-vaccination.

6.4.5.1.3 Vi-vaccines influence sRNA expression early after *S.* Typhi exposure

To look at the timing of vaccine-derived immunity (i.e. does it occur shortly after ingestion or after dissemination), sRNA expression was compared between control and Vi-vaccinated participants after challenge. sRNA feature expression differed between Vi-vaccine and control groups at D1, even after accounting for potential D0 differences between vaccine groups (Figure 76 B, Figure 81, Figure 82, Appendix Figure 42). Nineteen sRNA-dfs had significantly greater D0-D1 foldchanges in Vi-vaccinees than control participants (dark blue labelled points Figure 82A, Figure 82B).

Overall, these results suggest *S.* Typhi ingestion causes two groups of changes in sRNA expression. One group of sRNAs change irrespective of vaccination—and may represent an innate response to challenge or cross-reactive pre-existing immunity to *S.* Typhi antigens. These sRNAs are mainly miRNAs, e.g., Figure 82C. The other group of sRNAs change due to Vi-vaccine-induced immune responses (e.g., Figure 82B and Supplementary figure 6). These are mainly sRNA-dfs, presumably linked to Vi-specific B-cell response, as these are induced by the Vi-vaccines but not the control vaccine.

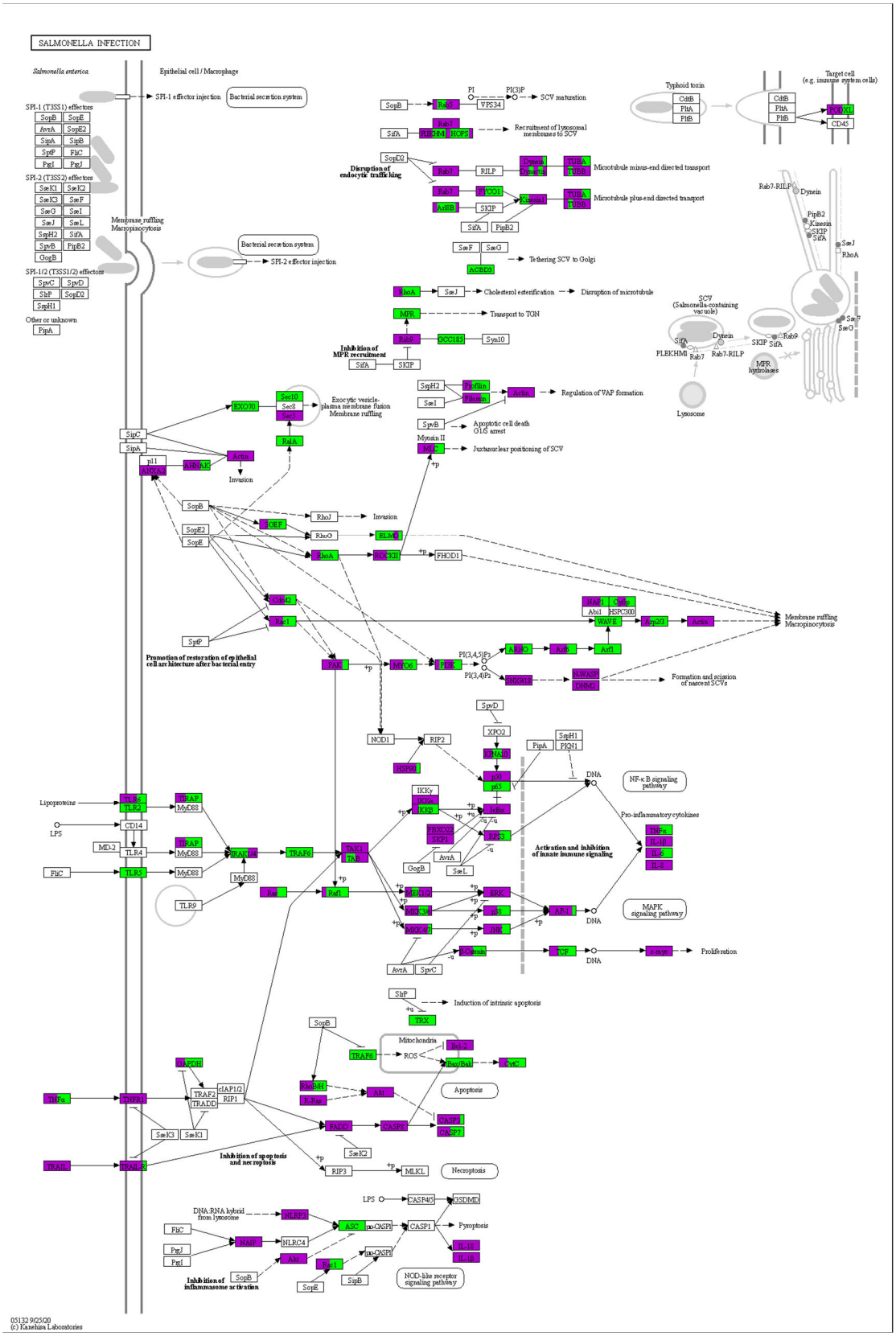


Figure 80 Genes involved in the Salmonella infection KEGG pathway. Genes in purple are targeted by miRNAs that are upregulated at D1. Genes in green are targeted by miRNAs that are downregulated at D1.

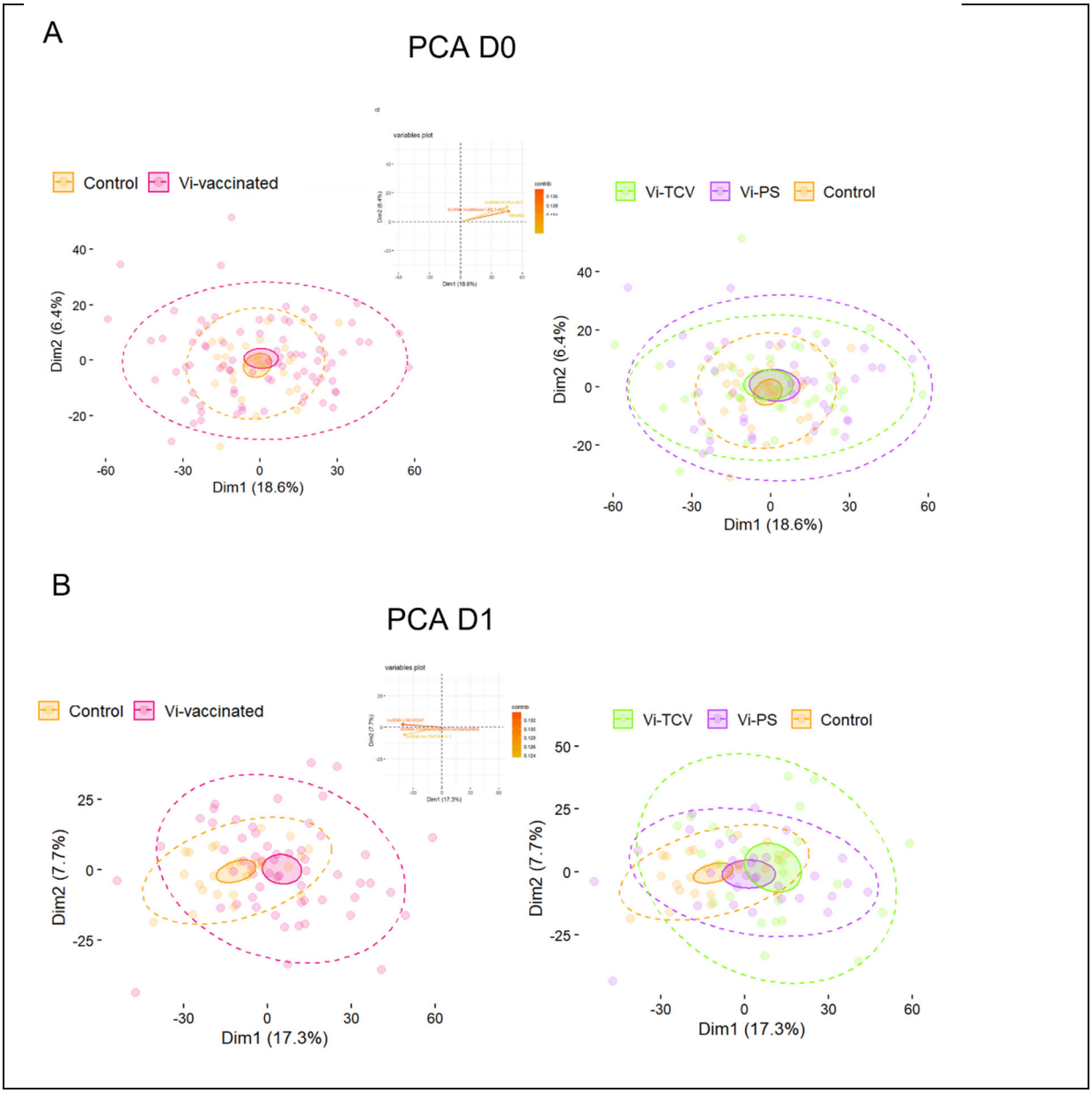


Figure 81 PCA of samples taken at D0 and D1 coloured by vaccine group. The outer dotted lines represent 95% confidence intervals of the data. The shaded ovals represent 95% confidence intervals of each timepoints' centroid.
A: PCA of samples taken at D0. The plot on the left shows samples coloured by whether a participant received a Vi-containing or control vaccine. Plot on the right shows samples broken down by vaccine. Vaccine type was not associated with PC1 or PC2
B: PCA of samples taken at D1. Plot on the left shows samples coloured by whether a participant received a Vi-containing vaccine or the control vaccine. Plot on the right shows samples broken down by vaccine.

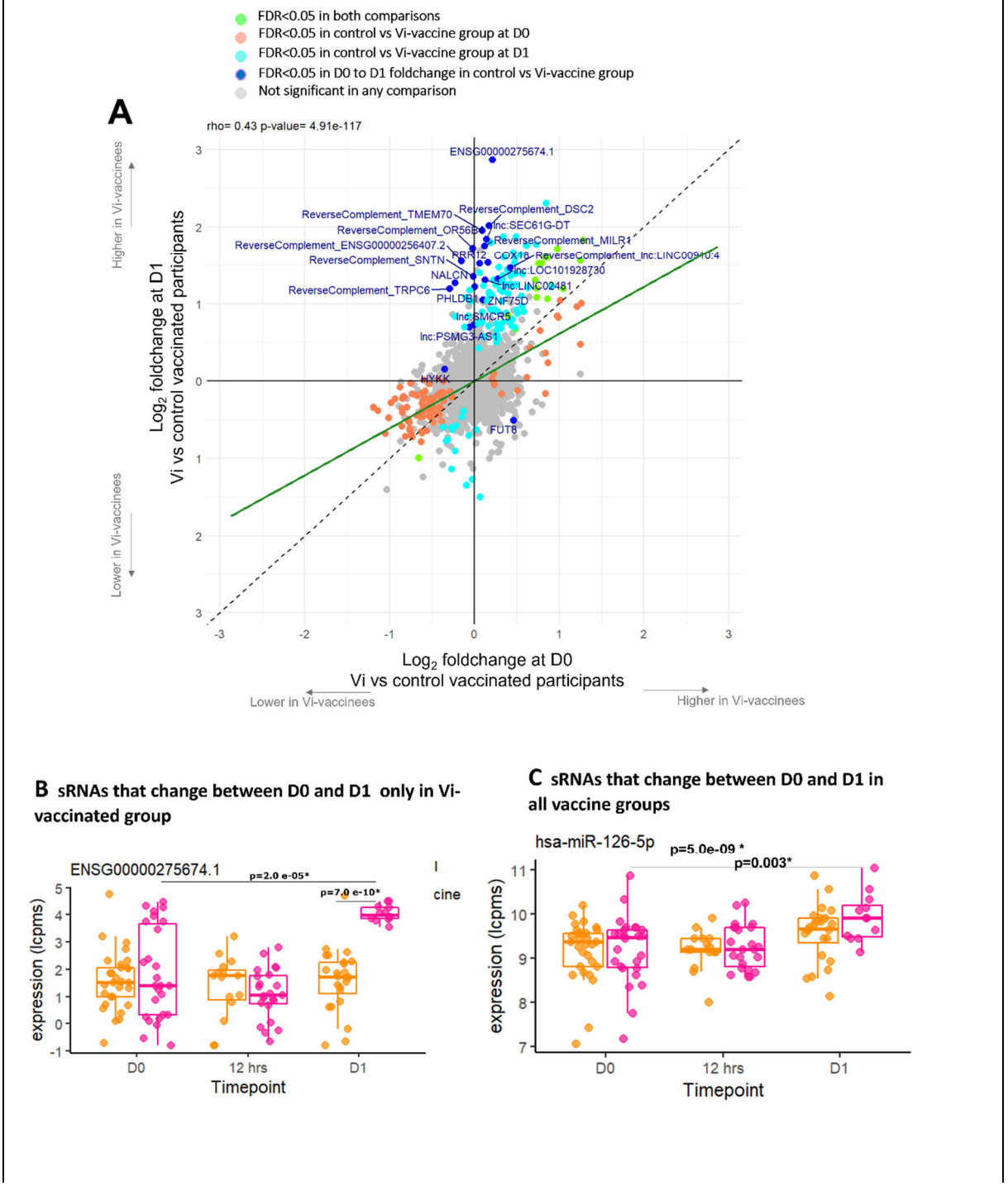


Figure 82

A: Agreement plot of log₂ foldchanges Vi- and control vaccinees at D0 (x-axis) and D1 (y-axis). sRNAs that were DE after adjusting for the difference in D0 expression are highlighted and labelled in dark blue. Green line = line of best fit

B: Example of a dot plot of an sRNA that changes at D1 in the Vi-vaccinees group only

C: Example of a dot plot of an sRNA that changes at D1 in all vaccine groups

6.4.5.1.4 A greater response to challenge at 24 hours is associated with protection from typhoid.

No sRNAs were DE between outcome groups within 24 hours after challenge (at FDR<0.05), even when restricting to features that changed by challenge. Despite this, perturbation in sRNA feature expression was more pronounced in non-diagnosed participants based on the distribution of effect sizes. This phenomenon was observed in all vaccine subgroups (Figure 83, Appendix Figure 43). In Vi-TCV vaccinated participants, there was also an association between challenge outcome and miRNA perturbation at the participant level when Euclidean distance was used to quantify each participant's D0 to D1 perturbation (Figure 83).

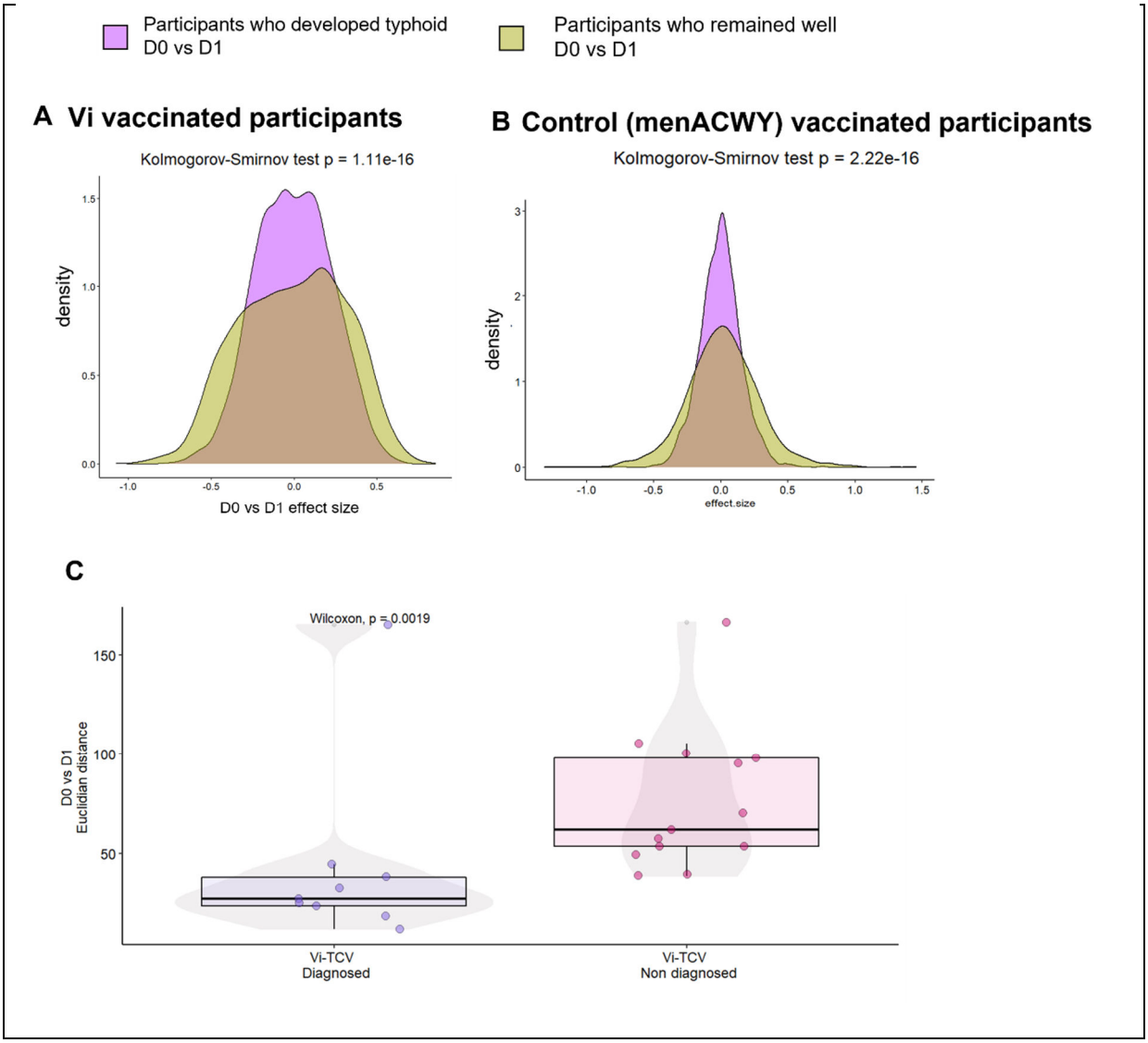


Figure 83 comparing effect sizes at D0 vs D1 between diagnosed and non-diagnosed groups.
A: Vi-vaccinees. Left panel—distribution of D0 vs D1 effect sizes in diagnosed (purple) and non-diagnosed (green) participants. Kolmogorov Smirnov statistic is used to compare the distributions.
B: Control vaccinees. Left panel—distribution of D0 vs D1 effect sizes in diagnosed (purple) and non-diagnosed (green) participants. Kolmogorov Smirnov statistic is used to compare the distributions.
C: Euclidean distances between participants' D0 and D1 sample in the Vi-TCV group only, broken down by outcome.

6.4.5.2 Does typhoid influence PBMC sRNA expression?

Longitudinal and cross-sectional analyses explored whether *S.* Typhi infection influences sRNA expression compared with health (D0) or 7 days after challenge

(D7) in those who remained well. Differential expression results are summarised in Figure 11. The control vaccine group was used to validate results where possible.

6.4.5.2.1 sRNA expression is markedly influenced by acute typhoid.

PCA shows samples taken at typhoid diagnosis cluster separately from the other samples (Figure 84). The overlap between D7 and D1 samples suggests similar profiles for the most variable genes. Differential expression was seen between typhoid diagnosis and D7 and D0 (Figure 85, See Appendix Table 10). The proportion of features validated in the control group varied from 40% to below 1%. This is partly due to a small number of D7 samples in the control group (n=6).

Hierarchical clustering based on DE features shows good separation of typhoid diagnosis samples from D0 and D7 samples (Figure 86A), with hsa-miR-1303 (the most upregulated miRNA at diagnosis) differentiating diagnosis samples from all other samples with good accuracy (Figure 85). This feature was selected as the only feature on component 1 in an sPLS-DA analysis trained to differentiate typhoid diagnosis from baseline, D0 and D7 samples, highlighting its importance in the immune response during typhoid (see Appendix Figure 44).

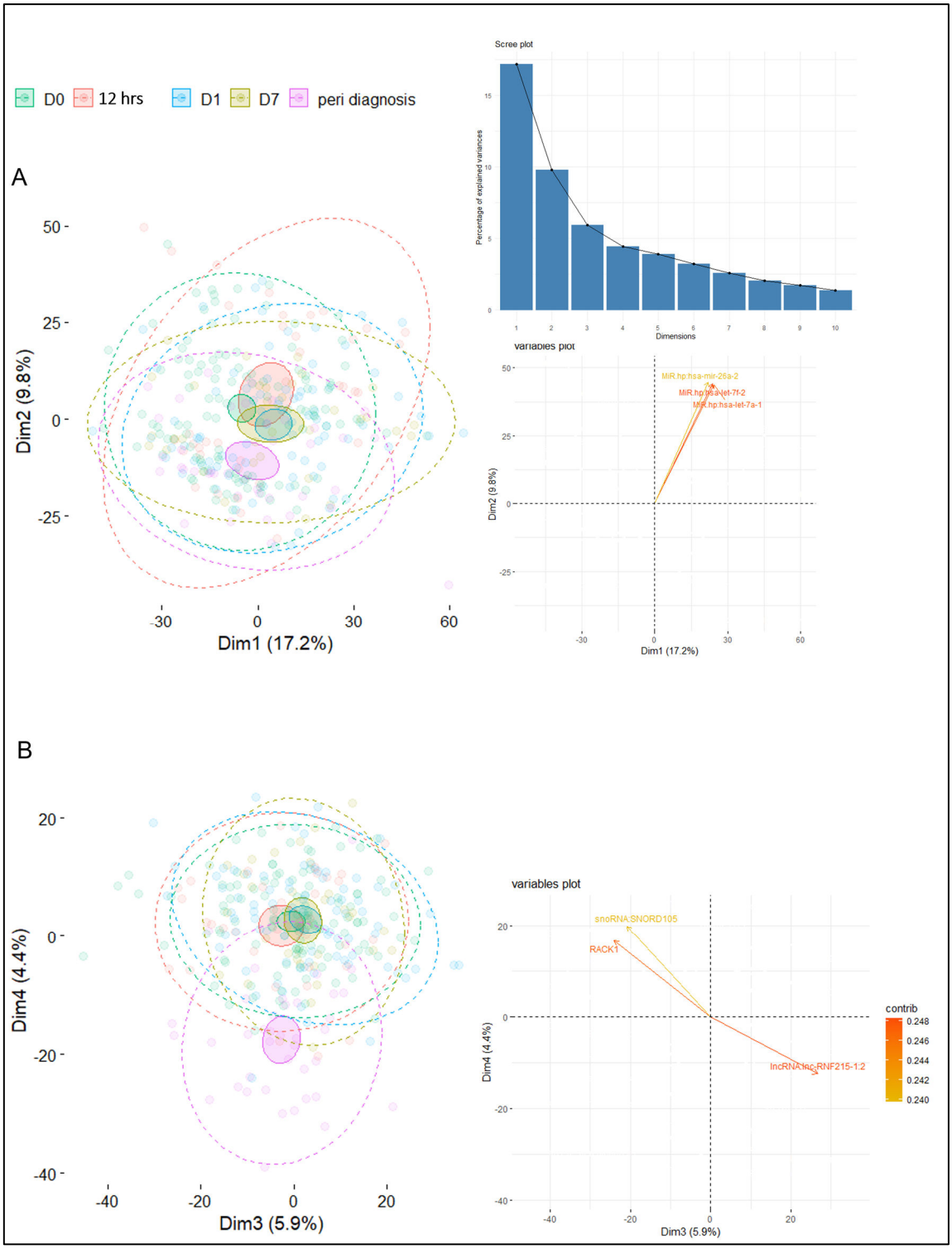


Figure 84 PCA of samples taken at D0, 12 hours, D1, D7 and at diagnosis. The outer dotted lines represent 95% confidence intervals of the data. The shaded ovals represent 95% confidence intervals of each timepoints' centroid.

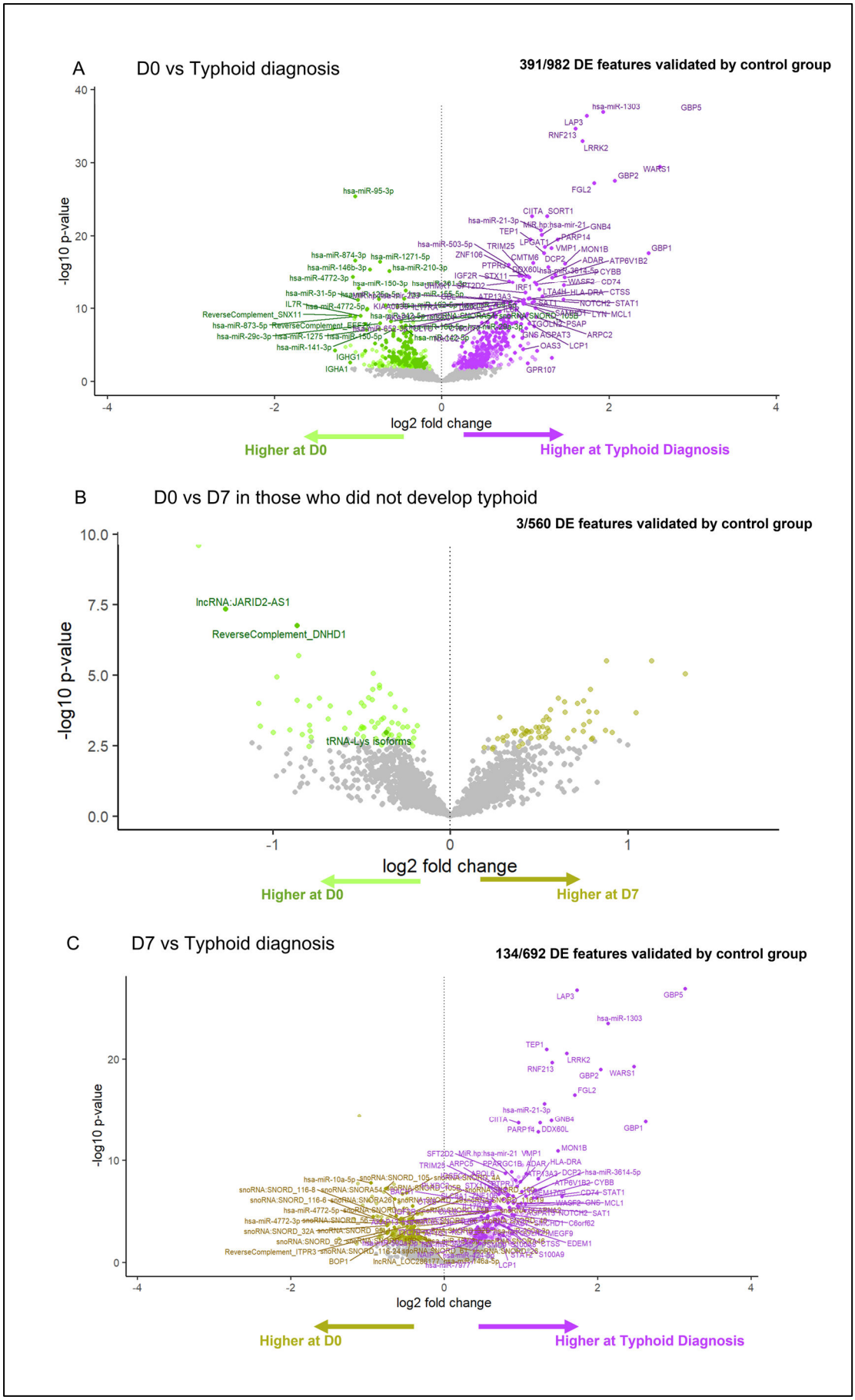

A
D0 vs Typhoid diagnosis
391/982 DE features validated by control group
-log10 p-value
log2 fold change
Higher at D0
Higher at Typhoid Diagnosis
B
D0 vs D7 in those who did not develop typhoid
3/560 DE features validated by control group
lncRNA:JARID2-AS1
ReverseComplement_DNHD1
tRNA-Lys isoforms
-log10 p-value
log2 fold change
Higher at D0
Higher at D7
C
D7 vs Typhoid diagnosis
134/692 DE features validated by control group
-log10 p-value
log2 fold change
Higher at D0
Higher at Typhoid Diagnosis

Figure 85 Changes in sRNA feature expression at D7 in those that remained well and at diagnosis in those that developed typhoid. Coloured dots represent sRNAs with FDR<0.05. Labelled dots were validated in the control group (i.e. the same direction of foldchange and p<0.05). Participant modelled as a random effect.
A: Volcano plot of log_2 foldchanges between D0 (all participates) and diagnosis in Vi-vaccinated participants.
B: Volcano plot of log_2 foldchanges between D0 (all participants) and D7 in Vi-vaccinated participants that remained well.
C: Volcano plot of log_2 foldchanges between diagnosis and D7 in Vi-vaccinated participants that remained well.

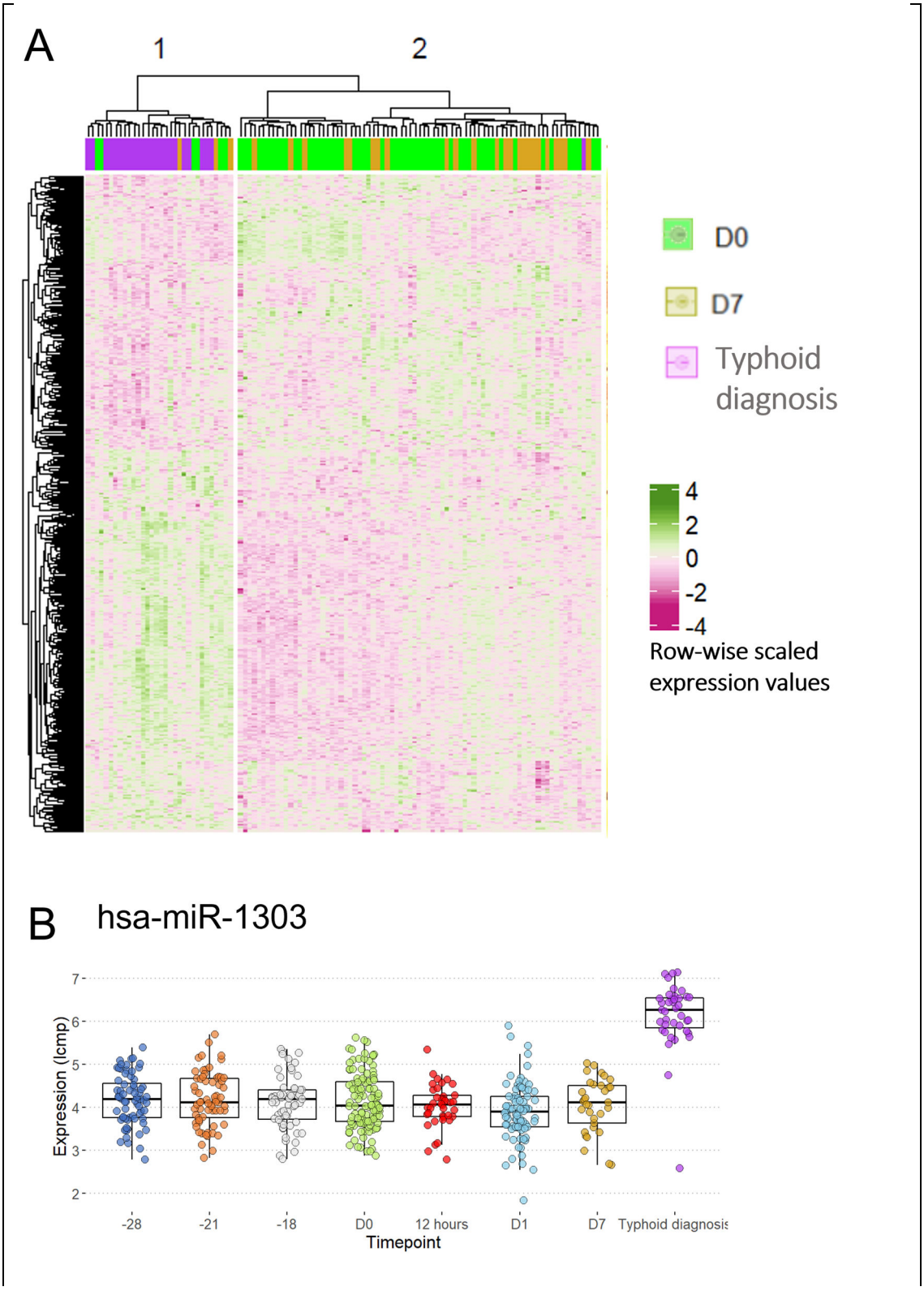


Figure 86
A: Heat map of D0, diagnosis and D7 samples, generated using DE sRNAs. Dendrograms represent hierarchical clustering based on Euclidian distances and Ward D2 method.
B: Dot plot of hsa-miR-1303 expression across samples. The miRNA is a good discriminator, apart from two outliers in the diagnosis group.

6.4.5.2.2 Pathways pertaining to T-cells are enriched in sRNA-dfs at typhoid diagnosis, but no pathways are enriched as miRNA targets.

ORA analysis on sRNA-dfs upregulated at typhoid diagnosis revealed 298 enriched pathways; these included terms relating to phagocytosis and lymphocyte activation and proliferation with some terms specific to T-cells but not B-cells. In keeping with this, enriched Reactome pathways included interferon signalling.

Despite significant perturbation in miRNA expression at typhoid diagnosis compared with D0, MIEAA2 pathway enrichment analysis did not identify any pathways as significantly positively or negatively enriched for targeting at FDR<0.05, and neither did an ORA approach. This could be due to different changes in different cells, obscuring individual signals. Interestingly, miR-146b and miR-155 were downregulated at typhoid diagnosis (Appendix Table 5) despite the former generally being associated with an anti-inflammatory phenotype, whilst miR-155 is generally pro-inflammatory. This potentially reflects mixed signals when looking at bulk cell populations.

In contrast to the lack of miRNA pathway enrichment at diagnosis, 229 pathways were depleted at D7 versus D0. These pathways mainly pertained to the immune system and carbohydrate metabolism. Even though miRNA profiles were similar at D1 and D7, there were differences at the pathway level; in particular, miRNA and RISC terms were enriched at D1 compared with both D0 and D7.

6.4.5.2.3 sRNA expression at typhoid diagnosis is not influenced by vaccine type

There were no differences between Vi-TCV and Vi-PS at typhoid diagnosis. Vi-vaccines and control vaccines could not be compared at D7 or typhoid diagnosis because they were run in separate batches, precluding robust biological conclusions.

6.4.5.2.4 Vaccination influences sRNA expression 7 days after challenge in those who do not develop typhoid.

Vaccination influences sRNA feature expression 7 days post-challenge in non-diagnosed participants (Figure 76 and Figure 87); 13/14 of DE features were miRNAs. In addition, DE features separate vaccine groups with high accuracy (Figure 87). The results indicate differences in how the immune system responds to S. Typhi after plain polysaccharide versus conjugate vaccine, presumably due to T-cell-dependent antibody responses elicited by Vi-TCV. Interestingly, D7 Vi-PS expression profiles are more similar to D1 than D7 Vi-TCV samples. The reason is unclear. Potentially, Vi-PS vaccinees experience ongoing exposure to S-Typhi in the gut, which Vi-TCV vaccinees cleared, but stool-shedding data did not support this. MIEAA2 miRNA enrichment analysis identified 979 terms negatively enriched in the Vi-TCV group compared with the Vi-PS at D7. These negatively enriched terms included T-cell terms, which were also negatively enriched compared with baseline—this encourages upregulation at the mRNA level. These findings suggest vaccine-specific differences in induction of different

adaptive immune responses, e.g. T-cells, but this was not identified in the CyToF or mRNA-seq data.

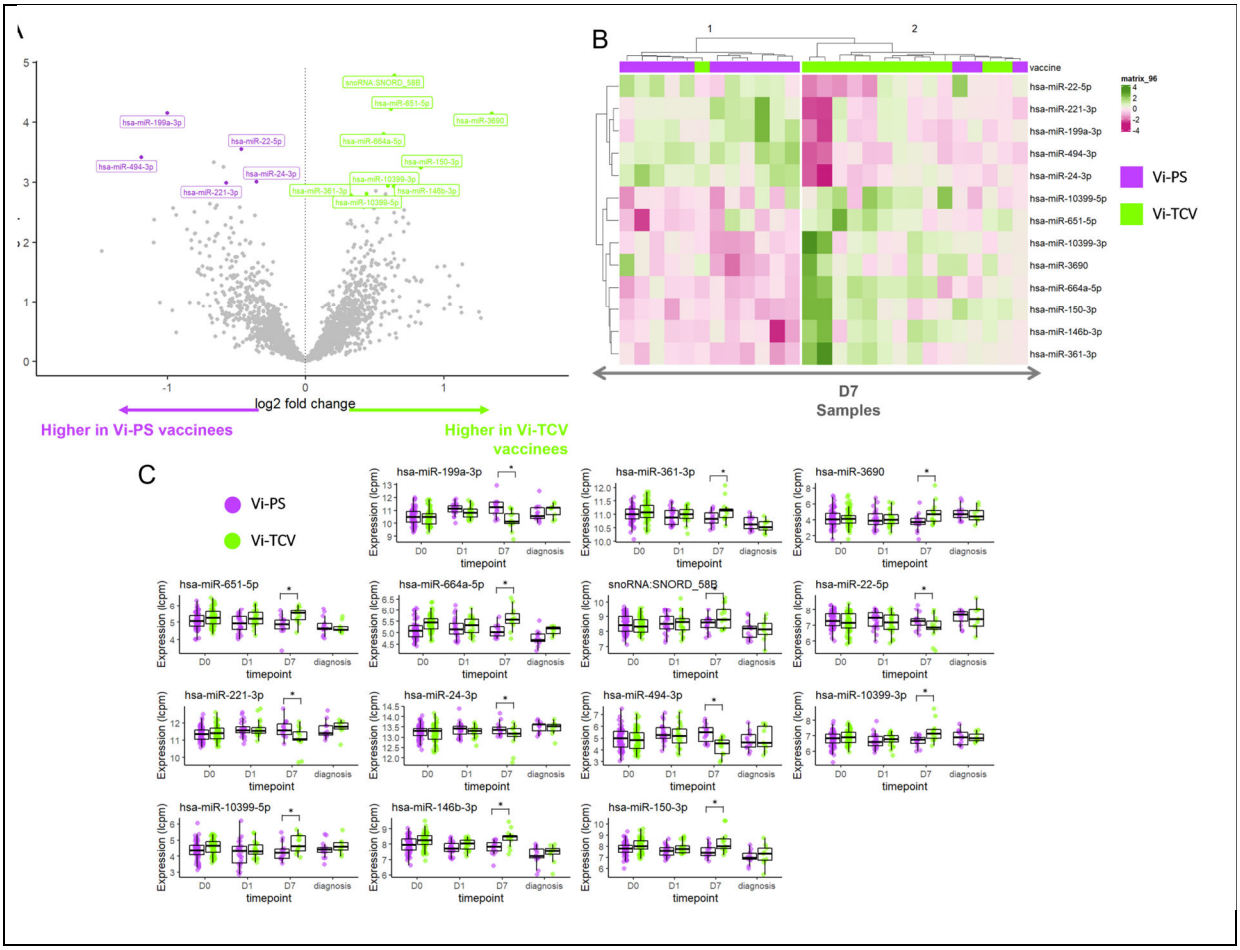


Figure 87 Comparing Vi-PS and Vi-TCV at D7 in those that did not develop typhoid.
A: Volcano plot of log_2 foldchanges between Vi-PS and Vi-TCV vaccinees at D7 after challenge. Coloured dots represent sRNAs with FDR<0.05. Labelled dots were validated in the control vaccinee group (i.e. the same direction of foldchange and p<0.05).
B: Heat map of D7 samples, broken down by Vi vaccine. Plot generated using DE sRNAs. Dendrograms represent hierarchical clustering based on Euclidian distances and Ward D2 method.
C: Dot plots showing the expression of differentially expressed sRNAs over time.

6.5 Discussion

PBMC sRNA expression was investigated in relation to vaccine responses and the host response to *S.* Typhi. Take-home answers to the research questions posed at the start of the chapter are summarised in Figure 88.

6.5.1 Theme 1: Does vaccination influence sRNA expression in PBMCs?

6.5.1.1 sRNA changes over time

The paucity of miRNA changes after vaccination compared with baseline mirrors the literature. The only publicly available dataset to assess PBMC miRNA changes after vaccination found no differences in miRNA expression after flu vaccination at an FDR<0.05[457]. In contrast to miRNAs, sRNA-dfs do change after vaccination. Notably, sRNA-dfs from the constant region of immunoglobulin genes (IGKC, JCHAIN, IGHA1) were upregulated 7 days (and 10 days in the case of JCHAIN sRNA-df) after vaccination and correlated with outcome. These changes were restricted to the Vi-TCV subgroup, probably reflecting a feature of T-cell dependent responses (antibody class switching, memory cell production, increased plasma cell production—see section 1.2.1.2.3.1), likely related to Vi, not just tetanus[458]. In the Vi-PS group, baseline expression of these sRNAs was associated with susceptibility to typhoid (and persisted 7 days after vaccination), whereas the opposite was true in the Vi-TCV group, potentially indicating people who respond best to a Vi-TCV vaccine differ from those who respond best to a Vi-PS vaccine, supporting the concept of personalised vaccine strategies.

6.5.1.2 MiRNA correlates of protection

The expression of miRNA correlates of protection was similar between vaccine groups; this contrasts with serological and cellular correlates, which were higher in Vi-TCV vaccinees[459–461]. The association between vaccine failure and increased miRNA repression of pathways relating to the immune system, RNA and DNA transcription, translation, modification and repair, and lipid metabolism and clotting, infers that non-protected individuals fail to control post-transcriptional regulation by miRNAs. miRNA differences may also reflect differences in cell frequencies within PBMC populations rather than intracellular miRNA expression. Unfortunately, there are no whole blood mRNA sequencing or CyTof data at this timepoint to pursue this idea. Single-cell dual mRNA:miRNA sequencing could resolve this issue, but the technology is not yet sufficiently high throughput.

Discovering divergence in miRNA expression is associated with protection provides a novel biological insight into effective vaccine responses but opens up several questions. What drives these divergent responses? Are divergent responses indicative of the immune response to vaccination or reflective of how people respond to antigenic challenge *per se*? It would be helpful to sequence samples from control vaccinees at 10 days post-vaccination, in case divergence reflects immunological competence to respond to antigen (and bacteria) rather than a Vi-specific response *per se*. It would be useful to assess whether divergent miRNA responses generalise to other vaccines (and species) and whether they can be overcome with tweaks to vaccine design. If this were the case, they could be valuable biomarkers of vaccine effectiveness to rationalise vaccine design in pre-clinical and early-phase clinical trials. The lack of studies

assessing miRNAs as correlates of protection needs to be addressed, particularly focussing on PBMC miRNA expression 10 days after vaccination.

6.5.2 Theme 2: Does exposure to *S.* Typhi influence sRNA expression in PBMCs?

6.5.2.1 Changes in miRNA and sRNA-df expression within 24 hours of challenge

Detectable changes in sRNA expression 12 hours after challenge are not necessarily surprising as systemic LPS challenge models show changes to immune profiles within 4 hours[462]. Nevertheless, given the first 24 hours of *S.* Typhi challenge is essentially asymptomatic (unlike LPS challenge) and less direct (S. Typhi must traverse the stomach and interact with host cells), detecting systemic perturbation at 12 hours was noteworthy and demonstrates asymptomatic exposure and subclinical infection with *S.* Typhi induces systemic signals.

Most changes seen at 12 hours strengthened at 24 hours—this is at odds with the whole blood mRNA-seq results from a previous Typhoid challenge study, which showed mRNA changes were greatest 12 hours post-challenge and resolved by 24 hours (unpublished). Differences may arise from the granulocytes in whole blood but not PBMCs, particularly as the 12-hour changes in the whole blood data were mainly linked to neutrophil processes. Running mRNA sequencing on VAST PBMCs at D0, 12 hours, and 24 hours after challenge would

resolve this question. Another explanation is temporal differences in transcriptomic change and post-transcriptional gene regulation by miRNAs. In line with this, the most enriched pathways for miRNA targeting at 24 hours post-challenge related to RNA-induced silencing, most of which pertained to miRNA processing and RISC siRNAs. Essentially, miRNAs seem to target their biogenesis, possibly as part of a feedback loop.

At 24 hours after vaccination, there was enrichment for miRNA targeting of pathways involved in the nervous system. The relevance of this is unclear. It may reflect the pleiotropic effects of miRNA gene silencing, dependent upon the cellular and temporal context. Whether the circulating factors (e.g., cytokines), which influence miRNA profiles within PBMCs, similarly influence miRNA expression in other cells is unclear; however, the literature supports interactions between the nervous system and the immune response, which extends to *Salmonella* species. For example, a protective GI tract response to *S.* Typhimurium is mediated by nociceptive neurons within the gut that release calcitonin gene-related peptide (CRGP) after detecting microbial product, which prevents invasion of the gut wall[463].

6.5.2.2 Vi vaccination influences sRNA expression after S. Typhi challenge

sRNAs that change after challenge, irrespective of vaccination, may represent an innate response to challenge or pre-existing cross-reactive immunity to *S.* Typhi antigens. These sRNAs were mostly miRNAs. sRNA changes restricted to Vi-vaccinated individuals mainly occurred in sRNA-dfs and presumably represent a

downstream effect of Vi-specific antibody-mediated effector functions or mobilisation of Vi-specific B-cell responses—this fits with the CytoF data showing changes in innate and adaptive responses 24 hours after challenge, and a modest association between vaccine status and PBMC profiles. Despite being vaccine-associated, no sRNAs were associated with protection after challenge within Vi-vaccinated groups. Lack of association with outcome does not necessarily rule out the possibility that these responses contribute to protection; it may just mean these ubiquitous responses are insufficient by themselves to prevent typhoid.

Post-challenge differences between Vi-PS and Vi-TCV groups were only present 7 days after challenge in non-diagnosed participants. Interestingly, miRNAs commonly identified in infection studies were DE with upregulation of hsa-miR-146b-3p and hsa-miR-150-3p in the Vi-TCV arm compared with the Vi-PS arm. These miRNAs are highly expressed in lymphocytes but are not lineage-specific[311,367]. Interestingly, terms relating to T-cell mediated immunity were negatively enriched in Vi-TCV vaccinees; this suggests vaccination might influence T-cell immune responses to *S.* Typhi. However, if this is the case, T-cell-related vaccine effects must be indirect, given Vi is not expected to induce T-cells because it is a polysaccharide, and T-cells respond to proteins (though exceptions exist) [61].

6.5.2.3 Predicting challenge outcome after early challenge and the problem of multiple signals.

No sRNAs were associated with challenge outcome, but average sRNA profiles were systematically more perturbed at 12 hours and D1 in the non-diagnosed group. This phenomenon was also seen in whole blood RNA sequencing data, suggesting that a more robust systemic response to *S.* Typhi is protective. The association between perturbation in sRNA expression and outcome does not relate to vaccine-mediated protection because it was also present in the control arm. Instead, the association could stem from innate or adaptive immune responses from previous exposure to antigens similar to those on *S.* Typhi. This finding, therefore, provides a novel insight into the early stages of typhoid pathogenesis.

The lack of vaccine-related correlates of protection identifiable within the first 24 hours of challenge could mean several things: 1) protective vaccine responses are not reflected at the sRNA level, 2) protective vaccine-related immunity comes into play more than 24 hours after challenge, e.g., after systemic dissemination, 3) the interaction between host susceptibility to typhoid, response to vaccination, and possibly chance, obscure vaccine-related protective signals. If baseline susceptibility to typhoid were better characterised, vaccine-related correlates would be easier to identify. Having said that, baseline correlates of susceptibility are challenging to identify in VAST because of independent signals of vaccine-mediated protection in the Vi-vaccine arms and the small number of innately resistant individuals in the control arm.

6.5.2.4 Acute typhoid and sRNA expression

Typhoid profoundly affects sRNA expression in PBMCs irrespective of vaccination. The most upregulated miRNA in all vaccine groups was hsa-miR-1303. *Mycobacterium Tuberculosis* (mTB) is the only infectious disease associated with hsa-miR-1303 in published studies—it is upregulated in PBMCs from TB patients compared with healthy control.[464] This is interesting because mTB is another facultative intracellular pathogen. According to miRNA cell expression atlases (FANTOM5 and the blood atlas), hsa-miR-1303 is lowly expressed in granulocytes and lymphocytes and highly expressed in monocytes[311,367]. CyToF data showed an increased expansion of macrophages at diagnosis, potentially explaining hsa-miR-1303 upregulation. It would be interesting to measure the expression of hsa-miR-1303 within PBMC cell subsets at D0 and typhoid diagnosis to confirm which cells produce it and identify whether intracellular levels increase.

It was possible to identify a miRNA signature that could differentiate health (i.e., D0 samples) and *S*. Typhi-exposed participants who were well from those with clinical infection. A clinically useful diagnostic signature needs to distinguish typhoid from other differential diagnoses, so the utility of this finding is limited. The merit of this finding is that it shows consistent changes are observable in acute typhoid despite host heterogeneity to *S.* Typhi infection. This consistency is essential if a miRNA-based signature of Typhoid is to be found that distinguishes typhoid from other infections.

6.5.2.5 Relating findings to the literature

This study is the first to profile miRNAs in PBMCs in Typhoid. Existing literature relates only to *S.* Typhimurium, which is used in mouse models of invasive *Salmonella* infection because it leads to a typhoid-like disease (in humans, *S.* Typhimurium is associated with gastroenteritis and is not usually invasive). Those studies implicated hsa-let-7 and hsa-miR-15 families in *Salmonella* Typhimurium pathogenesis[465,466]. hsa-let-7a is upregulated, and hsa-let-7i is downregulated in VAST participants at D1 and typhoid diagnosis. Downregulation of let-7a has been shown to confer protection against *Salmonella* Typhimurium by relieving translational repression of the pro-inflammatory cytokines IL-6 and IL-10[465]. let-7i-3p miRNA has been shown to inhibit *Salmonella* Typhimurium replication by targeting RGS2 mRNA, which alters endolysosomal trafficking and the vascular environment[467]. These studies suggest changes in let-7 miRNAs after challenge in VAST (if intracellular and unrelated to PBMC population change) would make cells more susceptible to infection. Another group of connected studies showed the miR-15 family prevents cell infection by *S.* Typhimurium in a human epithelial cell line (HeLa-229), a human intestinal epithelial line (HT-29) and a mouse macrophage-like cell line (RAW 264.7)[466]. They also found *S.* Typhimurium downregulates hsa-let-7i-3p and hsa-miR-15 family by inducing cells to secrete HMGB1, which downregulates the E2F1 transcription factor in neighbouring cells and results in downregulation of miRNAs. This increases cellular sensitivity to infection[468]. In VAST, miR-15b was upregulated at D1 and diagnosis, but the contextual difference between circulating PBMCS and epithelial tissue infected with *S.* Typhimurium limits meaningful comparison.

6.5.3 Study strengths

A key strength of this study is the controlled nature of the RCT set within a human challenge model. This allowed unbiased comparisons between Vi-vaccines and those who did and did not develop typhoid. The study also permitted evaluation of sRNA expression over time, including early after *S.* Typhi exposure. No previous study has evaluated the effect of the early innate and primed adaptive responses to a pathogen in man at the miRNA/sRNA expression level. Capturing this early time point, which may be the principal time point for understanding susceptibility to typhoid, is only possible in a human challenge model.

6.5.4 Study limitations

The control group was used as a validation cohort for post-challenge comparisons, which means Vi-vaccine-mediated effects cannot be validated. Another limitation relates to sRNA-seq of bulk cell populations because it is difficult to infer whether changes relate to differences in cell populations or intracellular expression—for example, CyTof data showed a reduction in NK cells and plasmacytoid dendritic cell, an increase in Th17 and cytotoxic T17 cells and an increase in memory B-cells homing to lymph nodes at D1 post-challenge.

Another limitation is a lack of overlapping timepoints in all datasets. The absence of CyToF or mRNA data 10 days after vaccination makes it hard to interpret miRNA correlates of vaccine-mediated protection in terms of cellular and humoral immunity. The strong miRNA signal may relate to cellular or mRNA correlates that have been missed. Future work should integrate the miRNA,

mRNA and CyToF data to look for correlation patterns across paired samples and use this to infer what may be occurring at the whole blood mRNA level and cellular composition of PBMCs 10 days after vaccination.

Despite their value, challenge models also have limitations. Complex set-up and risks, limit participant numbers. A high infection rate is needed in non-vaccinated participants to ensure adequate study power with these limited sample sizes, so a significant bacterial load must be ingested. This provides a very stringent test of vaccine efficacy, which may be artificially high compared with real-world *S*. Typhi exposure. Another limitation of VAST is that it was conducted in a non-endemic setting. Differences in the environmental and genetic background of individuals in endemic settings versus those of VAST participants could limit the generalisability of findings. Nevertheless, the preliminary efficacy data emerging from the Vi-TCV trial and the Oxford vaccine group are remarkably similar, potentially dispelling reservations around some of the aforementioned limitations.[469,470]

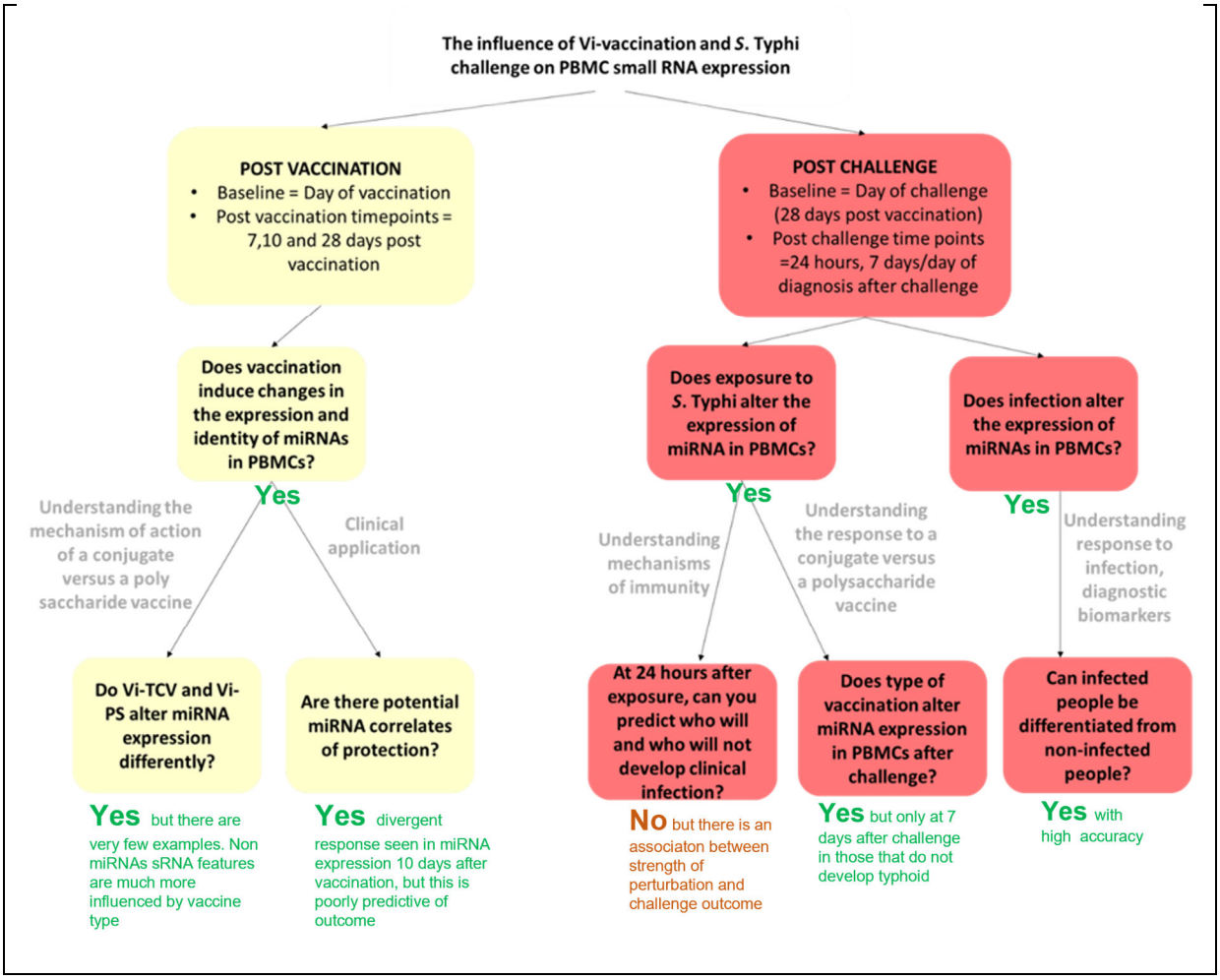


Figure 88 A summary of answers to the research questions posed at the outset of this Chapter.

6.6 Conclusion

This study has added biological insight into some of the challenges regarding typhoid set out in the chapter introduction. Firstly, vaccine failure appears to be linked to miRNA expression 10 days after vaccination. If these correlates translate to other vaccines and into mice, they could be used to prioritise the most promising vaccine candidate from an assortment of options during pre-clinical models. It would also be useful to understand the underlying mechanism driving the expression of miRNA correlates of protection to determine whether tweaks to vaccine design can improve vaccine effectiveness.

Secondly, heterogeneity in typhoid susceptibility appears linked to immunological hypo responsiveness to *S.* Typhi exposure, indicating that development of typhoid is not solely linked to chance. Thirdly, a miRNA signature of acute typhoid exists, but further work is required to determine whether it is specific to the disease. One limitation of this study was combining the sRNA and mRNA data as they come from different sample types. In the next chapter, sRNA-seq and standard RNA-seq were both conducted on whole blood in the context of a COVID-19 vaccine trial, facilitating inferences surrounding miRNA expression and mRNA post-transcriptional regulation.

7 The influence of COVID-19 on sRNAs in whole blood in the context of a ChAdOx1 nCoV-19 trial

7.1 Chapter preface

This chapter relates to work done in response to the COVID-19 pandemic. This work aligns with a University-wide refocus of work on COVID-19 and fits with University guidance for DPhil studies during the pandemic.

7.2 Introduction

On 11th March, the WHO declared the SARS-CoV-2 outbreak a pandemic[471]. Vaccines offered a vital way out of the pandemic. One of the first licenced vaccines in the UK was ChAdOx-1/nCoV-19 (initial tradename AZD1222, current tradename Vaxzervria; AstraZeneca), a recombinant viral vector-based vaccine against SARS-CoV-2 spike protein. Samples from the first clinical trials of ChAdOx-1/nCoV-19 provided a unique opportunity to study the influence of SARS-CoV-2 infection on sRNAs in whole blood in naïve individuals and recipients of ChAdOx-1/nCoV-19. This chapter reports that work.

7.2.1 SARS-CoV-2 and COVID-19

SARS-CoV-2 is a novel coronavirus. Coronaviruses are enveloped viruses with a positive-sense single-stranded RNA[472]. It is hypothesised that SARS-CoV-2 possibly originated in bats and crossed to humans via an intermediary animal[473].

SARS-CoV-2 is transmitted via exposure to virus contained in body fluids, on physical surfaces and in aerosols generated from speaking, sneezing and coughing from both symptomatic and asymptomatic infectees[474475]. Gold standard diagnosis is by a nucleic acid amplification test of a nasopharyngeal swab; however, rapid antigen tests (lateral flow tests) that detect nucleocapsid proteins are also widely used.

Infection varies from asymptomatic to fatal. Symptoms used to screen people for possible COVID-19 disease in the UK were a continuous cough, fever and loss of taste or smell. Common but less specific symptoms were headache, malaise, sore throat, coryza and diarrhoea[476,477]. Severe illness and death are associated with older age and co-morbidities, particularly obesity, diabetes and immune suppression[478]. In the UK, socio-demographic risk factors, including male sex, non-Caucasian ethnicity and deprivation, were also linked to severe illness[478–480].

The consequences of the COVID-19 pandemic are enormous. As of mid-September 2022, the John Hopkins study estimates 6.53 million COVID-19 deaths and 611 million cases worldwide[481]. Those who survive serious illness can experience sequalae from organ damage or increased frailty. Another sequalae, even after mild illness, is "long COVID", a poorly understood syndrome similar to other post-viral illnesses. The negative impacts of COVID-19 extend beyond direct morbidity and mortality, for example, secondary morbidity and mortality due to patient hesitancy and difficulty accessing healthcare. Lockdowns were crucial for managing infection rates and overwhelming pressure on UK hospitals; however, they led to a downturn in economic activity, had a negative influence

on mental health and caused adverse social effects on vulnerable people due to social isolation and a lack of the usual safety netting environment of schools and services that support vulnerable children[482–484]. Many raised concerns about the social and educational impacts of lockdown on babies through to teenagers due to deprivation of normal social interactions and increased barriers to quality education with possible long-term effects. The key to ending lockdowns was introducing highly effective vaccines, one of which was ChAdOx-1/nCoV-19.

7.2.2 ChAdOx-1/nCoV-19

ChAdOx-1/nCoV-19 was developed at the Jenner Institute and Oxford Vaccine Group at the University of Oxford. The vaccine comprises a recombinant, replication-deficient chimpanzee adenovirus that encodes the full-length wild-type SARS-CoV-2 spike protein. This antigen was selected based on previous work on coronaviruses, which showed spike protein mediates viral entry into cells through its receptor binding domain (RBD), is conserved and immunogenic[485].

Over 3 billion ChAdOx1 nCoV-19 doses have been given worldwide, including low- and middle-income countries, saving an estimated 6.3 million lives[486]. ChAdOx1 nCoV-19 has been a resounding success.

7.2.3 The ChAdOx-1/nCoV-19 trials

The Oxford vaccine group ran the first participant and observer-blinded, randomised controlled trials of ChAdOx1 nCoV-19 to establish its safety,

immunogenicity and efficacy. Subsequent phases became multi-centre international trials. These trials were collectively known as the "COV" studies. In the COV studies, participants were randomly assigned to receive either ChAdOx1 nCoV-19 or a control vaccine (MenACWY in the UK). Participants were followed up to see whether they developed COVID-19. Figure 89 contains an overview of the central study schema of the COV trials. Appendix 9.6.1 contains a summary of the COV studies pertaining to this chapter. Samples from these studies provided a unique opportunity to explore the immune response in COVID-19 and the effect of prior ChAdOx1 nCoV-19 vaccination upon this response.

7.2.4 A multi-omics study to assess the influence of ChAdOx-1/nCoV-19 on acute COVID-19

There were initial theoretical concerns regarding the possibility of vaccine-enhanced disease by ChAdOx-1/nCoV-19, which stemmed from the tragic 1960s RSV infant trial (see introduction 1.2.1.2.3.1)[66,67]. Preclinical models raised similar concerns. For example, unadjuvanted and alum-formulated vaccines based on spike protein and inactivated whole-virion vaccines were associated with increased lung eosinophilic immunopathology in SARS-CoV-infected mice[487]. Furthermore, a study in macaques implicated anti-spike IgG in increased lung pathology[488]. An advantage of the ChAdOx1 platform is that it induces T-cell mediated immunity and generates TH1-skewed responses, so because of this, the risk of vaccine-enhanced disease was felt to be very low[489]. In addition, non-human primate studies reporting prior to commencement of the in-human studies did not show safety concerns[490]. Prospective participants were counselled about this risk nonetheless. Before the efficacy outcome of the COVID-19 studies

was known, an exploratory safety study was set up to assess the immune response in COVID-19 in ChAdOx1 nCoV-19 vaccinees compared with control vaccinees using samples from the phase I to III COV001 and COV002 studies. This study involved whole-blood mRNA and sRNA-seq and serum cytokine profiling in participants with COVID-19 infection after either ChAdOx1 nCoV-19 or MenACWY vaccination. This study will be referred to as the COVID-19 Vaccine and infection multi-omics study (CoVI-multi-omics study).

The first efficacy results were published after starting the CoVI-multi-omics study, showing that ChAdOx1 nCoV-19 was safe and effective. Nevertheless, the exploratory study continued as it allowed several questions to be answered in the gold-standard setting of an RCT: firstly, what does the immune response during COVID-19 look like at the molecular level compared with health and illnesses with similar symptoms to COVID-19 and how does it evolve over the course of the infection? Secondly, does ChAdOx1 nCoV-19 influence the immune response during COVID-19?

7.2.5 An opportunity to understand the role of sRNAs in the immune response to viruses and vaccine-mediated responses during infection.

Little is known about the role of miRNAs in SARS-CoV-2 infection *in-vivo*. A handful of small-scale sRNA-seq studies have been done, but many were underpowered to find statistical differences after multiple testing corrections[491–498]. Furthermore, no studies compare COVID-19-specific immune responses to

other illnesses with a similar symptom profile, nor could they provide information on how vaccination influences the immune response[491–498].

Another advantage of the CoVI-multi-omics dataset is that, compared with the EUCLIDS and VAST studies presented in this thesis, this study profiles sRNAs and mRNAs from the same samples (whole blood). This improves the comparability of the two omics datasets, meaning integration could provide greater insights into the role of miRNA (and possibly other sRNAs) on post-transcriptional regulation during an immune response.

7.3 Aims

Chapter aims are to address the following research questions:

1) How does COVID-19 influence whole-blood expression of sRNAs compared with health and COVID-like illness?
2) Does ChAdOx1 nCoV-19 vaccination influence the above in break-through infection?
3) Is there evidence of transcriptional regulation by sRNAs on the transcriptome during SARS-CoV-2?

7.4 Methods

7.4.1 Samples and study set-up

Study samples from COV001 and COV002 came from baseline, symptom onset (S_0), and 7 days post symptom onset (S_7) visits (see Figure 89, Figure 90). S_0

visits were triggered when participants developed either cough, fever, or loss of smell or taste. At this visit, a COVID PCR test and blood samples were taken. These tests were repeated 7 days later at the S_7 visit.

The CoVI-multi-omics study had two stages (Figure 90). The initial stage had paired samples at baseline, S_0 and S_7, from people who did and did not have SARS-CoV-2 infection ("symptomatic controls") (Figure 90A). One symptomatic episode was included for each person, regardless of whether that episode was positive for SARS-CoV-2. Ideally, samples from COVID-19 and non-COVID-19 presentations would have been collected from the same people, but there were insufficient participants with both episode types. This stage allowed samples from acute COVID-19 ($S_0^{COVID-19}$ and $S_7^{COVID-19}$) to be compared with health (baseline) and COVID-like illnesses (S_0^{PCR-ve} and S_7^{PCR-ve}) and comparison of acute COVID-19 samples between vaccine groups (ChAdOx1 nCoV-19 vs menACWY).

The second stage of the CoVI-multiomics study focussed on health versus acute COVID-19, and ChAdOx1 nCoV-19 versus menACWY during acute COVID-19. It did not contain samples from symptomatic controls (Figure 90A).

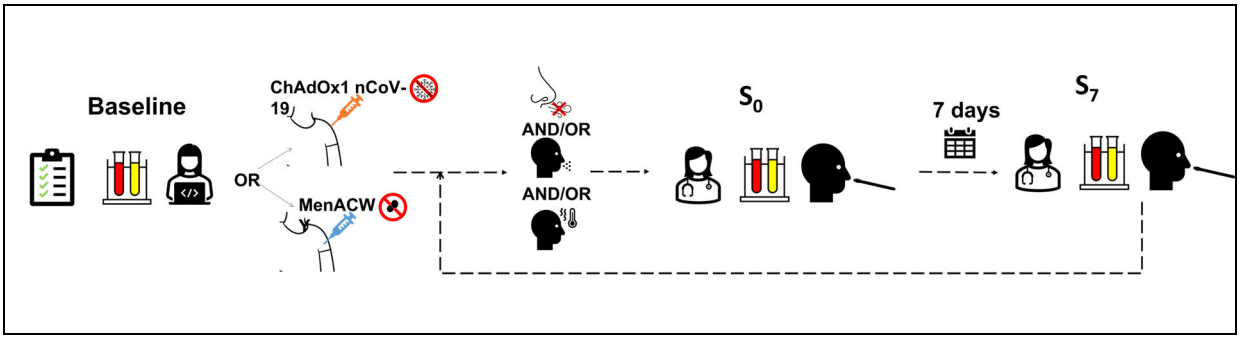


Figure 89 Visit schema. At baseline visits, participants were enrolled, blood was taken, and then they were randomised to receive ChAdOx1 nCoV-19 or MenACWY. Participants were then asked to contact the clinic if they had any key COVID-19 symptoms: loss of smell or taste, cough or fever. They then attended a symptomatic visit (S_0) where they were medically assessed, blood was taken, and a PCR COVID-19 test was done. Seven days later, the participants reattended for a repeat medical assessment, blood tests and another COVID-19 PCR test. Participants returned to standard follow-up until their next onset of COVID-like symptoms when they could have another S_0 and S_7. Not all participants were positive for SARS-CoV-2 at these visits. This means blood samples were available for participants with COVID-like symptoms who did (PCR +ve) and did not (PCR-ve, i.e., those with COVID-like illnesses) have a SARS-CoV-2 infection. Stage II visits were the same as stage I, except that most participants did not have a baseline sample as this ceased to be collected in later enrolments of the COV studies.

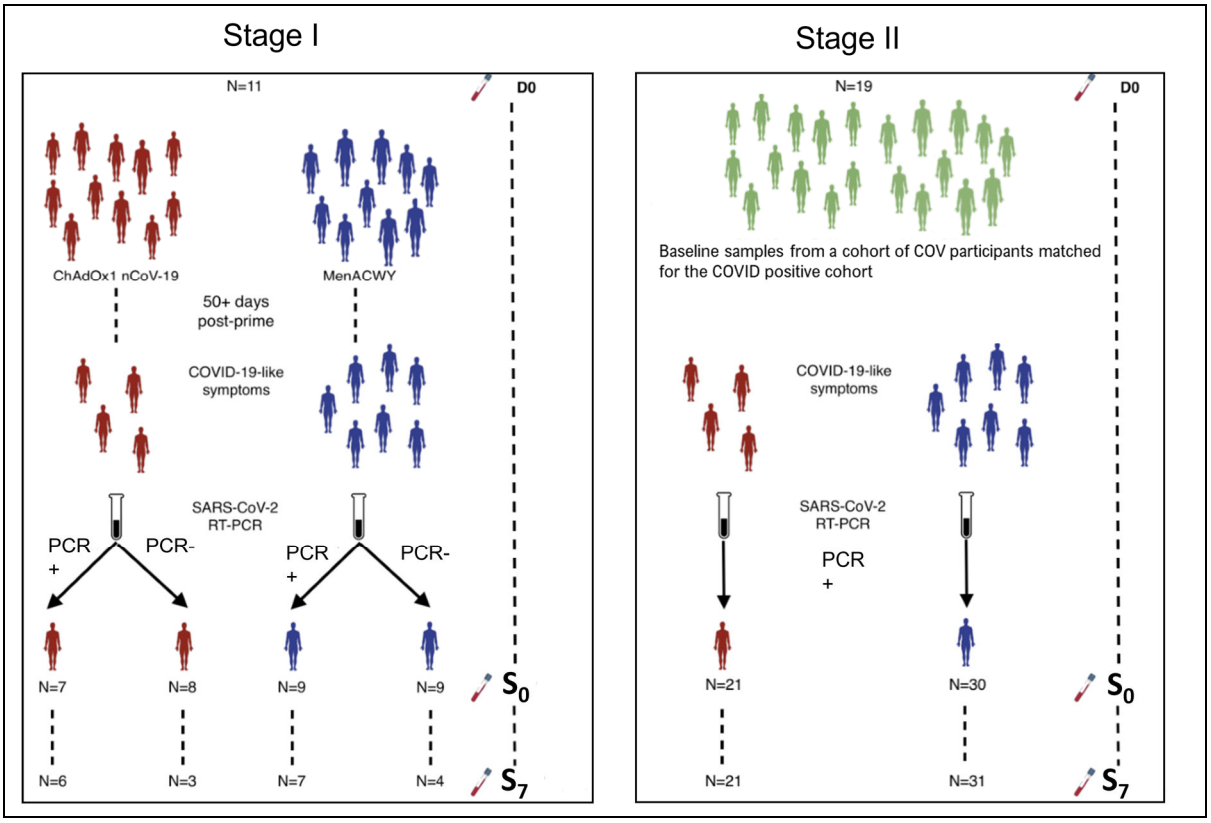


Figure 90 Schema of the two phases of this study amended from the study plan. Stage I had paired samples at baseline, S_0 and S_7, from people who did and did not have SARS-CoV-2 infection ("symptomatic controls"). This allows a comparison of acute COVID-19 versus health and illnesses with COVID-like symptoms. It also allowed the comparison of acute COVID-19 samples by vaccine type (ChAdOx1 nCoV-19 vs menACWY vaccine). The study's second phase focuses on the difference between health and acute COVID-19 and the difference between vaccine groups during acute COVID.

7.4.2 sRNA specific methods

7.4.2.1 sRNA-seq

RNA was extracted from whole blood collected in Paxgene tubes. See General Methods, section 2.2. In terms of validation—an FDR<0.05 in stage II was deemed validated by a p-value of <0.05 in stage I as the stage II cohort was bigger.

7.4.3 mRNA sequencing data

Others generated and analysed the standard RNA-seq data, but in brief, a ribodepletion library preparation was used to preserve coding and ncRNAs. cDNA libraries were paired-end sequenced (NovaSeq6000). rRNA, sex chromosome genes, mitochondrial RNA and haemoglobin genes were excluded from downstream analysis. Differential gene expression analysis was conducted in limma.

7.4.4 Cytokine and chemokine data

Cytokine and chemokine data were generated and analysed by others. In brief, a 20-plex Luminex panel (MILLIPLEX MAP Human Cytokine/Chemokine Magnetic Bead Panel—Immunology Multiplex Assay, Merck Millipore) measured the serum concentration of IL1β, IL2, IL3, IL4, IL5, IL6, IL8, IL10, IL12-p70, IL13, IL17A, IL18, sCD40L, G-CSF, GM-CSF, IFNγ, IP10 (CXCL10), MCP1, TNFα and VEGF-A. Wilcoxon tests (paired where possible) were used to test for significant differential expression and results subject to an FDR correction.

7.4.5 MiRNA pathway enrichment analysis

The General Methods Chapter details MiRNA pathway and miRNA set enrichment analyses (Chapter 2, section 2.7.2) and explains how to interpret the results of such analyses.

7.5 Results

7.5.1 COVI multi-omics Stage I sRNA results

7.5.1.1 Sample numbers and demographics

Sample numbers and participant and sample level information are contained in Table 24, Supplementary Table 11 and Supplementary Table 12. The majority of baseline samples were in the COVID-19-negative group. This was because SARS-CoV-2 infection was rare in the COV001 and COV002 trials at the point of sample selection. Therefore, most of the COVID-19-positive samples came from the large numbers of participants enrolled in COV002 when baseline samples were not taken.

Median and interquartile range of the CRP results were very similar between those with COVID-19 (median 4 mg/L, IQR: 4-7 mg/L, range 0.4-37 mg/L) and those with COVID-like illness (median 3 mg/L, IQR: 1-8 mg/L, range 0.3-77 mg/L) see Supplementary figure 12.

Table 24 Stage 1 sample numbers. Total number per subgroup in bold. Breakdown between vacccine groups in brackets (ChAdOx1 nCoV-19 : MenACWY) The vaccine group for the COVID-19 negative group is included for completeness; however, it is irrelevant to this analysis.

	D0		CT		CT7	
	Luminex	**sRNA + mRNA seq**	**Luminex**	**sRNA + mRNA seq**	**Luminex**	**sRNA + mRNA seq**
PCR +ve	**3** (2:1)	**3** (2:1)	**16** (7:9)	**16** (7:9)	**13** (6:7)	**13** (6:7)
PCR -ve	**8** (3:5)	**8** (3:5)	**17** (8:9)	**13** (5:8)	**7** (3:4)	**7** (3:4)
Total	**11**	**11**	**33**	**29**	**20**	**20**

7.5.1.2 The sRNA composition of whole blood samples

See Supplementary figure 7, Supplementary figure 8, and Supplementary Table 9 for sample QC plots. The stage I sequencing was of lower quality with smaller library sizes than usual, which were very variable compared with other studies in this thesis. This variability can be partially mitigated through normalisation; nevertheless, this noise may influence results, making it harder to discern differences between groups and timepoints.

The number of features pertaining to each class of RNA can be found in Table 25. The most abundant class are offset miRNAs (sequences that overlap with the miRNA hairpin) (Figure 91), followed by yRNAs and rRNAs. See Supplementary Table 15 for how the relative abundance of each RNA type compares across the substrates sRNA-sequenced in this thesis. The absence of significant differences in sRNA class abundances across time points means all features can be analysed together in a single differential expression analysis.

Table 25 Number of features pertaining to each class of RNA after noise removal

RNA class	Number of features
mature miRNA	337
offset miRNA	76
tRNA	38
snoRNA	154
snRNA	9
yRNA	3
lncRNA	133
vault	4
protein-coding	177
rRNA	8
viral mIRNA	1
Antisense to known feature	118
Total Features	**1058**

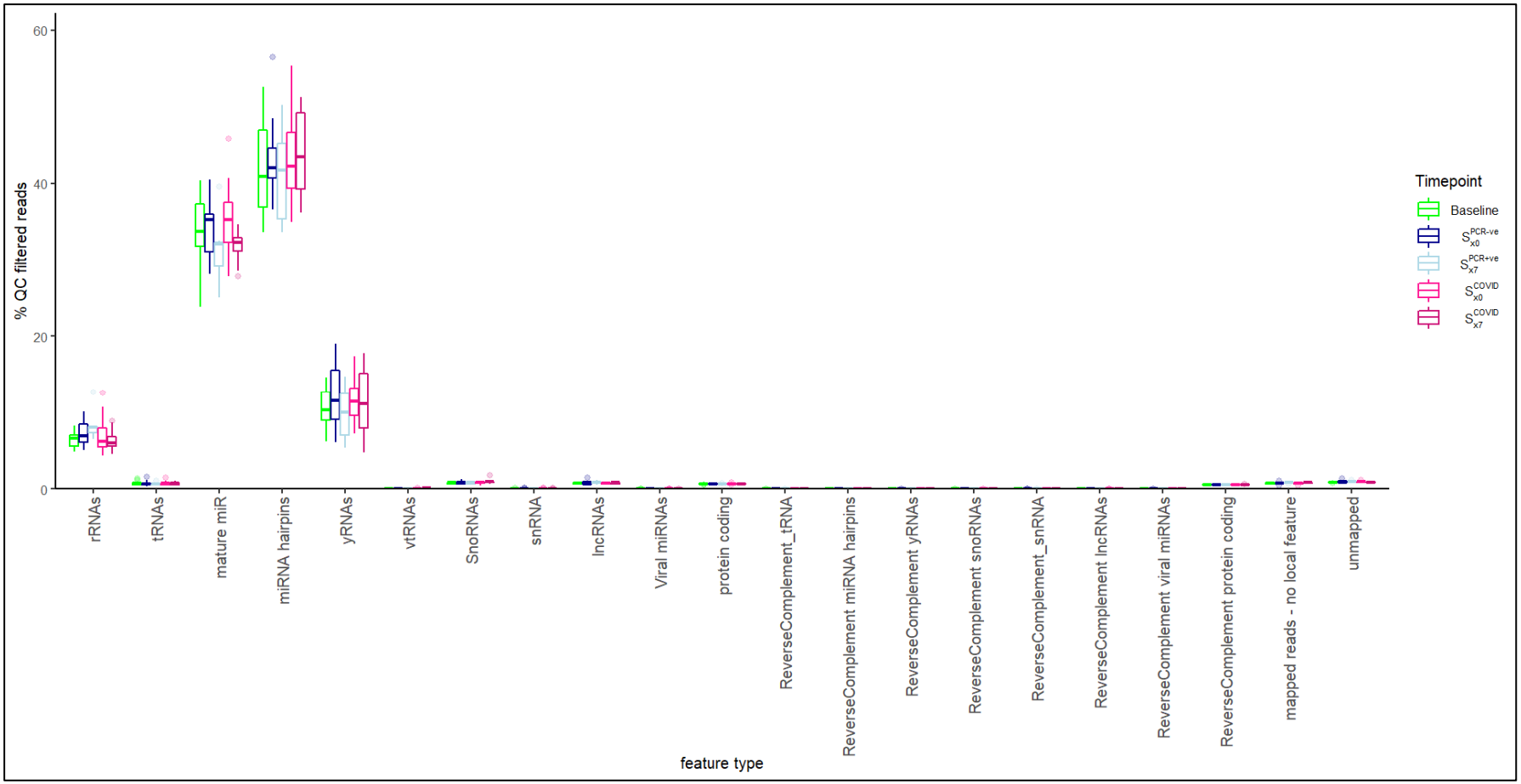


Figure 91 Distribution of sRNA-seq reads mapping to each class of RNAs. There were no significant differences between the timepoints.

7.5.1.3 Whole blood expression of miRNAs and sRNA-dfs differs between acute COVID-19 and health and COVID-like illnesses.

7.5.1.3.1 COVID-19 influences whole blood sRNA expression compared with health and over illness time-course

PCA analysis suggests COVID-19 has a limited influence on sRNA expression compared with health and COVID-like illness, as sample groups generally do not cluster discretely in the first 4 components (Figure 92).

Differential expression analyses, however, revealed COVID-19 influences sRNA feature expression compared with health, with 1 and 4 sRNAs DE between baseline and $S_0^{COVID+ve}$ and $S_7^{COVID+ve}$ visits in the menACWY group, respectively (see Figure 93, Appendix Table 11 and Appendix Figure 45). COVID-19 also influence sRNA expression over time with 7 sRNAs DE between $S_0^{COVID+ve}$ and $S_{7\text{-}COVID+ve}$. This small sample size meant these results were explored further in a larger independent cohort in stage II.

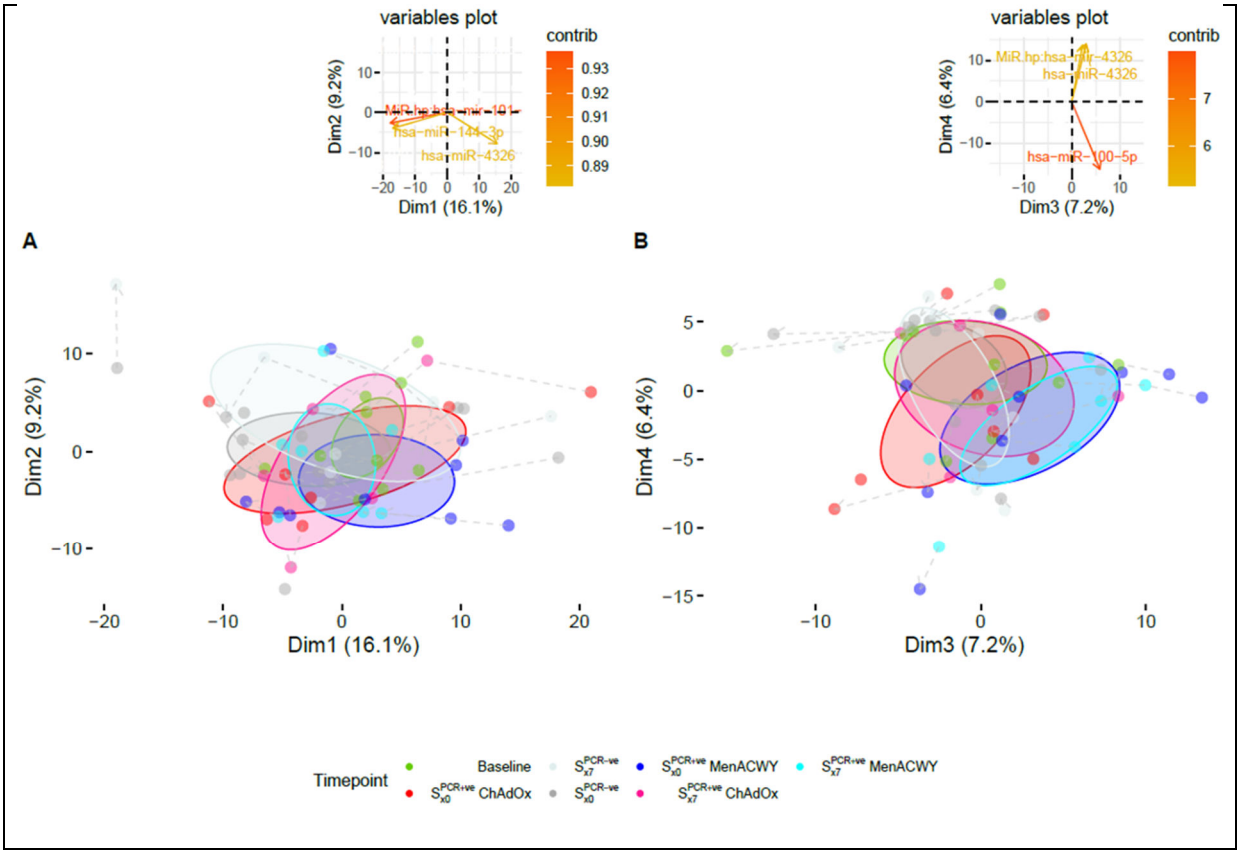


Figure 92 PCA analysis of all samples analysed in stage I. Samples are coloured by timepoint and vaccine group where relevant. A dashed line connects samples from the same person. The shaded ellipses mark the 95% confidence intervals of the centroids of each sample time point, helping to indicate the significance of group separation. Variable plots in the top right-hand corners show which sRNAs contribute the most to variation.

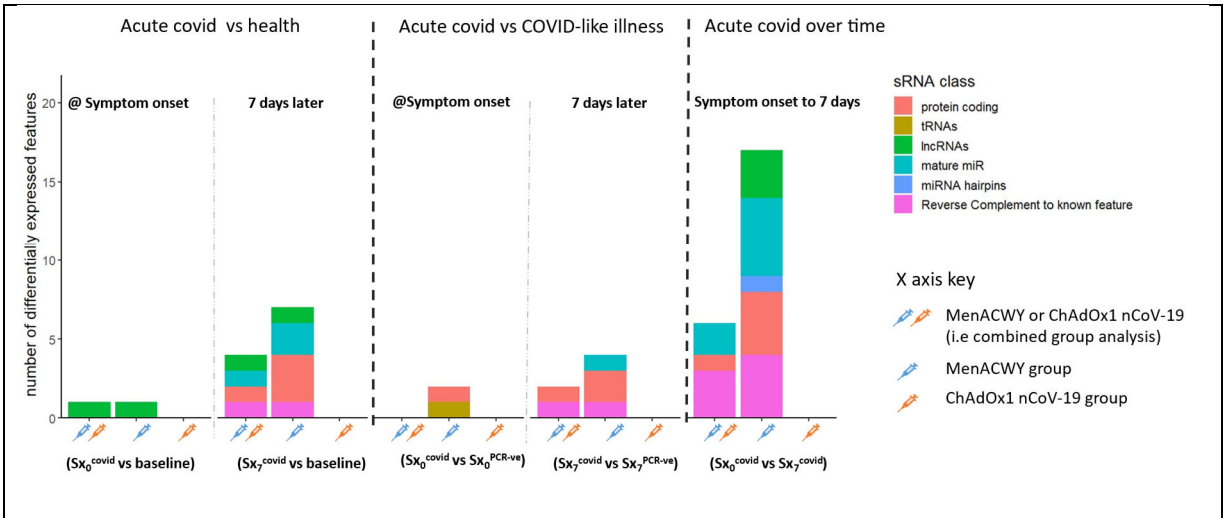


Figure 93. Summary of the number of DE sRNAs in different comparisons, broken down by sRNA class type.

7.5.1.3.2 COVID-19 influences sRNA expression compared with COVID-like illnesses with reduced miRNA targeting of immunological pathways.

COVID-19 is an unusual viral respiratory illness in terms of morbidity and mortality, and this is reflected in the different sRNAs it induces compared with COVID-like illnesses at symptom onset and 7 days into illness (Figure 93, Appendix Table 11, and Appendix Figure 45). MicroRNA pathway enrichment analyses showed pathways involved in the immune response were positively enriched for post-transcriptional regulation (inferring repression) at symptom onset in COVID-like illnesses compared with COVID-19 (Figure 94A). Many of these are related to immune processes, including B and T-cell proliferation and response, and TLR, chemokine and cytokine signalling. Analysing each disease-timepoint compared with health, reveals a clear skew towards enriched targeting of pathways in COVID-like illness but not in COVID-19 (Figure 94B). This is surprising given miRNAs were DE between baseline and $S_0^{COVID\text{-}19}$ but not between baseline and $S_0^{PCR\text{-}ve}$ (Appendix Figure 45). These results suggest a lack of co-ordinated pathway targeting by miRNAs in COVID-19.

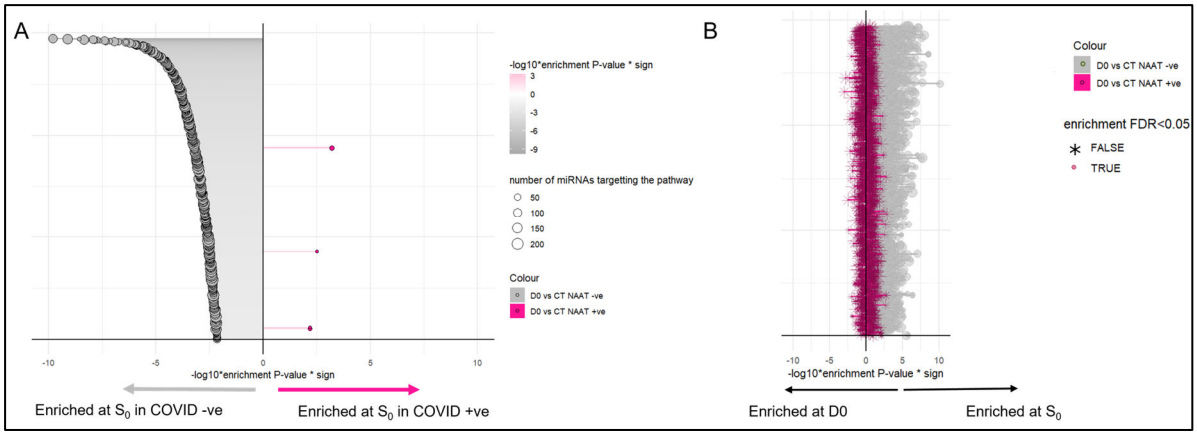


Figure 94 miRNA pathway enrichment analysis shows enrichment for targetting at symptom onset in participants with COVID-like illness compared with baseline and COVID-19. Each dot represents the result for a different pathway. Only pathways enriched at an FDR<0.05 are shown. The MIEA software does not return enrichment scores; therefore, the -log$_{10}$ of the p-values are presented alongside the direction of enrichment indicated. The shape shows whether the significance value was below 0.05 after applying fdr correction.
A: Results of the enrichment analysis comparing symptom onset in COVID-positive and COVID-like illness participants. Pathways positively enriched for targetting (i.e. pathways predicted to be subject to greater miRNA expression) in the COVID-like illness group (i.e. negatively enriched in the COVID-19 positive group) plot towards the left and are coloured grey. Pathways positively enriched for targetting in the COVID-19 positive group plot towards the right and are coloured pink. The magnitude of enrichment is represented by the distance from 0 on the X-axis and the colour of the dots and bar.
B: Enrichment analysis results comparing symptom onset to baseline in COVID-positive participants (results in pink) and COVID-like illness participants (results in grey) plotted on the same graph. Only the pathways presented in plot A are shown.

7.5.1.4 ChAdOx1 nCoV-19 ameliorates perturbation in sRNA expression during COVID-19

Interestingly, sRNA differences between COVID-19 and those with COVID-like illnesses were *only* present in control vaccinated COVID-19-positive participants at symptom onset. This suggests ChAdOx1 nCoV-19 ameliorates perturbation in sRNA expression during COVID-19. Strengthening this conclusion is the fact that more sRNAs were DE in every analysis when restricting to the control vaccine subgroup, and sRNA perturbation was greater in control subgroup analyses than in ChAdOx1 nCoV-19 subgroup analyses (Figure 93, Appendix Table 11,

Appendix Figure 45). Stage II of the COVID-19 multi-omics study was undertaken to see if these results could be validated.

7.5.2 Stage II: A larger confirmatory study comparing health with acute COVID-19 and the influence of ChAdOx1 nCoV-19 vaccination on COVID-19.

In total, 122 samples were run, from 19 healthy controls and 51 COVID-positive episodes, balanced as much as possible between vaccine groups (Table 26). Unpaired baseline samples from the COV studies ("healthy controls") were used due to a lack of available paired baseline samples. Baseline and COVID-19 samples were selected from participants matched for age, sex, ethnicity, co-morbidity and BMI to avoid systematic differences between timepoints and vaccine groups (Supplementary Table 13, Supplementary Table 14).

Table 26 Stage II sample numbers.

	Baseline		$S_0^{COVID-19}$		$S_7^{COVID-19}$	
	Luminex	**SRNA/ mRNA seq**	**Luminex**	**SRNA /mRNA seq**	**Luminex**	**SRNA /mRNA seq**
Total	**49**	**19**	**50**	**51**	**49**	**52**
ChAdOx1	18	12	19	21	19	21
MenACWY	31	7	31	30	30	31

See Supplementary Table 10, Supplementary figure 9 and Supplementary figure 10 for QC graphs. sRNA class distribution was similar to stage I, with offset miRNAs again being the most abundant sRNA type (Appendix Figure 47).

7.5.2.1 Replication of sRNA changes during COVID-19 with marked downregulation of miRNAs at symptom onset.

As seen in stage I, COVID-19 influences whole blood sRNA expression—this is more obvious in the stage II results as it can be seen in the PCA, which shows baseline and S_0 samples clustering separately with S_7 samples lying betwixt the two suggesting a general return to baseline 7 days into illness (Figure 95A:B). There were 381 sRNAs (36% of sRNAs) and 191 (18% of sRNAs) DE between baseline and S_0 and/or S_7, respectively (see Figure 95 C, Appendix Table 11 and Appendix Figure 48). Twenty-two of these were DE at $p<0.05$ in stage I results (Figure 96). The most DE RNA class was mature miRNAs (130/381 DE features), almost all of which were downregulated at S_0 (126/130) compared with baseline, suggesting a global lifting of miRNA post-transcriptional regulation at illness

onset. sRNA-dfs DE in COVID-19 arise from genes involved in metabolism and RNA silencing (see appendix 9.6.3).

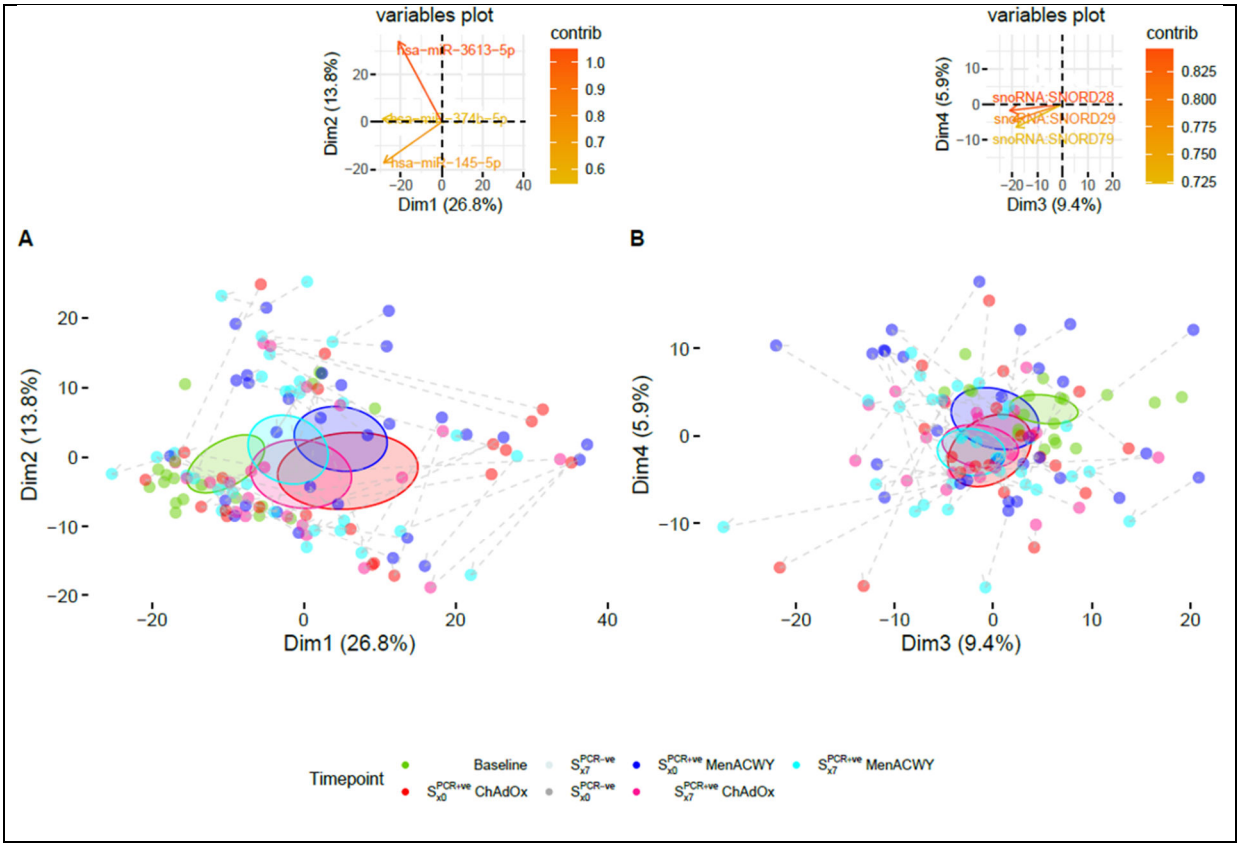


Figure 95 (continued)

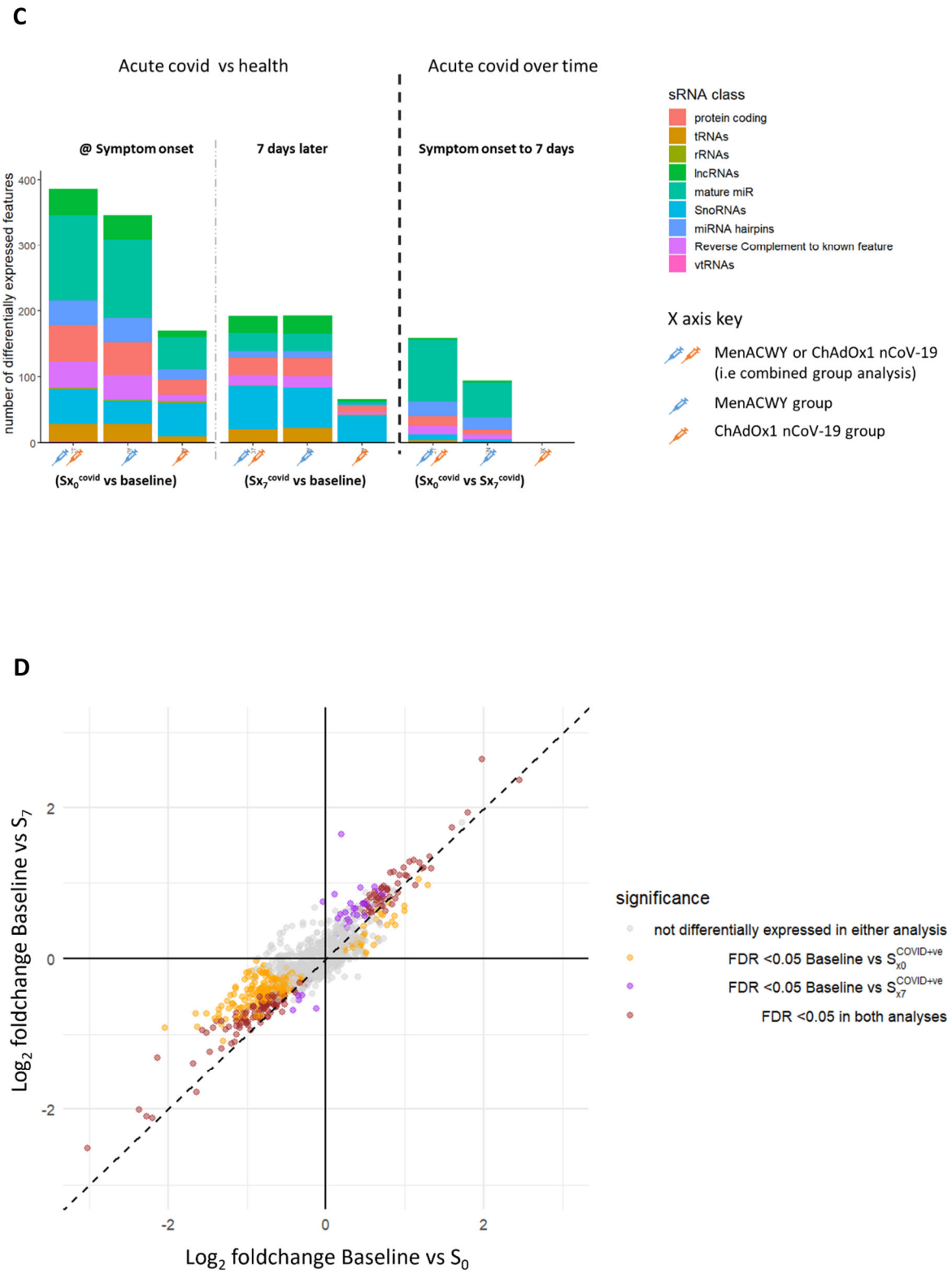


Figure 95 Stage I Differential expression results
Panels A+B: PCA analysis of all samples analysed in stage II. Samples are coloured by

Figure 95 caption (continued)
timepoint and vaccine group where relevant. A dashed line connects samples from the same person. The shaded ellipses mark the 95% confidence intervals of the centroids of each sample time point, indicating the significance of group separation. Variable plots in the top right-hand corners show which sRNAs contribute the most to variation.
Panel C: Summary of the number of DE sRNAs in different comparisons in stage II broken down by sRNA class type.
Panel D: agreement plot of the log_2 foldchange at baseline vs S_0 against baseline vs S_7. Dashed line represents y=x

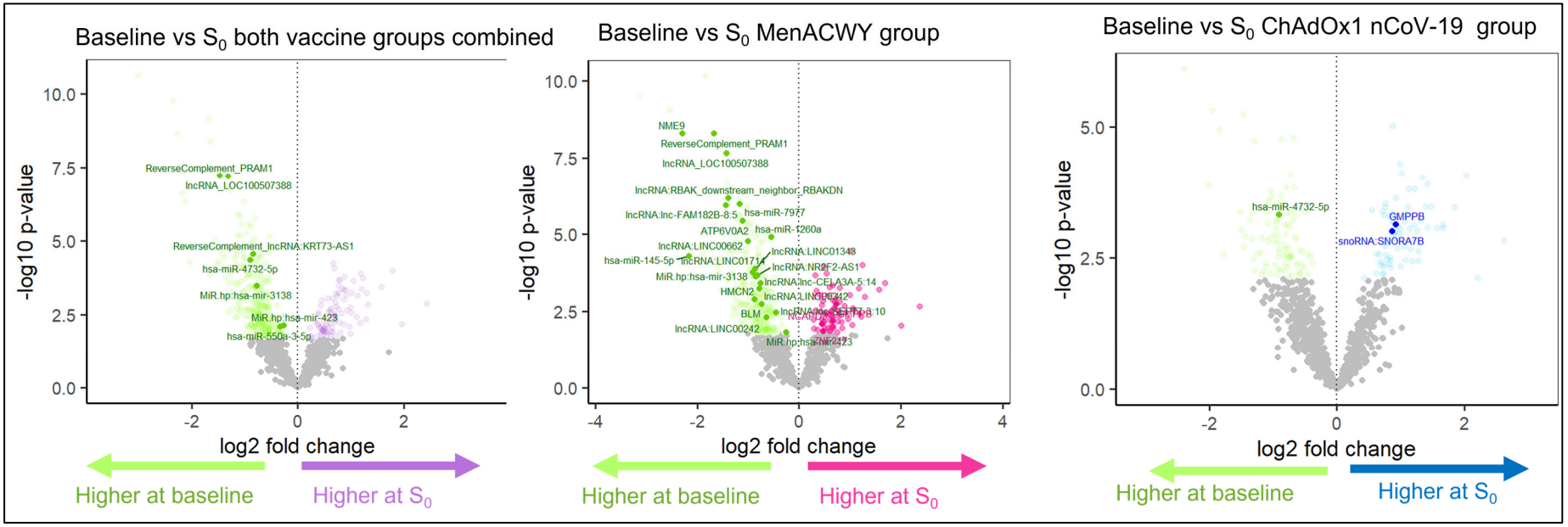


Figure 96 Volcano plots showing sRNAs that were DE at COVID-19 onset symptoms compared with baseline. Results are shown for both vaccine groups combined and split by vaccine group. Coloured dots are sRNAs with FDR <0.05. Labelled sRNAs were DE in the stage I data at p<0.05.

7.5.2.2 Changes in sRNA expression over the course of COVID-19 show resolving miRNA perturbation but increasing snoRNA perturbation (stage II results).

Compared with baseline, log_2 foldchanges at S_7 strongly agreed with S_0 (Figure 95D); however, more sRNAs were DE at symptom onset, indicating that though changes persist 7 days into infection, they are starting to resolve. This global picture, however, hides nuanced differences at the sRNA class level (Figure 97). miRNA changes were resolving by S_7—only 21% of miRNAs that were DE at S_0 remained significantly DE at S_7 compared with baseline. In contrast, differential expression of sRNA-dfs remained considerably perturbed 7 days into illness (58% were still DE at S_7 compared with baseline). In particular, snoRNA-dfs increase at S_7 compared with S_0 (Figure 95 C, Appendix Table 12, Figure 97). This latter point is reflected in the PCA, which shows S_7 samples lie furthest from baseline samples (Figure 95B) in PC4, with snoRNA-dfs contributing most to this component. In ChAdOx1 nCoV-19 vaccinees, snoRNA-dfs are already maximally upregulated at S_0, but this peak occurs later at S_7 in control vaccinees, suggesting snoRNA-df expression may be tied to the onset of adaptive immunity.

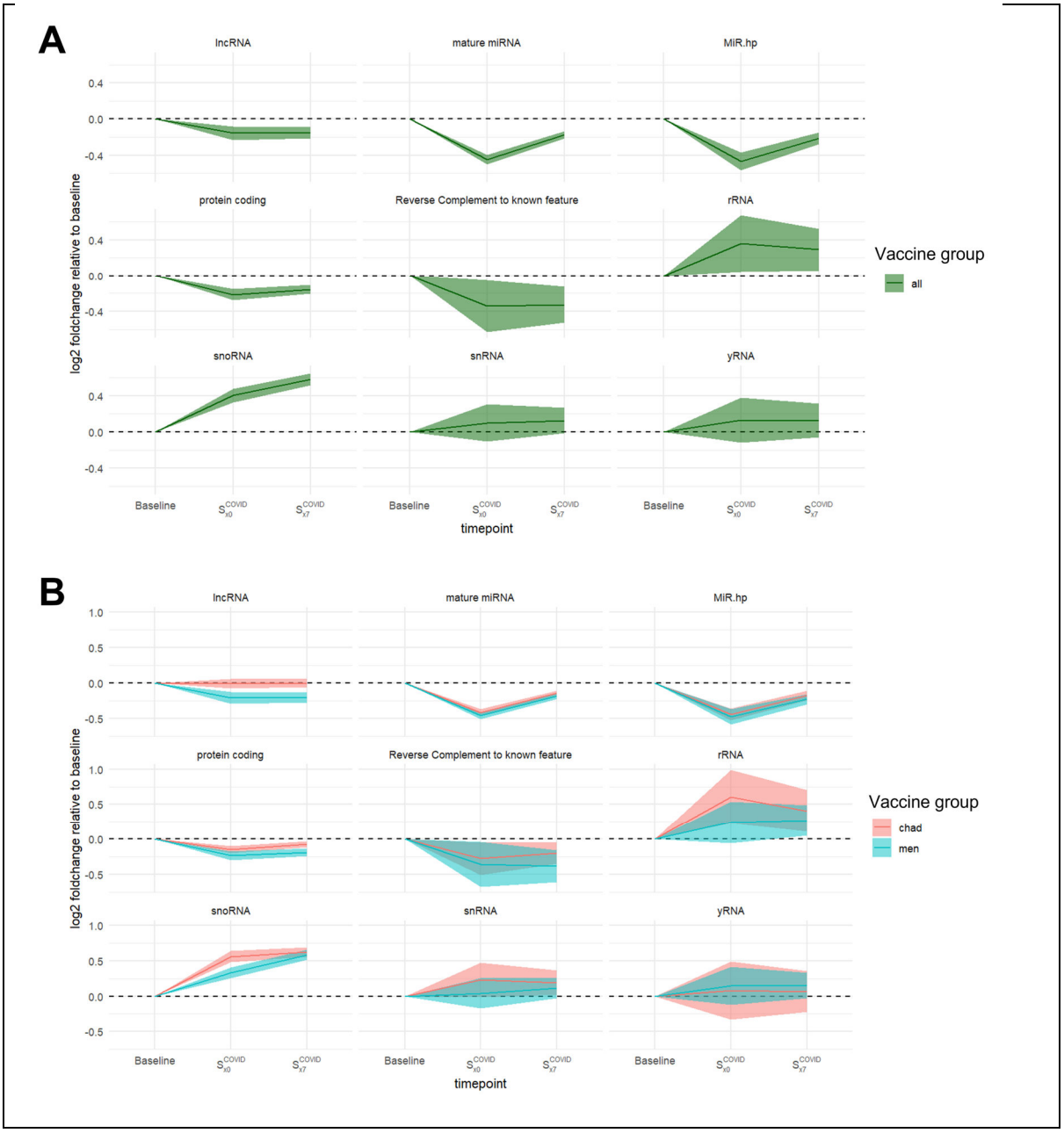


Figure 97 Class-wise changes in sRNA expression over time relative to baseline. Y-axis represents the mean log_2 foldchange of sRNAs within each class of sRNA with 95% confidence intervals shown in the shaded area. Panel A shows results split by vaccine group. Panel B shows results when both vaccine groups are analysed collectively.

7.5.2.3 Publicly available data confirm differential expression of a set of miRNAs in COVID-19

Eight studies have compared the whole blood micronome of COVID-19 hospitalised patients with healthy controls[491–498]. Public study results were

compared with baseline vs S_0 and baseline vs S_7 COVI-multi-omics results because the public studies recruited participants several days after symptom onset—i.e. nearer to S_7. Twenty-two of the 39 miRNAs DE in at least one public study and captured in the CoVI-multi-omics study were DE between baseline and S_0 and/or S_7 in the CoVI-multi-omics study at p<0.05 in the stage II study (11 had an FDR<0.05). Three were additionally DE in the stage I study (hsa-miR-6501-5p, hsa-miR-199b-5p and hsa-miR-361-3p) at p<0.05. No miRNAs were DE across all studies (Figure 98). A formal meta-analysis was not possible because the public studies reported gene lists but not associated metrics. Overall, corroboration of DE miRNAs supports the theory that COVID-19 influences miRNA expression.

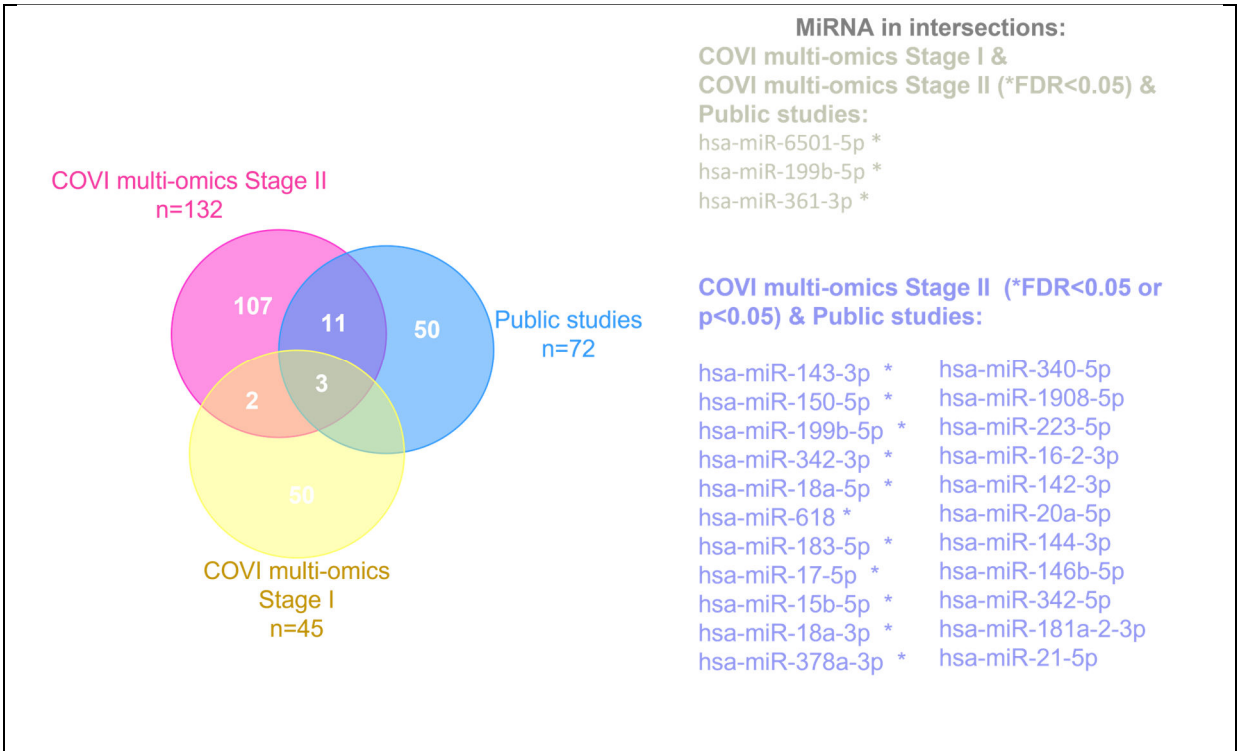


Figure 98 Overlap between miRNAs DE between baseline and S_0 and/or S_7, and miRNAs DE in publicly available datasets. An FDR of <0.05 was used as a significance cut-off for differential expression in the CoVI mutli omics studies for the Venn diagram. Differential expression in public studies was defined within those studies (usually based on a p-value of < 0.05, sometimes including a foldchange cut-off).

7.5.2.4 MiRNA repression of immunological pathways is lifted at COVID-19 onset.

To identify whether changes in miRNA expression during COVID-19 are likely to lead to increased or diminished repression of any pathways, GSEA style pathway analysis was done using MIEAA2 (General Methods Chapter 2 section 2.7.2 contains an explanation of how to interpret positive and negative enrichment).

7.5.2.4.1 Baseline versus symptom onset

Thirty-four gene ontology pathways were negatively enriched at S_0 compared with baseline, including terms relating to the immune response, e.g., TRIF-dependent toll-like receptor signalling pathway, negative regulation of I-kappa B kinase, NF-kappa B signalling, positive regulation of lymphocyte activation and cell stress responses including cellular response to hypoxia and regulation of JUN kinase activity (Figure 99*).* This should encourage upregulation of these pathways. No pathways were positively enriched. Enrichment results were similar in stage I, showing few enriched pathways. The fact that only 34 pathways were significantly enriched is surprising, as many miRNAs were DE between baseline and $S0^{COVID+ve}$. The paucity of enrichment reflects global miRNA downregulation resulting in general rather than specific pathway de-repression.

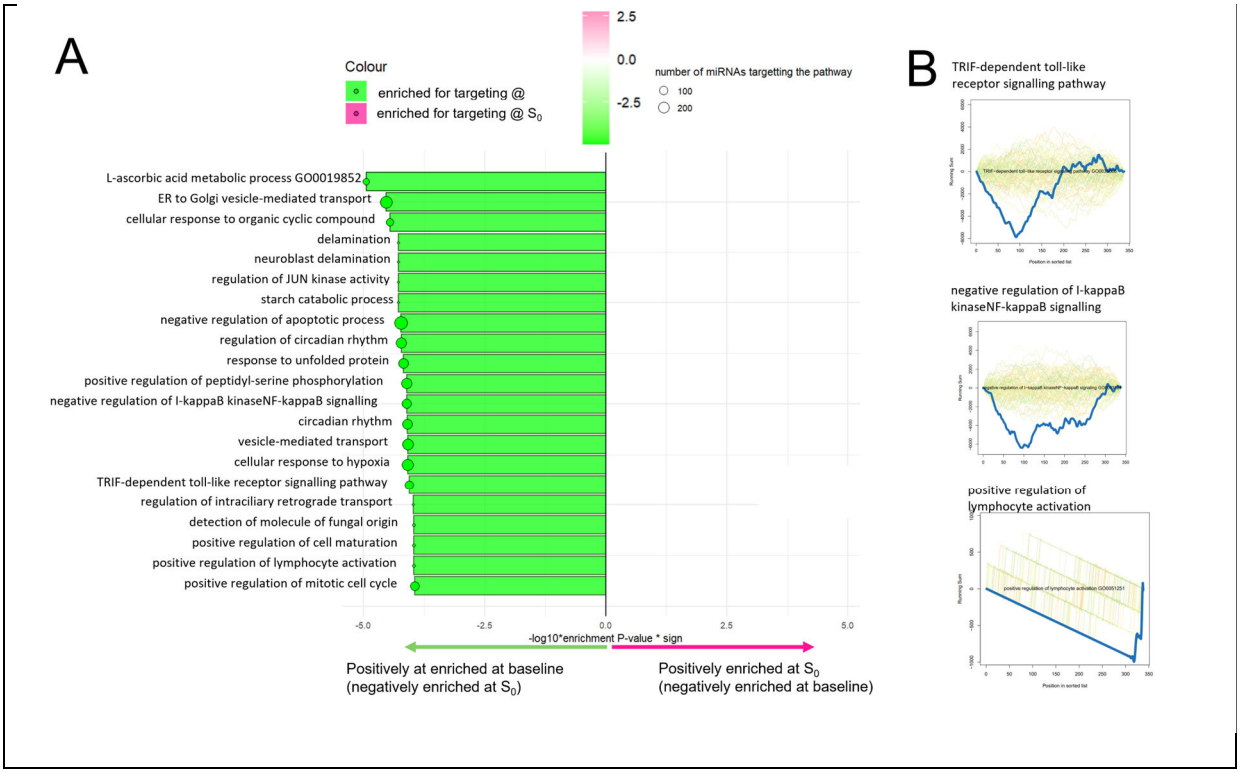


Figure 99 Enrichment results presented for significantly enriched pathways comparing symptom onset with baseline. Each dot represents the result for a different pathway. Only pathways enriched at an FDR<0.05 are shown. The MIEA software does not return enrichment scores; therefore, the -log10 of the p-values are presented alongside the direction of enrichment indicated. The shape shows whether the significance value was below 0.05 after applying fdr correction.
A: Results of the enrichment analysis comparing symptom onset in COVID-positive and baseline. Pathways positively enriched for targeting (i.e. pathways predicted to be subject to greater miRNA expression) at baseline (i.e. negatively enriched at COVID-19 symptom onset) plot towards the left and are coloured green. Pathways positively enriched for targeting at COVID-19 symptom onset plot towards the right and are coloured pink. The magnitude of enrichment is represented by the distance from 0 on the x-xis and the intensity of the colour of the dots and bar.
B: Examples of enrichment plots for three pathways are shown in plot A.

7.5.2.4.2 A co-ordinated resolution of miRNAs targeting pathways involved in the immune response occurs 7 days into COVID-19, potentially terminating the immune response (stage II results)

In the S_0 to S_7 analysis, 2,878 pathways were differentially targeted between the two timepoints. The vast majority of these pathways (2,842) were negatively

enriched, i.e., less targeted at S_0 than at S_7 (Figure 100A). These pathways pertained to cytokines (including TNF, IFN-γ, IL1, IL6, IL10, IL15, IL18), PRRs and their signalling pathways (including TLR2, 3, 4, 7 and 9), and leukocyte differentiation and activation (including macrophage differentiation, NK cell degranulation, positive regulation of B-cell activation, positive regulation of T-cell-mediated cytotoxicity and T-helper 17 cell lineage commitment) (Appendix Figure 49). The findings were similar to the stage I study.

Enrichment at S_0 versus S_7 is a relative phenomenon—are these pathways negatively enriched at symptom onset or positively enriched 7 days into illness? Given miRNAs were globally downregulated at S_0 compared with baseline and only starting to return to baseline at S_7, the most plausible scenario is global downregulation of miRNAs occurs at the onset of COVID-19, followed by selective recovery of miRNAs that target proinflammatory pathways. Average $\log_2$ foldchanges of miRNAs that target examples of negatively enriched pathways support this interpretation (Figure 100B).

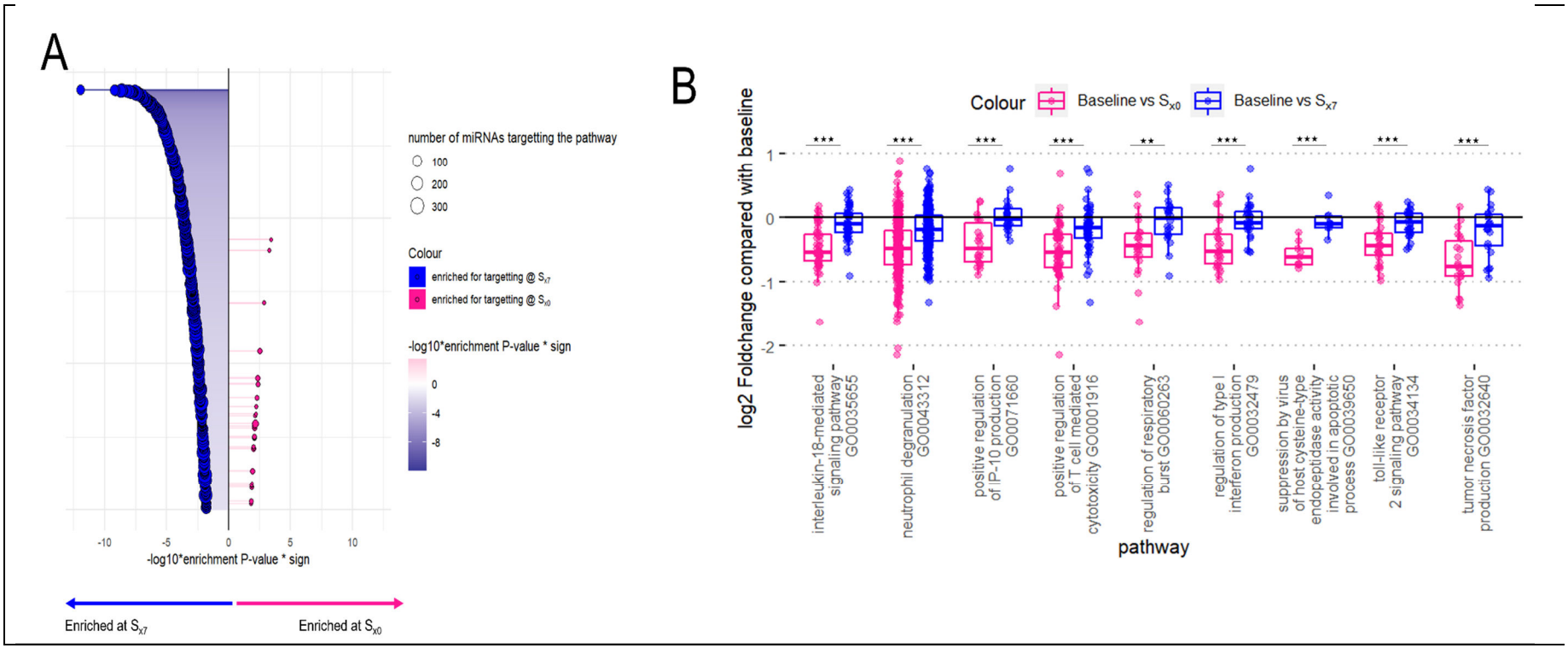


Figure 100 Enrichment results presented for significantly enriched pathways comparing 7 days into COVID-19 with symptom onset. Each dot represents the result for a different pathway. Only pathways enriched at an FDR<0.05 are shown. The MIEA software does not return enrichment scores; therefore, the -log$_{10}$ of the p-values are presented alongside the direction of enrichment indicated.
A: Results of the enrichment analysis comparing 7 days into COVID-19 with symptom onset. Pathways positively enriched for targeting (i.e. pathways predicted to be subject to greater miRNA expression) 7 days into COVID-19 (i.e. negatively enriched in the COVID-19 positive group) plot towards the left and are coloured blue. Pathways positively enriched for targeting in the COVID-19 positive group plot towards the right and are coloured pink. The magnitude of enrichment is represented by the distance from 0 on the x-axis and the intensity of the colour of the dots and bar.
B: The log2 foldchanges of miRNA targeting genes in the top pathways were visualised to understand the enrichment better.

7.5.2.5 Integration of microRNAs and mRNA analysis reveals evidence for post-transcriptional regulation during COVID-19.

If miRNA expression during COVID-19 promotes a rapid inflammatory response, there ought to be a relationship between the expression of miRNAs and the mRNAs/pathways they target. MiRNA, mRNA and their respective gene ontology results were therefore integrated to look for complementary trends in expression. This was assessed via miRNA target enrichment analysis and GSEA analyses.

7.5.2.5.1 miRNA target enrichment vs mRNA expression suggests a reciprocal relationship between miRNA and mRNA expression

The targets most (60%) miRNAs significantly downregulated at S_0 versus baseline were positively enriched in the mRNA results, i.e., their target mRNAs were generally upregulated at S_0 (Figure 101A). This is in line with the predicted reciprocal expression of miRNAs and their targets and suggests global downregulation of miRNAs at symptom onset releases mRNAs from repression, resulting in mRNA upregulation. In accordance with this, the collective mRNA targets of all downregulated miRNAs are positively enriched in the mRNA results (Supplementary figure 11).

The relationship between miRNAs and their target mRNAs was even more striking when comparing S_0 with S_7. The targets of most significantly upregulated miRNAs at S_7 compared with S_0 were negatively enriched in the

mRNA data—i.e., the targets of upregulated miRNAs were downregulated at the mRNA level (Figure 101B). This suggests that the global resolution of miRNA towards baseline values represses mRNAs causing them to reduce back towards baseline, reinstating immune homeostasis. In contrast to the strong connection between the miRNA and mRNA data, there was no relationship between sRNA-df and mRNA expression (appendix 9.6.3).

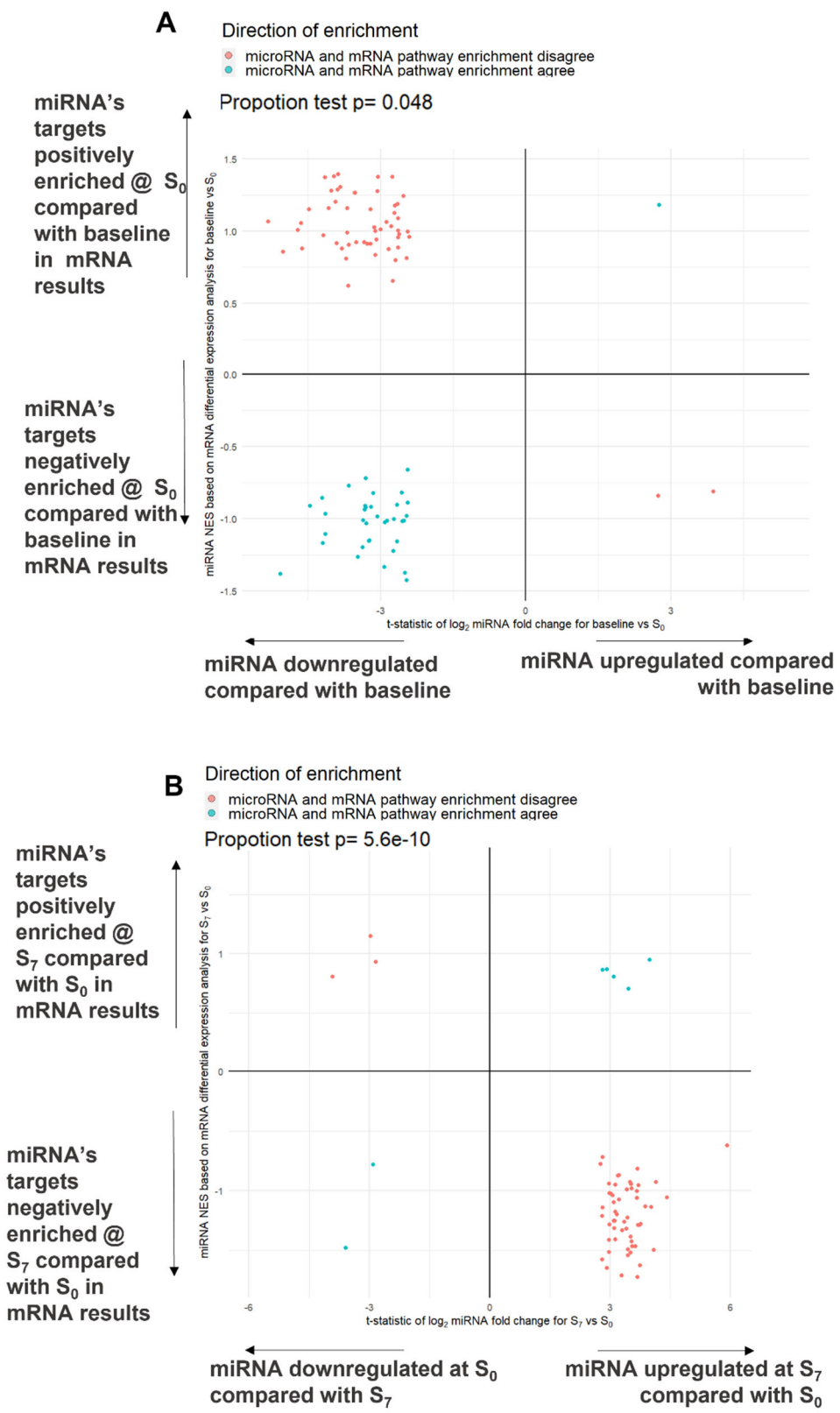


Figure 101 miRNA set enrichment analysis: comparing changes in miRNA expression with their targets in the mRNA data. The average t-statistic of the log_2 foldchange of each miRNA is plotted along the x-axis. t-statistic rather than log_2 foldchange is used because it captures the magnitude, significance and direction of foldchange. The enrichment

Figure 101 caption (continued)

score for each miRNA is plotted along the y-axis. Positively enriched miRNAs have mRNA targets, which are generally upregulated, and negatively enriched miRNAs have mRNA targets, which are generally downregulated. The colour indicates whether or not the sign of miRNA enrichment and miRNA foldchange agree (blue) or disagree (red). This can also be gauged from the quadrants of the graph a miRNA lies in.
A: Baseline vs symptom onset. Disagreement was significant based on a proportion test.
B: Symptom onset vs 7 days into infection. Disagreement in foldchange and enrichment sign was significant based on a proportion test.

7.5.2.5.1 GSEA analyses further support a reciprocal relationship between miRNA and mRNA expression

Potential antagonism of miRNA upon mRNAs was further explored at the pathway level by comparing the results of miRNA and mRNA GSEA style pathway analyses. This analysis again found a reciprocal relationship—pathways positively enriched at the miRNA level were generally negatively enriched at the mRNA level and vice versa. At baseline versus S_0, 59% of GO:BP pathways had disagreeing directions of enrichment in the miRNA and mRNA data. This disagreement in enrichment direction was strongly significant ($p<0.000001$) (Figure 102A). Furthermore, there was a significant relationship between the strength of enrichment and the probability of disagreement—the more positively or negatively enriched a pathway was in the mRNA data, the more likely it was to be enriched in the opposite direction in the miRNA data (Figure 102B, Appendix Figure 50, Appendix Figure 51). The same findings were seen in the stage I data.

Findings were even more robust when comparing S_0 with S_7 time points (Figure 103, Appendix Figure 52, Appendix Figure 53) and existed when comparing vaccine arms at S_0 and S_7 (Appendix Figure 54).

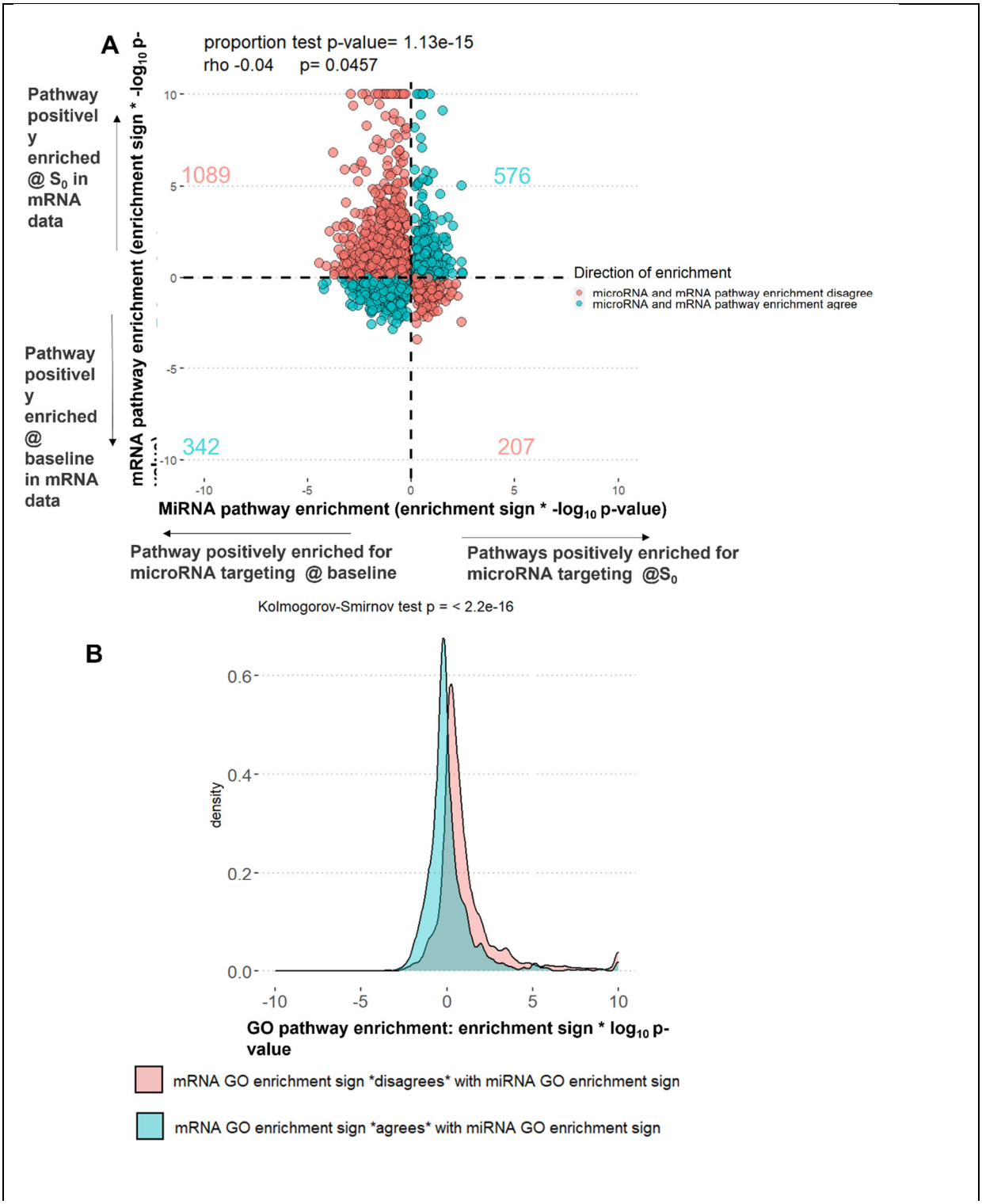


Figure 102 Baseline vs S_0 GO pathway enrichment comparisons.
Each dot represents the results of a pathway. The x-axis shows the -log₁₀ of the p-value of the pathway enrichment in the miRNA data. The y-axis shows the -log₁₀ of the p-value of the pathway enrichment in the mRNA data multiplied by the sign of enrichment (+1 if positive, -1 if negative). P-values are used because MIEAA does not give enrichment scores, only significance values. The colour indicates whether the sign of miRNA

Figure 102 caption (continued)
enrichment and mRNA enrichment results for a pathway agree (blue) or disagree (red). This can also be gauged from the quadrants of the graph the pathway lies in. The numbers in each quadrant show how many pathways fall in that quadrant. The results for all pathways tested in the miRNA and mRNA data are shown. B: Density plot shows the distribution of p-values (-log_{10} transformed) for pathways with enrichment signs that agree (blue) and disagree (red) in the miRNA and mRNA data.

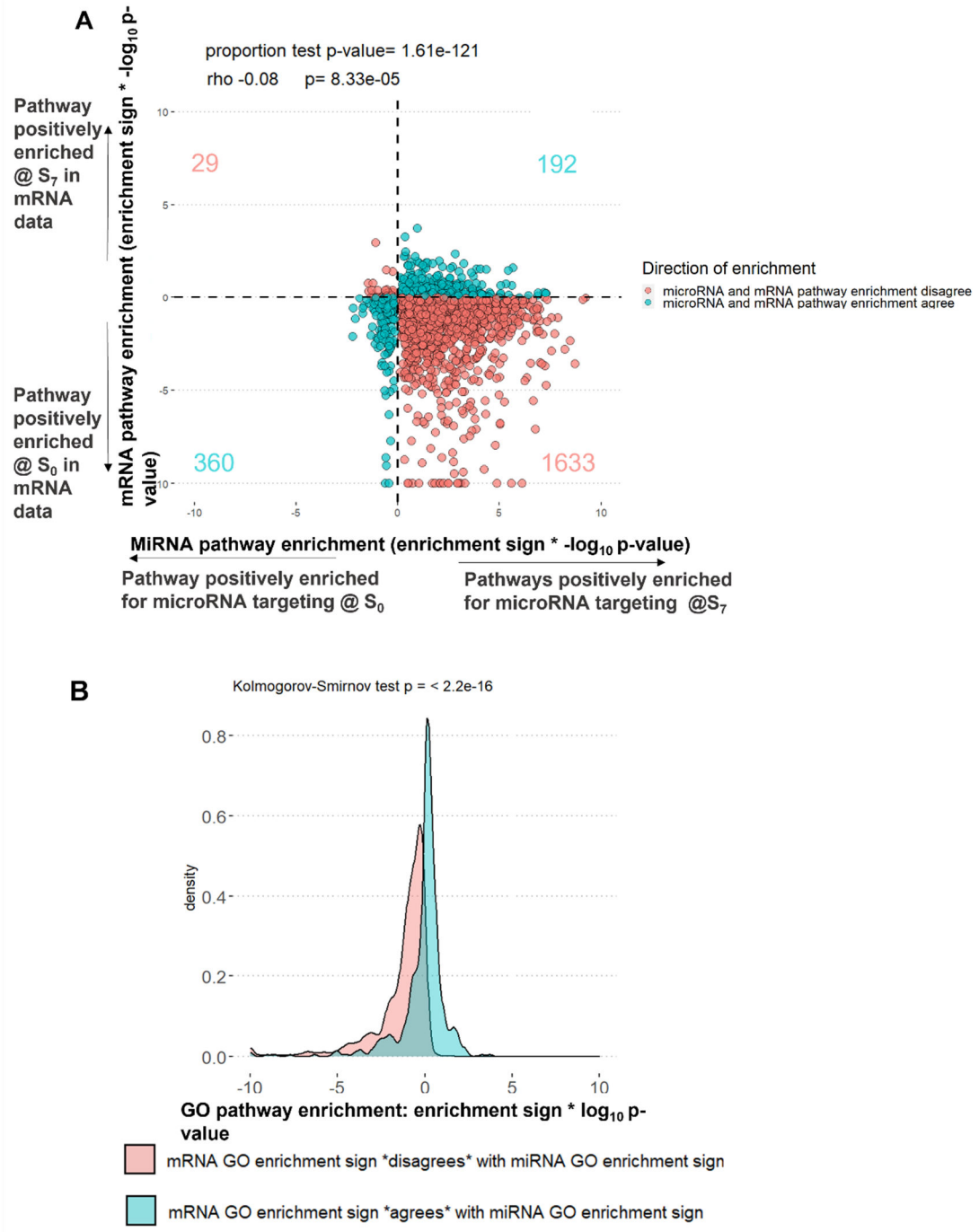


Figure 103 S_0 vs S_7 GO pathway enrichment comparisons
Each dot represents the results of a pathway. The x-axis shows the -log₁₀ of the p-value of the pathway enrichment in the miRNA data. The y-axis shows the -log₁₀ of the p-value of the pathway enrichment in the mRNA data multiplied by the sign of enrichment (+1 if positive, -1 if negative). P-values are used because MIEAA does not give enrichment scores, only significance values. The colour indicates whether the sign of miRNA enrichment and mRNA enrichment results for a pathway agree (blue) or disagree (red). This can also be gauged from the quadrant of the graph the pathway lies in. The numbers in each quadrant show how many pathways fall in that quadrant. The results

Figure 103 captions (continued)
for all pathways tested in the miRNA and mRNA data are shown.
B: This density plot shows the distribution of p-values (-log_{10} transformed) for pathways with enrichment signs that agree (blue) and disagree (red) in the miRNA and mRNA data. (Kolmogorov-Smirnov p-value for the difference in distributions is shown).

7.5.2.6 Vaccination attenuates sRNA perturbation in COVID-19.

The stage II study confirmed that ChAdOx1 nCoV-19 attenuates whole-blood sRNA perturbation in COVID-19. No sRNAs were DE between vaccine groups during COVID-19, but global-level differences were again present. There were fewer sRNAs DE at S_0 and S_7 when analyses included ChAdOx1 nCoV-19 vaccinees (Figure 95 C, Appendix Figure 48, Appendix Table 12), plus foldchanges were greater in menACWY subgroup analyses than in ChAdOx1 nCoV-19 subgroup analyses (Figure 104), and, at the pathway level, results were more profound in control vaccine arm sub-analyses. snoRNA-dfs are the only sRNA class to buck this trend.

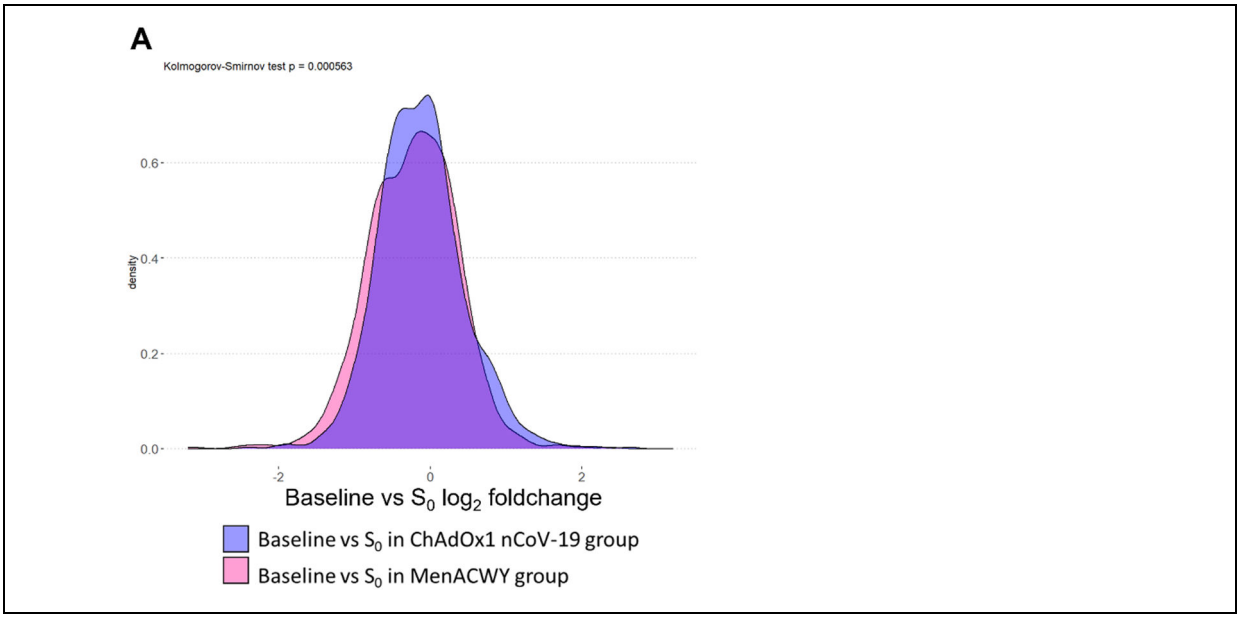


Figure 104 Global differences between ChAdOx1 nCoV-19 and control vaccine groups at COVID-19 symptom onset in sRNA average log_2 fold change.

7.5.2.7 Differences between vaccine groups 7 days into COVID-19 in miRNA targeting of T-cell pathways

Though there were no significantly DE sRNAs between vaccine groups, differences in pathway targeting were explored, given that global differences existed. The main timepoint for differences in pathway targeting was 7 days into infection, with 133 pathways differentially enriched for targeting between the vaccine groups. These pathways included; gamma-delta T-cell activation, innate immune response, negative regulation of I-kappaB kinase-kappaB signalling, negative regulation of JNK cascade, negative regulation of NF-kappaB transcription factor activity, negative regulation of interleukin-1-mediated signalling pathway, positive regulation of leukocyte migration, regulatory T-cell differentiation—all were positively enriched in the control vaccine group. The pathways relating to gamma-delta T-cell activation and regulatory T-cell differentiation were particularly interesting because, compared with baseline,

they were negatively enriched at day 7 following symptom onset in ChAdOx1 nCoV-19 vaccinees but positively enriched in the control vaccinees.

7.5.2.8 Multi-omics results identify genes involved in the cellular response to COVID-19 and converge on TNF alpha—a potential drug target

Links between serum cytokines DE at S_0 compared with baseline in the MenACWY group (IFN-γ, IL6, IL10, IL18, TNF-alpha and IP10) and DE RNAs and sRNAs are presented in Figure 106. These proteins, DE mRNAs, and targets of DE miRNAs were highlighted on the "Coronavirus Disease COVID-19" KEGG pathway. Many of the genes in this pathway were DE or targeted by miRNAs downregulated at S_0, often coinciding with upregulation at the mRNA level (Figure 105).

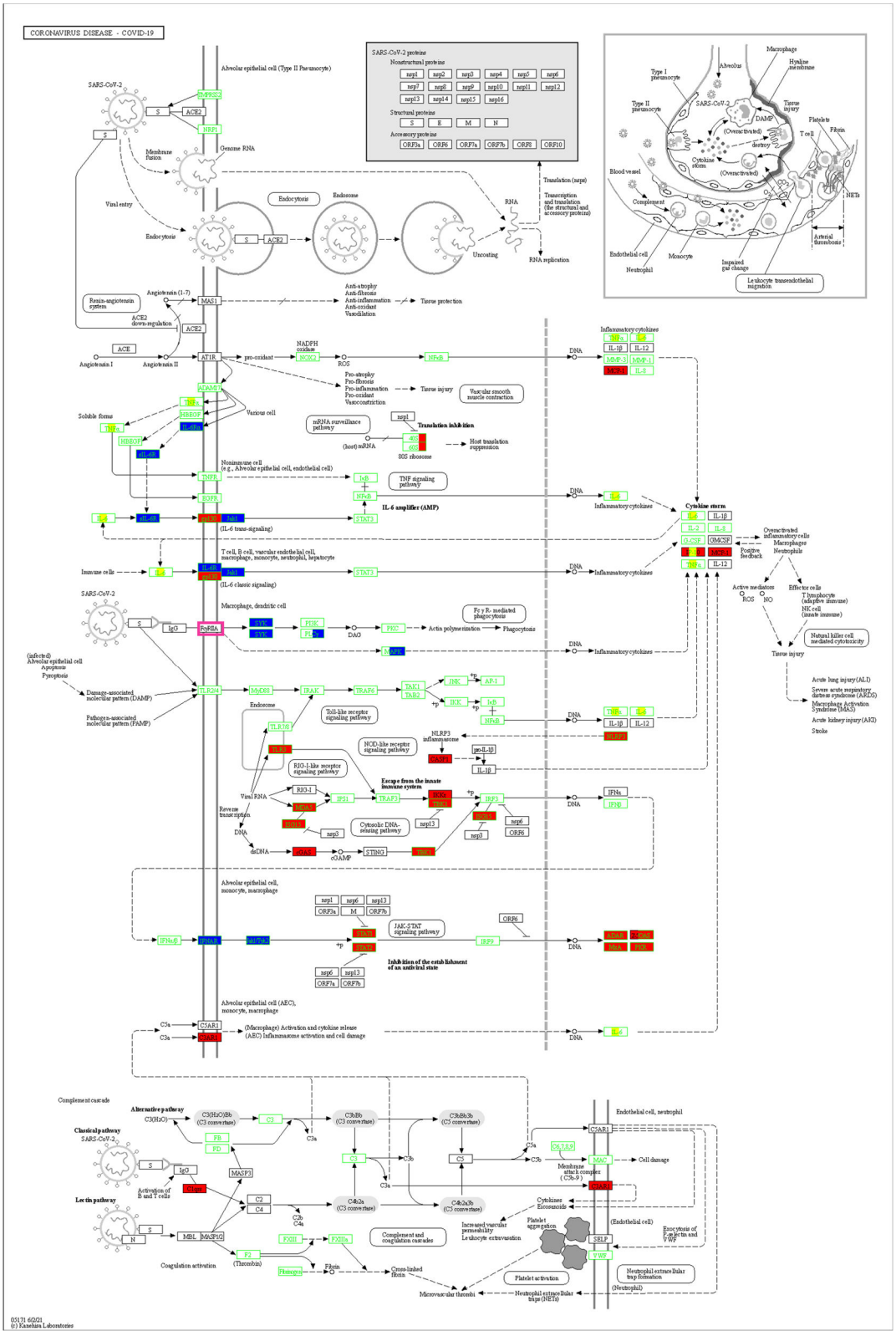


Figure 105 COVID-19 DISEASE KEGG pathway with DE features highlighted. Green outline and text = targets of DE miRNA, red fill = upregulated at the mRNA level, blue fill = downregulated at the mRNA level, internal yellow stripe = upregulated in serum at the protein level. Pink outline—an sRNA that lies reverse complement to FCGR2 (an antibody FC receptor) was downregulated.

Only three of the DE serum cytokines (TNF-alpha, IL18 and CXCL10) were detected at the mRNA level in whole blood—and of these, only IL18 and CXLC10 were DE (Figure 106). The disconnect between TNF-alpha at the mRNA and protein levels may suggest it is produced from another location in COVID-19 or released from storage in vesicles; however, post-transcriptional gene regulation by miRNAs is another possibility, particularly given TNF-alpha is targeted by miR-143-3p, which was downregulated at S_0 (Figure 106). Given that TNF appears in the COVID-19 Disease KEGG pathway and is being investigated as a potential drug target in COVID-19, it was reviewed in the pathway enrichment analyses. At the pathway level, "response to TNF-alpha (GO0034612)" is positively enriched in the mRNA data (p=0.022) and negatively enriched in the miRNA data (p=0.01). This is inline with TNF-alpha being a promising drug target in clinical trials[499].

IFN-γ was another overlap between the datasets. Though it was not detected in the mRNA data, pathways relating to interferon-gamma were positively enriched in the mRNA gene set expression analyses and negatively enriched in the miRNA pathway analyses. For example, interferon-gamma-mediated signalling pathway GO0060333 was enriched in both.

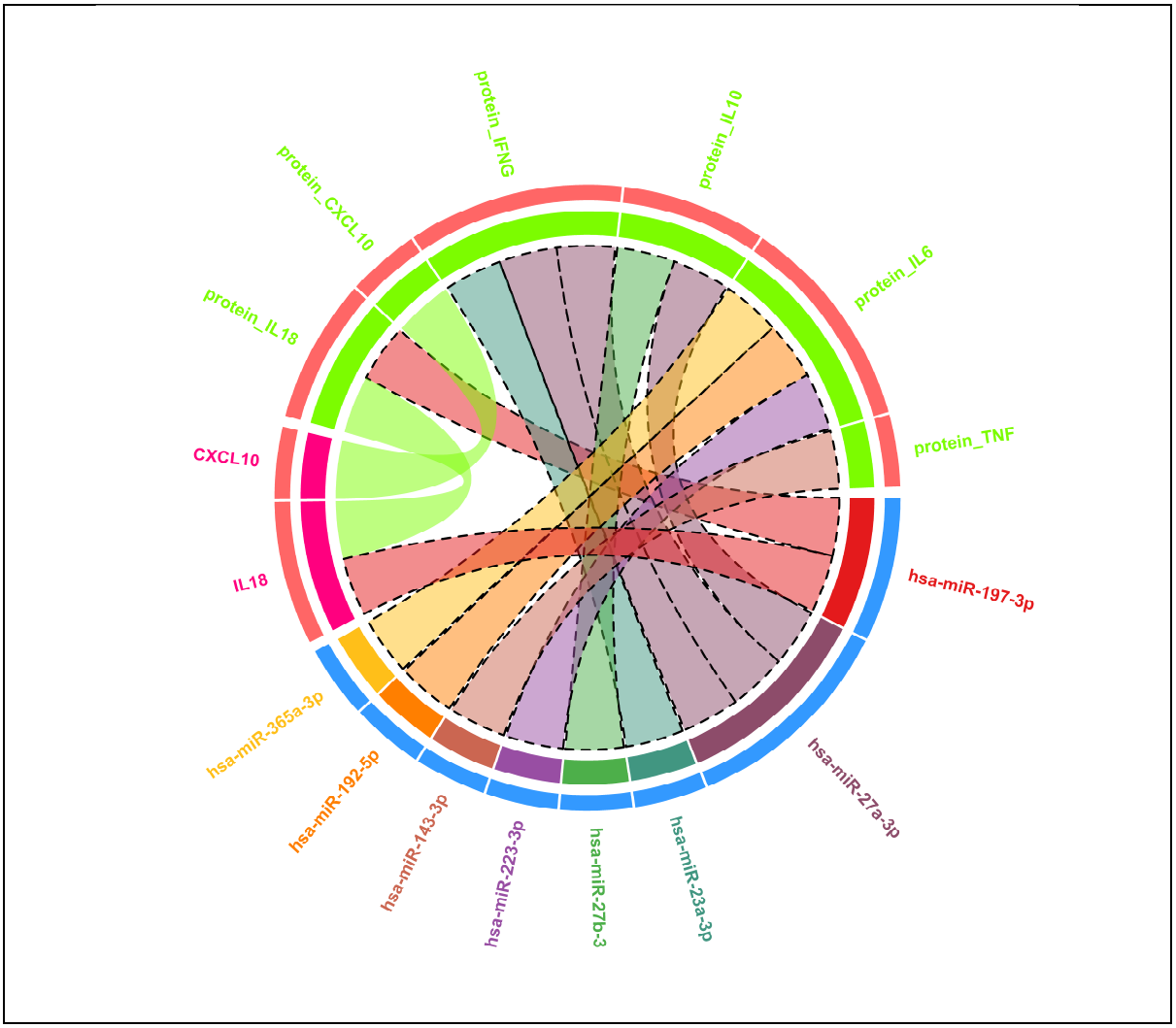


Figure 106 Chord plot showing the cross-over between DE proteins and DE mRNAs and miRNAs that encode and target them at S_0 vs baseline in the MenACWY group.
Track 1 (Outermost track) = feature name
Track 2 = whether that feature was up (red) or downregulated (blue) 24 hours after vaccination.
Track 3 = colour bar for that feature/feature type. Proteins are in bright green; mRNAs are in bright pink. miRNAs are multicoloured, so it is easier to see which protein is targeted by which miRNA. The length of the feature bar corresponds to the number of links the feature has with other features.
Linking lines ("chords") represent a connection between features. Links are drawn between mRNA and their protein counterparts. miRNAs are connected to their targets (downloaded from miRNet). Links with dotted edges connect features with opposite directions of foldchange. Links without dotted edges connect features with the same direction of foldchange. Note that the proteins are upregulated and only targeted by downregulated miRNAs.

7.6 Discussion

This study used whole blood samples from the randomised controlled ChAdOx1 nCoV-19 vaccine trials to understand how COVID-19 influences sRNA expression compared with health and COVID-like infection and over time in COVID-19. Each headline finding will be discussed in the broader context of the literature on COVID-19 and viral infections.

7.6.1 COVID-19 influences whole blood sRNA expression compared with health

The data show COVID-19 influences the expression of sRNAs within whole blood at the global class and individual sRNA feature levels. This was hinted at in the stage I study and confirmed in the stage II study. The top 5 DE miRNAs were hsa-miR-1260b, hsa-miR-5690, hsa-miR-7977, hsa-miR-30c-1-3p, and hsa-miR-5100. Hsa-miR-30c is associated with several infections, including Zika virus, Kaposi's sarcoma virus, LPS response and TB, and, interestingly, is predicted to target sequences within the SARS-CoV-2 genome (ORF1a) that encodes 16 non-structural-proteins[500–507]. MiR-7977 is a biomarker of complicated plasmodium vivax malaria infection and is upregulated in neurones infected with lyssavirus[508,509]. There is little literature on associations between miR-1260b, hsa-miR-5690, and hsa-miR-5100 and infection. DE miRNAs in acute COVID-19 also included those commonly identified in infection and leukocyte studies, e.g., hsa-miR-150-5p, hsa-miR-199b-5p, and hsa-miR-21-3p[35]. Other studies in this thesis also identified some of the DE miRNAs, e.g. hsa-miR-223-3p, is upregulated in infant plasma 24 hours after vaccination, though interestingly, it was downregulated in acute COVID-19 (see final discussion Chapter).

There was some validation of sRNAs in the stage I data. However, a small sample size and some limitations in sequencing quality of stage I mean results were compared with publicly available datasets to provide additional evidence of miRNA differential expression.

7.6.2 Studying the evolution of COVID-19 over the first week of illness

The results of stage II showed a global downregulation of miRNAs, which starts to resolve 7 days into infection. This may be an attempt to alleviate miRNA repression of leukocyte functions, causing rapid, widespread derepression of leukocyte effector functions. The ability of miRNAs to enable stand-by translation keeps cells in a poised state so cells can respond rapidly to stimuli by miRNA downregulation. This finding was supported by the reciprocal increase in the mRNAs that they target and potentially explains the overt inflammatory response responsible for COVID-19 severity. Notably, enrichment analysis showed more suppression of immunological pathways than in COVID-like illness.

The mechanism of global downregulation of miRNAs is unclear. At the gene expression level, DICER, the endonuclease essential for miRNA maturation, is downregulated at S_0 in the mRNA data. This has been linked to global downregulation of miRNAs in neutrophils in rheumatoid arthritis, so this may also be true in COVID-19, possibly in a wider array of leukocytes[510]. In the COVI multi-omics study, three AGO proteins, AGO1, AGO3, and AGO4, were down-regulated at the mRNA level. These proteins bind miRNAs in the RISC complex, so their downregulation could influence miRNA levels. Evidence supporting this

comes from an *in-vitro* study, which showed T-cell activation leads to global downregulation of miRNAs via AGO downregulation[511]. Another possibility is increased decay—RBX1 is upregulated at the mRNA level (log_2 foldchange of 0.66, FDR = $5.4x10^{-6}$) and plays a role in target-mediated miRNA decay (see section 1.2.2.4).

As time passes, miRNA levels return towards normal, though some changes persist. This return to normality potentially underlies reinstatement of immunological homeostasis as it corresponds to resolving perturbation of mRNAs. Enrichment analysis showed miRNAs that target pathways involved in immune responses recover most quickly, and this was associated with a corresponding reduction in those pathways at the mRNA level. Not all pathways identified in the miRNA enrichment analysis pertained to the immune response. Though miRNA expression patterns vary considerably amongst cell types, few are specific; thus, non-immune related enriched pathways may reflect the pleiotropic nature of miRNAs, the actions of which are context-specific, depending on which cells those miRNAs are being expressed within. It is possible that factors prompting changes in miRNAs in leukocytes prompt similar changes in the same miRNAs elsewhere, e.g. by activating or repressing a pattern of transcription factors. If this were true, theoretically, the enrichment results may provide a window into what may be happening within other cell types—particularly the nervous and cardiovascular systems, which are affected by SARS-CoV-2 infection, particularly in severe illness. This is highly speculative but does reflect similar findings in the VAST typhoid study.

One miRNA exception to the general miRNA pattern of differential expression was hsa-miR-6501-5p—interestingly, it increased 7 days into infection in both the stage I and II studies. Though hsa-miR-6501-5p was not DE between vaccine groups after multiple testing corrections (p=0.01, FDR 0.5), a review of its expression patterns show its rise was mainly restricted to the control vaccine group in stage I and II. There is little literature on hsa-miR-6501-5p. A PubMed search only identified two papers; interestingly, both relate to the upregulation of hsa-miR-6501-5p in the whole blood of COVID-19 patients[491,493]. The only miRNA expression atlas to give a breakdown of leukocyte miRNA content and contain an entry for hsa-miR-6501-5p is the FANTOM5 cell expression atlas, which shows it is exclusively expressed in T-lymphocytes and monocytes. This possibly suggests some aspect of T-cell responses, potentially specific for SARS-CoV-2 (given the absence of reference to any other condition in the literature), which are unique to unvaccinated individuals and present later in infection. It could be a facet of adaptive immunity that is not induced by vaccination, given it increases 7 days into infection yet shows no changes in ChAdOx1 nCoV-19 vaccinees at any time point.

The action of sRNA-dfs DE in COVID-19 (derived from tRNAs, protein-coding genes, lncRNAs, and sequences that lie antisense to protein-coding and lnc RNAs) is unclear. tRNA-dfs are induced by stress and are associated with global downregulation of translation (see introduction 1.2.2.7). A notable finding during acute COVID-19 was the global upregulation of snoRNA-dfs, which peaked at symptom onset in the ChAdOx1 nCoV-19 group and 7 days into illness in the control vaccine group. The absence of reciprocal changes in other sRNAs

suggests this is a true effect, not an artefact of stochasticity in RNA-seq. Little is known about snoRNA-dfs. DICER and FUS (HNRP) can cleave snoRNAs into fragments, some of which have been shown to act as miRNAs[262,263]. Interestingly, FUS was upregulated at the mRNA level at S_0 and S_7. The earlier peak in the ChAdOx1 nCoV-19 group possibly links snoRNA-dfs to induction of the adaptive immune response, which would be predicted to appear more quickly in the ChAdOx1 nCoV-19 group due to clonal expansion of spike-specific T and B-cells induced by vaccination. The changes seen in snoRNA-derived sRNA are supported by the findings of Wilson et al., who described this phenomenon later in COVID-19 disease[512]. The role of snoRNA fragments in the immune response is just coming to light with *in-vitro* and clinical biomarker studies identifying their differential expression in flu, TB and multiple sclerosis compared with healthy controls[513–515]. The flu study showed that snoRNA-dfs do not correlate with the expression of their full-length RNA counterparts, which was also true in the CoVI-multi-omics study. Data on the mechanism or biological function of snoRNA-dfs is lacking. Further work is required to characterise the driving force behind changes in snoRNA-df expression in COVID-19 and elucidate their biological function.

7.6.3 Comparing COVID-19 with COVID-like illness

The symptomatic controls in the stage I study could offer insight into the specific aspects of COVID-19 that make it different from other illnesses, though caution is warranted when drawing conclusions given the small sample size. Using baseline expression as a reference for interpreting foldchanges, tRNA-Gly, and an sRNA-df derived from hemicentin2 (HMCN2—a secreted extracellular matrix

protein) are downregulated in COVID-19 compared with COVID-like illness at symptom onset. hsa-miR-6501-5p and sRNAs derived from the sequences mapping to antisense to ENTPD4, ZC3HAV1 and KLF13 were upregulated in COVID-19 compared with COVID-like illness 7 days into the illness. In all these cases, differences between COVID-19 and COVID-like illness appear to be driven by changes in the COVID-19 arm rather than the COVID-like illness arm.

The miRNA pathway enrichment analysis showed immunological pathways were positively enriched for post-transcriptional regulation at symptom onset in COVID-like illnesses compared with both health and COVID-19. A unique aspect of pathology in COVID-19 may be an excessive immune response arising from a lack of miRNA-imposed gene and pathway regulation. There are alternative explanations, however, so this hypothesis requires further investigation in a larger cohort with better symptomatic controls.

The small number of differences between COVID-19 and COVID-like illness may be due to the small sample size. Alternatively, biological noise stemming from heterogeneity in underlying diagnoses may also obscure differences.
Given fever, cough, or loss of smell or taste was sufficient to trigger a symptomatic visit, COVID-like illnesses can range from non-infectious (e.g., hay fever) to respiratory infections (e.g., RSV) and non-respiratory illnesses, e.g., urinary tract infections. COVID-like illness may therefore be somewhat of a misnomer. Ideally, a group with a similar syndrome to COVID-19—e.g., respiratory tract infections—would be an ideal comparison to isolate the immunological response to SARS-CoV-2. Samples from the multi-omics COMBAT

study would be useful to do sRNA-seq on given the well-phenotyped comparison groups, which include flu and sepsis cohorts and healthy participants[516].

7.6.4 The effect of ChAdOx1 nCoV-19 vaccination on COVID-19

ChAdOx1 nCoV-19 vaccination reduces perturbation at the sRNA level, and this finding was repeated across the protein and mRNA datasets, which showed reduced perturbation of COVID-19 severity biomarkers. This was despite a lack of difference in viral load between ChAdOx1 nCoV-19 and control vaccinees. The reason is unclear; possibly mild infection results in less stimulation of the immune response and, therefore, fewer changes at the molecular level. The findings biologically underpin real-world clinical data showing COVID-19 vaccines reduced disease severity.

7.6.5 Study Strengths

Though there are eight published studies on miRNA expression in COVID-19 in whole blood, the COVI multi-omics study had several advantages: it was adequately powered to find statistical differences after multiple testing, compared COVID-19 to COVID-like illnesses, assessed miRNAs changes over the time course of COVID-19, had paired mRNA data so that integration could be done, and explored all of the above in the wider sRNAome. Additional strengths included the very early time point that participants were recruited. Participants were strictly instructed to attend for testing within 24 hours of symptoms, and blood samples were taken before the results of COVID-19 swabs were known. Most of the studies surveying the miRNAome recruited hospital patients who

would have been ill for a while as the average time to hospital presentation is between 4 to 10 days[517,518]. Investigating the influence of prior vaccination with ChAdOx1 nCoV-19 miRNA expression using the unbiased context of a blinded RCT is a particular strength of the CoVI-multi-omics study. Non-RCT cohort studies comparing SARS-CoV-2 vaccinees with non-vaccinees are at high risk of confounding given age, comorbidities, and occupation influence vaccine take-up along with disparities in access and vaccine confidence across regions and population demographics.

7.6.6 Study limitations and considerations

The generalisability of these results to serious COVID-19 infection is limited, given the infections within this study were mild. Serious illness was rare in the COV trials when samples were being selected. Partly this is down to the population of people participating in the trials; enrolment was only later widened to include a pragmatic, relevant population covering older age and co-morbidities, and those with serious unstable illness and severe frailty were not enrolled. Whether some of these findings are SARS-CoV-2 variant specific is unclear as all cases in this study involved the alpha variant.

A further limitation of the CoVI-multi-omics study was the lack of paired baseline samples available for stage II. Efforts were made to match them in age and co-morbidities to those with COVID-19 samples included in stage II, but other differences could be present. It also means there is less power for comparisons with health. Compounding this is the lack of full blood count data available for baseline results. Red cells are enriched for miRNAs, and red cell counts vary

between participants; therefore, including them in differential expression models is helpful. This could not be done for baseline comparisons, further limiting power. A lack of pairing and red cell count control might explain why the miRNA pathway analysis found many more significant pathways when comparing S_0 to S_7 than S_0 to baseline.

A further limitation of the study was the small size of the stage I results. Usually, the second cohort—stage II in the study—acts as a validation cohort; however, the sample size meant stage II had the greatest power, so validating significant results meant taking FDR significant findings in stage II and using a cut-off of an unadjusted p-value below 0.05 in stage I. A further issue was the sequencing quality in stage I results, with larger than usual variation in library sizes—this limits the usefulness of stage I results as a good validation cohort. Reassuringly, however, the take-home messages were consistent between the two stages. Furthermore, using publicly available studies of variable sizes corroborated differential expression of 22 miRNAs.

A final limitation to consider is the interpretation of the results. What underlies changes in sRNA expression is not clear. Some or all of the changes in the sRNAome may represent changes in the underlying circulating populations rather than changes within those populations. The COV studies revealed that there is a relative increase in circulating monocytes, memory T-cells and plasmacytoid dendritic cells and a decrease in naïve B-cell and neutrophils at COVID-19 symptom onset (based on CIBERSORT deconvolution analyses). Further work is warranted to pursue the mechanism and biological effects of sRNA changes, for

example, by delving into particular leukocyte subsets and *in-vitro* work. Unfortunately, single-cell miRNA sequencing is not at the high throughput level required to investigate these phenomena at the individual cell level.

7.7 Conclusion

In conclusion, the COVI multi-omics study demonstrates that COVID-19 influences the expression of miRNAs and sRNA-dfs derived from other RNA species. MiRNAs are downregulated at the onset of COVID-19, which is inferred to lead to widespread de-repression of mRNAs and pathways, including, but not limited to, those that target pro-inflammatory immune pathways. A return towards baseline levels occurs later in infection, possibly as a mechanism of re-establishing homeostasis. The study has offered insights into the real-time protection afforded by ChAdOx1 nCoV-19 at the molecular level. This study helps probe the pathophysiology of COVID-19 and the protective effects of vaccination and may also offer broader insights into the immune response to viral infections.

8 Discussion

This discussion summarises key thesis results, drawing together findings across three diverse studies to offer new insights into the roles and associations between sRNAs and vaccination and infection. Findings will be set within the context of the broader literature, and avenues for further research will be identified.

8.1 Key study findings

8.1.1 EUCLIDS

The first study in this thesis assessed plasma expression of sRNAs and proteins before, 4 and 24 hours after administration of vaccines in 4 and 12-month-old participants of the EUCLIDS study—a randomised, descriptive, open-label study exploring the influence of concomitant 4CMenB on the whole blood transcriptome with a focus on vaccine reactogenicity.

The study showed for the first time that vaccination alters plasma expression of miRNAs and sRNA-dfs derived from tRNAs, snoRNAs and protein-coding genes 24 hours after vaccination. The study found miRNAs likely arising from leukocytes (probably myeloid cells) and liver were generally upregulated and downregulated, respectively. This was particularly true for miR-122-5p—a finding recapitulated in a mouse vaccination model. Deconvolution of plasma sRNA profiles suggested a relative decrease in sRNAs carried by circulating

ribonucleic acid protein complexes and an increase in sRNAs contained in low-density vesicles.

Luminex and mass spectrometry revealed that vaccination induces the hepatic acute phase response. In addition, a network of DE plasma miRNAs, proteins, and whole blood RNAs and their interactors suggested a link between serum IL-6 upregulation and hepatic miRNA expression. This *in-silico* analysis was supported by *in-vitro* work, which found IL6 stimulation of a hepatocyte cell line reduced hsa-miR-122-5p release, recapitulating findings of the infant study. These results tentatively infer a link between hsa-miR-122-5p and the hepatic acute phase response. Additional support for the association between hsa-miR-122-5p and the acute phase response comes from the COVID study. At COVID-19 onset, there was a non-significant drop in miR-122-5p in whole blood, followed by a significant increase back to baseline level (at $p<0.05$) 7 days into COVID-19 illness. Testing whether this reflects the plasma fraction of whole blood or intracellular levels within granulocytes or platelets could inform hypotheses on why plasma miR-122-5p drops after vaccination.

The EUCLIDS study suggested a link between vaccine-induced miRNA expression and host factors (infant age) and vaccine reactogenicity, with changes in plasma sRNA expression more profound in young infants and after concomitant administration of the reactogenic 4CMenB vaccine. In addition, at 4 months of age, plasma miRNAs correlated with maximal fever, a finding mirrored in the mouse model and expression of hsa-miR-122-5p, hsa-miR-223-3p and hsa-miR-375-3p correlated with post-vaccination neutrophil abundance.

8.1.2 VAST

The second study in this thesis assessed the influence of vaccination on sRNA expression in PBMCs in a typhoid vaccine study embedded within a human challenge model of *S*. Typhi. Vaccination subtly affected miRNA expression over time in PBMCs at 7, 10 and 28 days after vaccination. The paucity of longitudinal change in miRNA expression may partly arise due to heterogeneity in vaccine responses as divergent expression of miRNAs occurred 10 days after vaccination in those that remained well compared with those that developed typhoid. The association between challenge outcome and miRNA expression is the first evidence of potential miRNA-based correlates of protection.

An increase in sRNA-dfs arising from immunoglobulin genes occurred 7 days after Vi-TVC vaccination, especially in those not developing typhoid. These sRNA-dfs positively correlated with Vi-specific B-cell frequency and higher IgG antibody titres to Vi and tetanus (more so to Vi). The absence of these findings in the Vi-PS group links these immunoglobulin sRNA-dfs to T-cell dependant responses. Baseline differences of these sRNAs were associated with susceptibility to typhoid in the Vi-PS group (and persisted 7 days after vaccination), whereas this was not the case in the Vi-TCV group, potentially indicating people who respond best to a Vi-TCV vaccine differ from those who respond best to a Vi-PS vaccine.

S. Typhi induces systemic signals very early—with sRNA changes occurring within 12 hours. The study provides the first evidence in man that pathogen exposure in the absence of systemic disease can profoundly influence miRNA

responses in PBMCs. Vi-vaccination induced additional changes in sRNA expression after challenge, indicating vaccine-related immunity shapes the early response to pathogen exposure.

Interestingly, the peak timepoint for differential expression of miRNAs early after challenge comes later than that of mRNAs in whole blood (which peaks at 12 hours). Alongside this, enrichment analysis suggested miRNA expression 24 hours after challenge negatively regulates post-transcriptional silencing by miRNAs. If true, this could be part of a feedback loop designed to maintain homeostasis of post-transcriptional regulation in the face of immunological challenge—a concept partly described by Mehta et al.[25].

Though the expression of individual sRNAs did not correlate with outcome, global foldchanges and effect sizes did (more so in the control group). This indicates hypo-responsiveness to *S*. Typhi on initial exposure is associated with susceptibility, but immunity conferred by Vi-vaccination can somewhat overcome this effect. Acute typhoid had a profound effect on sRNA expression. sRNA profiles could differentiate samples at typhoid diagnosis from those taken before and 7 days after challenge (in those that remained well). Though this signature is not clinically valuable (because a useful signature of typhoid discriminates differential diagnoses, not health from disease), consistency in sRNA response in acute typhoid opens up the possibility of a diagnostic signature. Interestingly the most DE miRNA was hsa-miR-1303, which has only been linked to one other infection—TB, another facultative intracellular pathogen, indicating miRNAs

might discriminate between pathogen types, e.g. extracellular bacteria, intracellular bacteria, viral and fungal infections.

8.1.3 COVID-19 vaccination and infection study

The final study in this thesis assessed the influence of COVID-19 on whole-blood sRNA expression in participants of a randomised controlled trial designed to test the efficacy of ChAdOx1 nCoV-19 versus a control vaccine. This study showed COVID-19 influences whole-blood expression of miRNAs and sRNA-dfs arising from tRNAs, rRNAs, lncRNAs, snoRNAs, miRNA hairpins and protein-coding genes, and transcripts derived from loci lying antisense to known features.

The CoVI-multi-omics study showed widescale upregulation of pro-inflammatory mRNAs coincided with global downregulation of miRNAs at symptom onset. The inference is that loss of miRNA post-transcriptional regulation facilitates a rapid and robust inflammatory response, potentially contributing to COVID-19 pathogenicity. Over time, miRNA deregulation resolved, particularly amongst miRNAs targeting pathways involved in immune responses, and this coincided with a resolution of mRNA perturbation. This was particularly evident when evaluating TNF-alpha and IFN-λ at the protein and pathway level. Assessing miRNA expression at the onset of more severe COVID-19 infections would be interesting to see whether severe disease is linked to further widespread downregulation of miRNAs[109,516].

The study also circumstantially tied snoRNA-df expression to the onset of adaptive immunity as snoRNA-dfs continued to increase 7 days into COVID-19

with an earlier peak in ChAdOx1 nCoV-19 vaccinees. SnoRNA-dfs comprised a small fraction of whole-blood sRNAs (~1% of reads) but a relatively high proportion in PBMCs, according to the VAST study (~ 16% of reads). A rise in snoRNA-dfs in COVID-19 thus fits with a rise in a PBMC cell population, e.g. lymphocytes. Caution is required when comparing studies because RNA was extracted using different kits and sequenced in different batches, which can influence relative sRNA abundance[158].

The CoVI-multi-omics study showed that ChAdOx1 nCoV-19 vaccination ameliorates sRNA perturbation during breakthrough infection—mirroring the vaccine's effect on disease severity.

8.1.4 Comparing the distribution of RNA classes across studies

There are differences in the distribution of reads mapping to each RNA class across the three studies. Notably, mature miRNAs were the most abundant in PBMCs and second most abundant in whole blood but less expressed in plasma. Offset miRNAs were the most abundant class in whole blood but barely detected in plasma and PBMCs. tRNA-dfs were highly abundant in plasma, moderately abundant in PBMCs and barely detected in whole blood. snoRNA-dfs were highly expressed in PBMCs but barely detected above noise in plasma and whole blood. The abundances of yRNAs and rRNAs were fairly similar across all studies. See Supplementary Table 15 for how the relative abundance of each RNA type compares across the substrates sRNA-sequenced in this thesis. Because each study was sequenced at a different time, batch differences (e.g. from size selection and adapters) cannot be excluded as a source of these differences.

Nevertheless, it could suggest that miRNA isomiR abundance is very different between PBMCS and granulocytes and/or red cells and shows that plasma sRNAs are not a reflection of intracellular RNAs from leukocytes.

8.2 Answering the key research questions of this thesis with a view across the three studies

This section collates findings from the three thesis studies to answer the research questions posed at the start of this thesis.

8.2.1 Do vaccination and infection alter miRNA expression?

Yes: Vaccination and infection clearly influence miRNA and sRNA-df expression across the three studies. This thesis is the first to show vaccination and infection alter the expression of sRNA fragments derived from lncRNAs and protein-coding genes.

8.2.1.1 Comparing results across thesis studies

In total, 1,846 sRNAs were DE in at least one comparison in at least one study (at FDR<0.05). Two hundred and seventy-five sRNAs were DE in two studies (15%), of which 106 were mature miRNAs; the rest were sRNA-dfs from tRNAs, vault RNAs, protein-coding genes, lncRNAs, snoRNAs and loci that lie antisense to protein-coding genes. Five features were DE in all three studies—hsa-miR-223-5p, hsa-miR-223-3p and sRNA-dfs from tRNA-Val, tRNA-Arg and tRNA-Asp.

8.2.1.1.1 tRNA-dfs and infection

RSV and mTB infection and cellular stress alter tRNA expression by induction of Angiogenin, which cleaves full-length tRNAs[269,283,284]. These tRNA fragments can act as miRNAs and globally downregulate translation as a cell stress response [270]. Interestingly, these fragments were upregulated in infant plasma after vaccination and at typhoid diagnosis but were downregulated in COVID-19. Downregulation of these tRNA-dfs in COVID-19 may contribute to widespread loss of post-transcriptional repression in COVID-19. Given this was not seen in typhoid, it further suggests the pathogenesis of COVID-19 may be linked to a reduction in gene regulation.

8.2.1.1.2 miR-223-5p and miR-223-3p and infection

As discussed in Chapter 4, the top cellular source for hsa-miR-223-5p and hsa-miR-223-3p are CD15+ cells—a myeloid-specific marker, but they are also present in platelets and across all leukocytes, explaining why they were identified in VAST PBMCs. These miRNAs were downregulated in COVID-19 but upregulated in infant serum after vaccination and at typhoid diagnosis. Interestingly, hsa-miR-223-5p was upregulated 10 days after vaccination in participants who did not develop typhoid compared with those that did. These miRNAs have been identified in other infection studies, including mTB, dengue, H-pylori and sepsis, where they correlate with outcome[519–523,524,244,525]. Most studies look at the role of miR-223-3p in neutrophils, but it also plays a role in macrophage polarisation, promoting an anti-inflammatory M2 phenotype[186]. IL6 has been shown to promote miR-223-3p enrichment in macrophage exosomes,

potentially explaining why vaccination prompts upregulation of this miRNA in infant plasma.

8.2.1.1.3 Other sRNA-dfs in the immune response

There is no literature on the biological role of DE sRNA-dfs. Some sRNA-dfs DE across multiple studies arise from genes with well-characterised roles in the immune response. For example, S100A8 and S100A9 form the heterodimer calprotectin, which chelates transitional metals, starving bacteria of access to these elements. These sRNA-dfs were upregulated in plasma after infant vaccination and in PBMCs 1 day after *S.* Typhi challenge and at typhoid diagnosis.

8.2.2 Can miRNAs or sRNA-dfs predict or shape response to vaccination and/or infection?

Yes: The EUCLIDS, VAST, and COVID studies associate sRNA expression with vaccine responses in terms of reactogenicity, effectiveness, and amelioration of disease in breakthrough infection, respectively. More work is required to determine whether sRNAs modulate or simply correlate with vaccine responses.

8.2.2.1 How these findings relate to the literature

There is little literature on miRNA association with vaccine responses. One study found that 8 miRNAs were induced in an in-vitro fever model in Th-P1 cells[526]. None of these miRNAs was associated with fever in EUCLIDS. That said,

intracellular expression of heat-stressed THP-1 cells and plasma miRNA expression after vaccination are poorly comparable systems.

Two previous studies equated miRNA expression and antibody titres (see section 1.2.2.6.7), but none investigated miRNAs as correlates of protection. hsa-miR-107, hsa-miR-20b-5p and hsa-miR-15a-5p are more highly expressed in measles-stimulated B-cells from people with high versus low measles titres[527]. Paradoxically, these miRNAs were more highly expressed 10 days after vaccination in PBMCs from VAST participants who developed typhoid versus those that did not.

8.2.3 Do miRNAs play a functional role in the immune response to infection/vaccination?

Yes, potentially: Integration across multi-omics data sets and findings within this thesis suggest miRNA could play a functional role in the immune response to vaccination and infection. For example, integrating miRNA and mRNA data in the COVID-19 study revealed a global drop in miRNA expression in whole blood at COVID-19 onset coincided with an increase in their mRNA targets. This finding may partly explain the pathogenicity of COVID-19, particularly given this phenomenon was not seen at typhoid onset.

Another potential functional insight was discovering that hsa-miR-122-5p was downregulated in plasma after vaccination. It is not unreasonable that a highly abundant miRNA in hepatocytes and within plasma could participate in the acute phase response either through post-transcriptional regulation of hepatocyte

acute phase response genes or as an intercellular regulator of gene expression within leukocytes. The *in-vitro* model in Chapter 6 should be further exploited to see how changes in miR-122-5p release relate to intracellular levels, the post-transcriptional regulation of acute phase response proteins, and gene expression in leukocytes via co-culture models.

8.2.3.1 The Role of sRNA-dfs in immune responses

There is increasing interest in the action of sRNA-dfs. As the introduction Chapter covers, sRNA-dfs derived from snoRNAs, yRNAs, and tRNAs are not degradation products and play roles in gene regulation at the transcriptional, post-transcriptional, and translational levels. Their role in infections is only starting to be investigated. This is the first study investigating the relationship between sRNA-dfs and vaccine responses. Validated differential expression of lnc and protein-coding sRNA-dfs in this thesis suggests some RNA fragments may be biologically functional molecules (e.g. endo-siRNAs) rather than biological noise. The frequent lack of correlation between sRNA-dfs and their RNA counterparts, as seen in the COVID study, adds weight to this hypothesis. For example, KLF13 is upregulated one day after *S.* Typhi challenge, at typhoid diagnosis, and 7 days into COVID-19, but this gene is not DE in either mRNA dataset.

No published studies have addressed associations between sRNA-dfs and aspects of immune responses in such comprehensive datasets as in this thesis, especially not datasets derived from clinical samples with associated paired RNA-

sequencing data. These datasets are, therefore, an important contribution to exploring the possible roles of sRNA-dfs in the immune response.

Future work is needed to determine biological mechanisms (if any) underlying associations between DE sRNA-dfs in this thesis and immune responses. In the first instance, this requires further bioinformatic analysis to look at features and motifs at the loci from which sRNA-dfs derive. The secondary structures of these features need assessing to understand whether they represent novel miRNAs, piRNAs or something else. Publicly available datasets of RIP-sequencing could provide information on which RNA-binding proteins bind DE sRNA-dfs. Such proposals open a pandora's box of research questions that extend beyond the initial scope of this thesis.

8.3 Potential translation of study findings to the clinic

Diagnostic biomarkers

The *ex vivo* stability of miRNAs makes them excellent candidates for diagnostic signatures. Given miRNAs curate cell responses, their expression may summarise latent patterns in transcriptomic response—essentially acting as feature selection tools. Many studies have identified miRNA-based diagnostic signatures and biomarkers of disease severity and outcome, but their potential hinges on whether they are sufficiently discriminative[35]. The multi-cellular expression and pleiotropic roles of each miRNA may mean they are not specific enough to act as good biomarkers – potentially explaining why miRNA-based biomarkers in infectious diseases are not as developed as mRNA-based ones currently in clinical trials.

miRNA-based therapeutics

Future work should seek a mechanistic understanding of why miRNAs are associated with vaccine and infection responses across these studies. In the first instance, this means isolating the cells miRNAs derive from and manipulating those miRNAs *in-vitro* or *in vivo* to ascertain whether there is a mechanistic link between miRNAs and vaccine and infection responses. If there is, vaccines could be tweaked to induce a different systemic/cellular miRNA response, or adjuncts or treatments could be developed that target (or promote) particular miRNAs to encourage more effective host response. Trials of miR-122-5p antagonists as Hepatitis C treatments show this approach is clinically feasible[259,260,528].

miRNA-based biomarkers of vaccine reactogenicity and effectiveness

Even if there is no mechanistic link between miRNAs and vaccine/infection outcomes, these results may still offer use as biomarkers. For example, in Chapter 3, the expression of miR-122-5p was associated with concomitant administration of the reactogenic 4CMenB and fever. Similar associations were also seen in infants using a multi-omics signature in Chapter 5. The subjective nature of symptom reporting, the fluctuating nature of fever and interoperative differences in measuring temperatures create noise in data. An objective signature that could identify the reactogenic potential of a vaccine in pre-clinical models or a smaller number of people could help prioritise vaccine candidates for further development and enable the discard of vaccines likely to be unacceptably reactogenic.

miRNA biomarkers of vaccine effectiveness found in VAST (Chapter 6) are worth pursuing. Even if they do not turn out to be mechanistic correlates themselves, they ought to correlate with the mechanistic correlate—thus, the pursuit of what drives divergence in miRNA expression 10 days after typhoid vaccination could identify that mechanistic correlate. If further work shows miRNAs identified as potential correlates of vaccine effectiveness translate to other vaccines and species, then they might help predict whether a vaccine is likely to produce a protective response in pre-clinical development and early-phase clinical trials, rationalising vaccine design.

8.4 Thesis limitations

Study-specific limitations were discussed in the relevant results chapters. Limitations common to more than one study and are addressed here.

8.4.1.1 sRNA-seq

As noted in Chapter 1, sequence-dependent ligation and amplification mean some sequences are overestimated, whilst others are underestimated or not sequenced. Another issue is the inability to detect and exclude PCR duplicates created during library preparation. This is because whole transcripts are sequenced; thus, there is no unique "fingerprint" of random fragmentation to identify duplicates in sRNA-seq. This is a particular issue in low-input RNA sequencing studies, e.g. biofluids, in the EUCLIDS study. It is less of a problem in this thesis because the same biases affect samples equally within a project, but it does limit comparisons across thesis studies. Future sRNA-seq studies

could use unique molecular identifiers (UMIs) and adapters incorporating random nucleotides to enable identification and exclusion of PCR duplicates and capture more of the sRNAome[292,529].

Another issue relates to the alignment of multi-mapping reads and data processing, including noise estimation and normalisation. No standardised approach exists, and all methods have drawbacks. Reassuringly, experimentation with different pipelines, alignment methods, normalisation procedures and differential analysis methods did not affect the take-home messages of studies in this thesis. Before putting effort into follow-up studies of interesting sRNAs, a researcher would like to be confident that the signal behind DE sRNAs is genuine and not due to incorrect mapping of multi-mapping reads. This is why a flexible assignment pipeline was developed that allows researchers to quickly check alternative alignments of reads mapping to their sRNA of interest. It also allows researchers to assess whether the expression profile of their sRNA of interest indicates degradation product—particularly relevant for sRNA-dfs.

8.4.1.2 Bulk sequencing of PBMC populations.

The COVID and VAST studies involved bulk sequencing of PBMCs, making it difficult to infer whether changes relate to differences in cell populations or intracellular expression. Furthermore, this leads to mixed competing signals that may be an issue for identifying enriched pathways—this may explain why the number of pathways enriched in MIEAA2 miRNA pathways analysis did not correlate with the number of DE miRNAs, for example, COVID-19 and typhoid

onset were associated with dramatic changes in miRNA expression but a surprisingly low number of enriched pathways.

An alternative to using bulk cell populations would be to sequence sorted cells, but then a decision must be made about which cells to sequence, and even then, subsets exist within subsets. Single-cell sRNA-seq is not sufficiently developed to be applied to samples in this thesis. Deconvolution algorithms can infer changes within cells in mixed cell populations, e.g. CIBERSORTX but depend on genes uniquely expressed within a cell population, which miRNAs tend not to be. The increase in miRNA cell atlases may mean methods like CIBERSORTX can be used in the future. Following up interesting findings, e.g. what drives hsa-miR-1303 upregulation in acute typhoid, can be done by performing RT-PCR on sorted cells from samples taken at baseline and diagnosis. In addition, prime-flow assay technologies that combine flow cytometry with miRNA probe-based detection could also be used to identify which cells a DE miRNA arises from whilst inferring whether or not changes in expression are due to changes in intracellular expression.

Identifying cell types that drive differential expression signals would be especially beneficial in VAST because CyTOF or mRNA sequencing data were not collected 10 days after vaccination in VAST, yet the miRNA data suggested that is the optimal time to look for correlates of protection. Integrating sRNA, CyToF and mRNA data across VAST samples, e.g. correlation and clustering analysis, may infer what happens at the transcriptomic and cellular level 10 days post-vaccination. Work is underway to address this.

8.4.1.3 Descriptive analysis

Much of this thesis' analyses were descriptive, limiting the understanding of the functional roles of specific miRNAs in host response to vaccination and infection. Despite a prioritisation of bioinformatics analysis over *in-vitro* functional work in this thesis, some insights on the potential functional role of miRNA in the host response to vaccination and infection could be gleaned from miRNA pathway analyses, the integration of omics datasets within studies, and comparison of findings across studies. For example, the COVID dataset showed widescale de-repression of mRNA and pathways at COVID-19 onset is linked to a general downregulation of miRNAs. However, this was not seen at the onset of acute typhoid nor in COVID-like illness, potentially providing an insight into the unique pathogenicity of COVID-19.

8.4.1.4 Generalisability of results to other populations

Thesis studies were based on relatively small sample sizes with an over-representation of demographic features seen within study catchment areas (Oxfordshire/South central England)—e.g. minimal racial diversity. Indeed only Caucasian infants were enrolled in the EUCLIDS study. This potentially limits the generalisability of the results to other populations, given population ancestry influences miRNA expression[530,531]. This is a significant limitation, particularly in VAST. Typhoid mainly affects children from typhoid-endemic settings in Asia and Africa, whereas VAST participants were adults, most of whom were of white European ancestry.

8.5 Future work

Avenues for future work (in addition to those described above) include:

1) Identifying the cellular source of DE miRNAs by sorting cell populations or using prime-flow assays
2) Application of machine learning algorithms to identify subnetworks of variation within datasets and over time points, e.g. WGCNA analysis, tensor decomposition, independent component analysis, Time omics and maSigPro—this is especially important for teasing apart miRNA regulatory networks. These methods may infer transcriptomic and cellular responses associated with protection 10 days after Vi-vaccination in VAST.
3) Comparison of sRNA-dfs and their full-length counterparts at sample level as associations may be masked by comparing expression at the average foldchange level
4) GWAS and EQTL analysis to identify genetic variation, which could explain some of the observations within these studies, e.g. correlates of vaccine response. Genotyping studies are underway for all thesis studies.
5) Functional studies—follow up of interesting findings *in-vitro* and possibly *in-vivo* using some of the methods outlined in the introduction, e.g. knockout models.
6) Ideally, some study findings would be confirmed in independent cohorts. The costs of setting up such studies make this unfeasible, but publicly available datasets may emerge, which could act as independent cohorts.

8.6 Conclusion

The goal of this thesis was to use vaccine and infection studies to interrogate the role of sRNAs in the immune response. These studies covered various ages, vaccines, and infections, providing a broad view of sRNAs in the immune system. The findings of this thesis expand upon current knowledge and emphasise the importance of miRNAs in the host response to vaccines and infections. In addition, this thesis opens up exciting avenues for investigating the role of sRNA-dfs in the immune response, hitherto overlooked.

At the start of this thesis, I proposed that it was insufficient to rely on transcriptomics to understand the immune system. Instead, a holistic understanding must extend to knowledge of how those genes are regulated, e.g. by miRNAs. The findings of this thesis support this assertion. Much work is needed to see whether these findings translate into clinical applications. Regardless, the insights of this thesis represent a small but incremental step towards a deeper understanding of immune responses to infection and vaccines. In a world threatened by emerging pandemics and rising antibiotic resistance, the quest for a holistic understanding of the immune response is vital. This thesis contributes to that quest, a quest that will inform the design of novel treatments, diagnostics, and vaccines—improving the health of all.

9 Appendix

9.1 Appendix for Chapter 2

Supplementary Table 1 R packages used for data analysis

mixOmics	Data integration	data.table	Data wrangling	GEOquery	Retrieval of external Data
preprocessCore	Data normalisation	dplyr	Data wrangling	rstatix	Statitistical tests
circlize	Data visualisation	magrittr	Data wrangling	stringi	Text manipulation
cowplot	Data visualisation	plyr	Data wrangling	stringr	Text manipulation
factoextra	Data visualisation	reshape	Data wrangling	textclean	Text manipulation
formattable	Data visualisation	reshape2	Data wrangling	igraph	Data visualisation and network analysis
ggnewscale	Data visualisation	tidyr	Data wrangling		
ggplot2	Data visualisation	edgeR	Differential expression analysis		
ggpubr	Data visualisation	limma	Differential expression analysis		
ggrepel	Data visualisation	readr	File input		
ggvenn	Data visualisation	readxl	File input		
Glimma	Data visualisation	rrvgo	Gene set analysis		
gplots	Data visualisation	fgsea	Gene set analysis		
pheatmap	Data visualisation	GOSim	Gene set analysis		
plotly	Data visualisation	gprofiler2	Gene set analysis		
RColorBrewer	Data visualisation	GSA	Gene set analysis		
rgl	Data visualisation	utils	generic		
Rtsne	Data visualisation	vcd	Gene set analysis		
UpsetR	Data	pacma	Package		

	visualisation	n	management
RCy3	Data visualisation and network analysis	biomaRt	Retrieval of external Data

9.2 Appendix for Chapter 3

Appendix Table 1 RT-qPCR assay IDs

Assay name	**miRCURY LNA miRNA PCR Assays. Cat. No. / ID: 339306 Assay ID:**	**TaqMan Advanced miRNA assays - product number:**
Hsa-miR-122-5p/ Hmu-miR-122-5p	YP00205664	477855_mir
Hsa-miR-223-5p	YP00204529	477984_mir
Hsa-miR-483-5p	YP00205693	478432_mir
Hsa-miR-4497	(Did not work)	479836_mir
Hsa-miR-576-3p		478164_mir
Hsa-miR-128-3p	YP02108212	477892_mir
Hsa-miR-320b	YP00205995	478588_mir
Hsa-miR-378a-3p	YP02119299	478349_mir
Hsa-miR-24-3p	YP00205946	477992_mir
Hsa-miR-27a-3p	YP00204260	478384_mir
Hsa-miR-29a-3p	YP00206038	478587_mir
Mmu-30b-5p	YP00204698	-
Mmu-103a-3p	YP00204765	-
UniSp6		-
UniSp4	YP00203953	-
Cel-miR-39-3p	YP00204063	-
U6 SnRNA (mouse/human)	YP00203952	-

Appendix Table 2 Demographics of participants whose samples were assayed for sRNAs and/or proteomics

					4 Month Cohort								12 Month Cohort							
Participant ID (anonymous)	Vaccine Group	Post vaccination timepoint	Sex	Meningococcus Group B SBA titre at V6/V7	Fever above 38 °C	Max Temp in first 24 hours post vaccination (°C)	paracetamol given in 1st 24hrs?	Number paracetamol doses (1st 24hrs)	sRNA sequencing	RT-PCR	Mass Spec Proteomics	Luminex	Fever above 38 °C	Max Temp in first 24 hours post vaccination (°C)	paracetamol given in 1st 24hrs?	Number paracetamol doses (1st 24hrs)	Max Temp in first 24 hours post vaccination (°C)	sRNA sequencing	RT-PCR	Mass Spec Proteomics	Luminex
8768	Routine	24 hours	Male	<2	No Fever	37.5	Yes	2	No	No	No	No	No Fever	36.75	Yes	1	#N/A	No	No	Yes	Yes
153	Routine	24 hours	Female	#N/A	No Fever	37.125	#N/A	NA	No	No	No	No	No Fever	37.875	No	NA	#N/A	No	No	Yes	Yes
8526	4CMenB + Routine	24 hours	Female	4	Fever	38	Yes	1	No	No	No	No	Fever	38	No	NA	#N/A	No	No	Yes	Yes
88	4CMenB + Routine	24 hours	Male	#N/A	Fever	38.75	Yes	1	Yes	Yes	No	No	Fever	38	No	NA	#N/A	No	No	Yes	Yes
91	4CMenB + Routine	24 hours	Female	64	Fever	38.125	#N/A	NA	Yes	Yes	No	No	Fever	38.25	Yes	1	#N/A	No	No	No	No
138	4CMenB + Routine	24 hours	Female	128	No Fever	37.25	No	NA	No	No	No	No	Fever	38.375	Yes	1	#N/A	No	No	Yes	Yes
61	Routine	24 hours	Male	4	No Fever	37.625	Yes	1	Yes	Yes	No	No	Fever	38.5	No	NA	#N/A	No	No	No	No
121	4CMenB + Routine	24 hours	Male	#N/A	Fever	38	#N/A	NA	No	No	No	No	Fever	39.25	No	NA	#N/A	No	No	Yes	Yes
38	Routine	24 hours	Female	<2	No Fever	37.875	No	NA	No	No	No	No	No Fever	37.875	No	NA	37.875	Yes	Yes	Yes	Yes
2	Routine	24 hours	Female	#N/A	No Fever	37.25	No	NA	No	No	No	No	Fever	38.25	No	NA	38.25	Yes	Yes	Yes	Yes
69	Routine	24 hours	Male	2	Fever	38.75	No	NA	No	No	No	No	Fever	38.5	No	NA	38.5	Yes	Yes	Yes	Yes
39	4CMenB + Routine	24 hours	Female	#N/A	Fever	38.125	No	NA	No	No	No	No	Fever	39	#N/A	NA	39	Yes	Yes	Yes	Yes
70	4CMenB + Routine	24 hours	Male	128	Fever	38.125	Yes	1	No	No	No	No	Fever	39.125	Yes	1	39.125	Yes	Yes	Yes	Yes
93	4CMenB + Routine	24 hours	Male	32	Fever	38.625	Yes	1	No	No	No	No	Fever	39.25	No	NA	39.25	Yes	Yes	Yes	Yes
90	4CMenB + Routine	24 hours	Male	128	Fever	39.625	Yes	2	No	No	No	No	Fever	39.25	Yes	1	39.25	Yes	Yes	No	No
84	Routine	4 hours	Female	8	No Fever	37	No	NA	No	No	No	No	No Fever	36.875	Yes	1	#N/A	No	No	Yes	Yes
77	4CMenB + Routine	4 hours	Female	64	No Fever	37.125	Yes	NA	No	No	No	No	No Fever	37	#N/A	NA	#N/A	No	No	Yes	Yes
122	Routine	4 hours	Female	2	No Fever	37.125	#N/A	NA	No	No	No	No	No Fever	37.125	No	NA	#N/A	No	No	Yes	Yes
110	Routine	4 hours	Male	4	No Fever	37.5	Yes	1	No	No	No	No	No Fever	37.125	Yes	1	#N/A	No	No	Yes	Yes
140	Routine	4 hours	Female	<2	Fever	38.5	#N/A	NA	No	No	No	No	No Fever	37.25	No	NA	#N/A	No	No	Yes	Yes
8935	4CMenB + Routine	4 hours	Female	2	Fever	38.625	Yes	1	No	No	No	No	No Fever	37.5	Yes	2	#N/A	No	No	Yes	Yes
127	Routine	4 hours	Female	4	No Fever	37.125	No	1	No	No	No	No	No Fever	37.625	No	NA	#N/A	No	No	Yes	Yes
21	4CMenB + Routine	4 hours	Female	4	Fever	39.125	Yes	NA	No	No	No	No	No Fever	37.625	Yes	1	#N/A	No	No	Yes	Yes
139	Routine	4 hours	Female	4	Fever	39.25	Yes	2	No	No	No	No	No Fever	37.75	No	NA	#N/A	No	No	Yes	Yes
58	Routine	4 hours	Male	8	#N/A	#N/A	No	NA	No	No	No	No	No Fever	37.75	No	NA	#N/A	No	No	Yes	Yes
94	4CMenB + Routine	4 hours	Male	128	Fever	38.25	Yes	1	No	No	No	No	Fever	38.125	Yes	1	#N/A	No	No	Yes	Yes
40	Routine	4 hours	Male	#N/A	No Fever	37.875	Yes	2	No	No	No	No	Fever	38.25	No	NA	#N/A	No	No	Yes	Yes
135	4CMenB + Routine	4 hours	Male	8	No Fever	37.5	#N/A	NA	No	No	No	No	Fever	38.375	No	NA	#N/A	No	No	Yes	Yes
56	Routine	4 hours	Male	2	Fever	38	No	NA	No	No	No	No	Fever	38.375	No	NA	#N/A	No	No	Yes	Yes
106	4CMenB + Routine	4 hours	Male	32	Fever	38.375	#N/A	NA	No	No	No	No	Fever	38.5	Yes	1	#N/A	No	No	Yes	Yes
8593	4CMenB + Routine	24 hours	Male	#N/A	Fever	38	Yes	1	No	No	Yes	Yes	No Fever	35.875	Yes	1	38	No	No	No	No
112	4CMenB + Routine	24 hours	Male	16	Fever	38.375	#N/A	NA	Yes	Yes	Yes	Yes	No Fever	37.375	No	NA	#N/A	No	No	No	No
27	4CMenB + Routine	24 hours	Female	32	Fever	38.875	#N/A	NA	Yes	Yes	Yes	Yes	No Fever	37.5	No	NA	#N/A	No	No	No	No
123	Routine	24 hours	Female	4	No Fever	37.5	No	NA	Yes	Yes	Yes	Yes	No Fever	37.625	#N/A	NA	#N/A	No	No	No	No
25	Routine	24 hours	Female	#N/A	No Fever	37.5	#N/A	NA	Yes	Yes	Yes	Yes	No Fever	37.625	#N/A	NA	#N/A	No	No	No	No
152	Routine	24 hours	Female	2	No Fever	37.125	No	NA	Yes	Yes	Yes	Yes	Fever	39.25	Yes	1	#N/A	No	No	No	No
8123	Routine	24 hours	Male	#N/A	No Fever	36.5	No	NA	Yes	Yes	Yes	Yes	No Fever	37	No	NA	37	Yes	Yes	Yes	Yes
143	Routine	24 hours	Female	<2	No Fever	38.125	No	NA	Yes	Yes	Yes	Yes	No Fever	37	#N/A	NA	37	Yes	Yes	Yes	Yes
103	4CMenB + Routine	24 hours	Male	#N/A	No Fever	36.125	Yes	2	Yes	Yes	Yes	Yes	No Fever	37.125	Yes	2	37.125	Yes	Yes	Yes	Yes
23	4CMenB + Routine	24 hours	Female	128	Fever	39.75	#N/A	NA	Yes	Yes	Yes	Yes	No Fever	37.5	#N/A	NA	37.5	Yes	Yes	Yes	Yes
133	4CMenB + Routine	24 hours	Female	#N/A	#N/A	#N/A	No	NA	Yes	Yes	Yes	Yes	No Fever	37.5	No	NA	37.5	Yes	Yes	Yes	Yes
8892	Routine	24 hours	Male	4	No Fever	36.875	Yes	1	Yes	Yes	Yes	Yes	No Fever	37.625	Yes	1	37.625	Yes	Yes	Yes	Yes
19	4CMenB + Routine	24 hours	Female	256	No Fever	37.75	Yes	1	Yes	Yes	Yes	Yes	No Fever	37.625	No	NA	37.625	Yes	Yes	Yes	Yes
33	Routine	24 hours	Male	<2	Fever	39.25	No	NA	Yes	Yes	Yes	Yes	No Fever	37.75	No	NA	37.75	Yes	Yes	No	No
97	Routine	24 hours	Female	2	No Fever	37.5	Yes	2	Yes	Yes	Yes	Yes	Fever	38	No	NA	38	Yes	Yes	Yes	Yes
8087	4CMenB + Routine	24 hours	Female	16	Fever	38.25	Yes	1	Yes	Yes	Yes	Yes	Fever	38	Yes	1	38	Yes	Yes	Yes	Yes
53	4CMenB + Routine	24 hours	Male	64	No Fever	37.75	No	NA	Yes	Yes	Yes	Yes	Fever	38.375	No	NA	38.375	Yes	Yes	Yes	Yes
149	Routine	24 hours	Female	2	No Fever	37.5	Yes	1	Yes	Yes	Yes	Yes	Fever	38.5	No	NA	38.5	Yes	Yes	Yes	Yes
52	Routine	24 hours	Female	4	Fever	38	#N/A	NA	Yes	Yes	Yes	Yes	#N/A	#N/A	No	NA	NA	Yes	Yes	Yes	Yes
46	4CMenB + Routine	4 hours	Male	16	No Fever	37.375	Yes	1	No	Yes	Yes	Yes	No Fever	36.75	Yes	1	#N/A	No	No	Yes	Yes
30	Routine	4 hours	Male	<2	No Fever	37.25	No	NA	No	Yes	Yes	Yes	No Fever	36.875	No	NA	#N/A	No	No	No	No
3	Routine	4 hours	Male	<2	No Fever	37.625	Yes	1	No	Yes	Yes	Yes	No Fever	36.875	No	NA	#N/A	No	No	Yes	Yes
8274	Routine	4 hours	Male	<2	No Fever	37.625	No	NA	No	Yes	Yes	Yes	No Fever	36.875	No	NA	#N/A	No	No	No	No
102	Routine	4 hours	Male	2	No Fever	36.5	#N/A	NA	No	Yes	Yes	Yes	No Fever	37.25	No	NA	#N/A	No	No	No	No
146	Routine	4 hours	Male	16	No Fever	37.5	No	NA	No	Yes	Yes	Yes	No Fever	37.625	No	NA	#N/A	No	No	No	No
47	Routine	4 hours	Female	4	No Fever	37.875	Yes	NA	No	Yes	Yes	Yes	No Fever	37.875	No	NA	#N/A	No	No	No	No
82	4CMenB + Routine	4 hours	Male	#N/A	No Fever	37.5	Yes	1	No	Yes	Yes	Yes	Fever	38.375	Yes	2	#N/A	No	No	No	No
14	4CMenB + Routine	4 hours	Female	128	Fever	38	Yes	1	No	Yes	Yes	Yes	Fever	38.375	Yes	2	#N/A	No	No	Yes	Yes
20	Routine	4 hours	Male	#N/A	No Fever	36.875	Yes	1	No	Yes	Yes	Yes	Fever	38.875	Yes	2	#N/A	No	No	Yes	Yes
65	4CMenB + Routine	4 hours	Male	#N/A	Fever	38.5	Yes	1	No	Yes	Yes	Yes	Fever	39.25	Yes	2	#N/A	No	No	Yes	Yes
137	Routine	4 hours	Male	<2	No Fever	37.125	No	NA	No	Yes	Yes	Yes	#N/A	#N/A	Yes	2	#N/A	No	No	No	No
134	4CMenB + Routine	4 hours	Male	32	No Fever	37.875	No	NA	No	Yes	Yes	Yes	#N/A	#N/A	No	NA	#N/A	No	No	No	No

9.2.1 Associations between miRNA expression and degree of sample haemolysis.

An association between sample haemolysis and miRNA expression was analysed for pre- and post-vaccination samples separately to see if miRNAs that were DE after vaccination were associated with sample haemolysis before or after vaccination. This helps to isolate the effect of haemolysis on the miRNAome from the effect of vaccination

ppendix Table 3 miRNAs significantly (FDR <0.05) associated with the degree of sample haemolysis in 4-month-old baseline vaccine samples after controlling for participant and vaccine status. Highlighted miRNAs were also associated with haemolysis at 12-months of age at p<0.05. The greatest associations were found in pre-vaccine samples—this may be because of partial obscuration by a vaccination signal in post-vaccine samples

miRNA	logFC	log CPM	LR	p-value	FDR
hsa-miR-10a-5p	-1.41	11.17	49.53	0.00	0.00
hsa-miR-375-3p	-1.71	10.68	42.73	0.00	0.00
hsa-let-7g-5p	1.19	14.23	40.20	0.00	0.00
hsa-miR-30a-5p	-1.25	10.75	40.18	0.00	0.00
hsa-miR-483-5p	-1.60	11.19	40.17	0.00	0.00
hsa-miR-483-3p	-4.14	7.12	39.77	0.00	0.00
hsa-miR-122-5p	-1.30	16.42	38.01	0.00	0.00
hsa-miR-3168	-1.44	11.46	35.76	0.00	0.00
hsa-miR-182-5p	1.65	10.18	34.91	0.00	0.00
hsa-miR-1183	-1.46	10.38	32.10	0.00	0.00
hsa-let-7i-5p	0.98	14.92	31.73	0.00	0.00
hsa-let-7b-5p	1.01	15.36	29.52	0.00	0.00
hsa-miR-134-5p	-1.37	10.40	28.69	0.00	0.00
hsa-miR-485-3p	-1.90	7.54	26.62	0.00	0.00
hsa-miR-4433b-5p	-2.40	6.18	26.15	0.00	0.00
hsa-miR-10b-5p	-1.08	12.47	25.16	0.00	0.00
hsa-miR-493-5p	-1.29	8.83	24.85	0.00	0.00
hsa-miR-409-3p	-1.16	11.61	24.69	0.00	0.00
hsa-miR-183-5p	1.63	9.73	24.45	0.00	0.00
hsa-miR-125a-5p	-1.10	10.49	23.35	0.00	0.00
hsa-miR-155-5p	-1.12	10.35	22.34	0.00	0.00
hsa-miR-7-5p	1.15	10.16	22.27	0.00	0.00
hsa-miR-106b-3p	1.02	9.64	22.00	0.00	0.00
hsa-miR-382-5p	-1.00	10.79	20.80	0.00	0.00
hsa-miR-200a-3p	-1.76	6.26	20.53	0.00	0.00
hsa-let-7e-5p	-1.09	10.36	20.26	0.00	0.00
hsa-miR-323b-3p	-1.40	8.41	20.19	0.00	0.00
hsa-miR-215-5p	-1.82	7.37	19.49	0.00	0.00
hsa-miR-27b-3p	-0.79	12.24	19.07	0.00	0.00
hsa-miR-17-5p	1.14	8.93	18.60	0.00	0.00
hsa-miR-191-5p	0.76	11.56	18.32	0.00	0.00
hsa-miR-126-3p	-0.84	11.81	17.90	0.00	0.00
hsa-miR-370-3p	-1.27	7.74	17.10	0.00	0.00
hsa-miR-20a-5p	1.33	10.38	16.73	0.00	0.00

hsa-miR-532-5p	0.76	10.96	16.44	0.00	0.00
hsa-miR-25-3p	1.12	16.25	16.19	0.00	0.00
hsa-miR-6773-5p	-1.60	6.38	16.17	0.00	0.00
hsa-miR-224-5p	-1.08	9.72	16.00	0.00	0.00
hsa-miR-432-5p	-1.14	10.95	15.62	0.00	0.00
hsa-miR-16-5p	1.20	11.17	14.31	0.00	0.00
hsa-miR-486-3p	1.39	11.09	14.23	0.00	0.00
hsa-miR-127-3p	-0.83	10.67	14.19	0.00	0.00
hsa-miR-889-3p	-1.40	6.67	14.01	0.00	0.00
hsa-miR-93-5p	1.29	10.28	13.95	0.00	0.00
hsa-miR-433-3p	-0.99	8.15	13.94	0.00	0.00
hsa-miR-199a-3p	-0.68	11.13	13.72	0.00	0.00
hsa-let-7d-5p	0.71	11.14	13.51	0.00	0.00
hsa-miR-4433b-3p	-1.40	8.87	13.49	0.00	0.00
hsa-miR-185-5p	0.95	12.78	13.35	0.00	0.00
hsa-miR-146a-5p	-0.56	12.52	13.06	0.00	0.00
hsa-miR-92a-3p	0.89	16.23	13.04	0.00	0.00
hsa-miR-26b-5p	0.87	10.03	13.02	0.00	0.00
hsa-miR-423-5p	0.64	16.03	12.85	0.00	0.00
hsa-miR-2110	0.81	8.49	12.77	0.00	0.00
hsa-miR-654-3p	-1.44	6.97	12.58	0.00	0.00
hsa-let-7a-5p	0.68	15.49	12.14	0.00	0.00
hsa-miR-323a-3p	-0.96	8.01	12.06	0.00	0.00
hsa-miR-363-3p	1.18	10.15	11.95	0.00	0.00
hsa-miR-101-3p	0.91	13.58	11.77	0.00	0.00
hsa-miR-99b-5p	-0.82	9.87	11.76	0.00	0.00
hsa-miR-99a-5p	-0.67	10.78	11.36	0.00	0.00
hsa-miR-342-5p	0.75	8.20	11.12	0.00	0.00
hsa-miR-379-5p	-1.07	8.42	10.76	0.00	0.00
hsa-miR-629-5p	0.80	10.55	10.62	0.00	0.00
hsa-miR-142-5p	0.79	10.08	10.58	0.00	0.00
hsa-miR-136-3p	-1.47	6.35	10.24	0.00	0.00
hsa-miR-221-3p	-0.64	10.85	9.39	0.00	0.01
hsa-miR-361-3p	0.62	8.85	9.36	0.00	0.01
hsa-miR-1294	1.89	5.58	9.23	0.00	0.01
hsa-miR-455-5p	-1.24	6.26	9.02	0.00	0.01
hsa-miR-451a	1.29	13.56	8.85	0.00	0.01
hsa-miR-103a-3p	0.64	10.29	8.83	0.00	0.01
hsa-miR-98-5p	0.76	8.73	8.80	0.00	0.01
hsa-miR-1260a	-1.43	5.32	8.73	0.00	0.01
hsa-miR-130b-5p	-0.97	5.98	8.62	0.00	0.01
hsa-miR-199a-5p	-0.71	9.34	8.58	0.00	0.01
hsa-miR-452-5p	-1.68	5.48	8.37	0.00	0.01
hsa-miR-106b-5p	1.41	7.06	8.27	0.00	0.01
hsa-miR-425-3p	1.07	6.05	8.25	0.00	0.01
hsa-miR-1285-3p	1.20	5.79	8.15	0.00	0.01
hsa-miR-186-5p	0.93	9.66	8.11	0.00	0.01

hsa-miR-125b-5p	-0.67	8.15	7.93	0.00	0.01
hsa-miR-144-3p	1.55	8.01	7.89	0.00	0.01
hsa-miR-139-5p	-1.10	6.59	7.55	0.01	0.02
hsa-miR-11400	-1.70	5.37	7.52	0.01	0.02
hsa-miR-4508	1.74	5.88	7.24	0.01	0.02
hsa-miR-576-3p	1.14	6.55	7.15	0.01	0.02
hsa-miR-5010-5p	0.85	7.07	6.94	0.01	0.02
hsa-miR-206	-0.89	7.43	6.91	0.01	0.02
hsa-miR-543	-0.72	8.30	6.91	0.01	0.02
hsa-miR-485-5p	-0.75	7.93	6.78	0.01	0.02
hsa-miR-20b-5p	1.94	7.46	6.56	0.01	0.03
hsa-miR-199b-5p	1.88	5.56	6.38	0.01	0.03
hsa-miR-24-3p	-0.43	11.80	6.24	0.01	0.03
hsa-miR-941	0.70	8.05	6.22	0.01	0.03
hsa-miR-152-3p	-0.52	9.02	6.14	0.01	0.03
hsa-miR-29a-3p	-0.47	9.16	6.06	0.01	0.03
hsa-miR-652-3p	1.40	5.62	5.84	0.02	0.04
hsa-miR-30e-5p	0.66	11.29	5.74	0.02	0.04
hsa-miR-4446-3p	-1.15	6.94	5.68	0.02	0.04
hsa-miR-501-3p	0.65	8.47	5.63	0.02	0.04
hsa-miR-25-5p	0.89	6.49	5.62	0.02	0.04
hsa-miR-193b-5p	-0.80	7.44	5.61	0.02	0.04
hsa-miR-1180-3p	1.08	6.87	5.59	0.02	0.04
hsa-miR-30a-3p	-0.65	7.90	5.58	0.02	0.04
hsa-miR-27a-3p	-0.40	10.87	5.47	0.02	0.04
hsa-miR-486-5p	0.87	18.28	5.45	0.02	0.04
hsa-miR-126-5p	-0.55	8.29	5.42	0.02	0.04
hsa-miR-106a-5p	1.68	5.53	5.41	0.02	0.04
hsa-miR-381-3p	-0.76	7.53	5.39	0.02	0.04
hsa-miR-3131	-0.93	6.46	5.28	0.02	0.05
hsa-let-7d-3p	-0.40	11.32	5.28	0.02	0.05
hsa-miR-222-3p	-0.56	9.56	5.21	0.02	0.05
hsa-miR-505-5p	1.02	6.20	5.20	0.02	0.05
Hs	0.48	14.59	5.12	0.02	0.05

Appendix Table 4 miRNAs significantly (FDR <0.05) associated with the degree of sample haemolysis in 4-month-old baseline vaccine samples

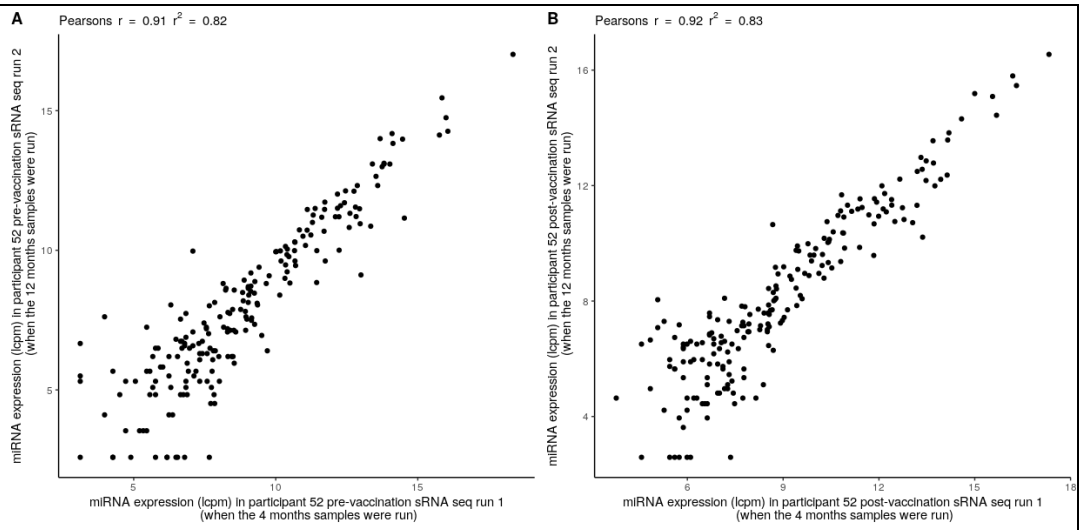
miRNA	logFC	logCPM	LR	PValue	FDR
hsa-miR-7-5p	1.45	10.14	58.57	0.00	0.00
hsa-miR-532-5p	1.18	11.05	55.48	0.00	0.00
hsa-miR-182-5p	1.77	9.77	52.88	0.00	0.00
hsa-miR-106b-3p	1.30	9.78	51.12	0.00	0.00
hsa-miR-16-5p	1.58	11.15	51.09	0.00	0.00
hsa-miR-485-3p	-2.47	7.46	44.57	0.00	0.00
hsa-miR-25-3p	1.45	16.20	42.82	0.00	0.00
hsa-miR-486-3p	1.82	11.12	38.61	0.00	0.00
hsa-miR-183-5p	1.92	9.50	37.85	0.00	0.00
hsa-miR-101-3p	1.22	13.60	37.00	0.00	0.00
hsa-let-7i-5p	0.97	14.98	36.35	0.00	0.00
hsa-miR-185-5p	1.20	12.76	36.16	0.00	0.00
hsa-miR-92a-3p	1.13	16.22	35.67	0.00	0.00
hsa-miR-4508	2.10	5.98	33.18	0.00	0.00
hsa-miR-125a-5p	-1.10	10.40	32.68	0.00	0.00
hsa-miR-10a-5p	-1.07	11.15	30.77	0.00	0.00
hsa-miR-363-3p	1.34	10.04	30.22	0.00	0.00
hsa-miR-26a-5p	-1.23	12.28	29.98	0.00	0.00
hsa-miR-629-5p	1.11	10.49	28.49	0.00	0.00
hsa-miR-93-5p	1.15	10.29	28.07	0.00	0.00
hsa-miR-493-5p	-1.52	8.77	27.28	0.00	0.00
hsa-miR-20a-5p	1.13	10.43	26.97	0.00	0.00
hsa-miR-126-3p	-0.97	11.76	26.67	0.00	0.00
hsa-miR-423-5p	0.84	15.99	25.88	0.00	0.00
hsa-miR-99b-5p	-0.97	9.85	25.34	0.00	0.00
hsa-miR-486-5p	1.57	18.32	24.97	0.00	0.00
hsa-miR-186-5p	1.10	9.66	24.22	0.00	0.00
hsa-miR-2110	0.92	8.50	23.95	0.00	0.00
hsa-miR-30a-5p	-1.03	10.78	23.02	0.00	0.00
hsa-miR-155-5p	-1.03	10.23	23.00	0.00	0.00
hsa-let-7g-5p	0.76	13.97	21.71	0.00	0.00
hsa-miR-451a	1.49	13.41	21.38	0.00	0.00
hsa-miR-134-5p	-1.17	10.26	21.31	0.00	0.00
hsa-miR-323b-3p	-1.30	8.31	20.85	0.00	0.00
hsa-miR-502-3p	1.50	5.95	20.08	0.00	0.00
hsa-miR-576-3p	1.39	6.42	19.05	0.00	0.00
hsa-miR-221-3p	-1.09	10.91	18.94	0.00	0.00
hsa-miR-3168	-0.92	11.24	17.59	0.00	0.00
hsa-miR-1294	1.80	5.75	17.39	0.00	0.00
hsa-let-7b-5p	0.78	15.22	17.32	0.00	0.00
hsa-miR-4433b-5p	-1.86	6.14	17.30	0.00	0.00
hsa-miR-199a-5p	-1.09	9.42	17.14	0.00	0.00

hsa-miR-136-3p	-1.66	6.39	16.96	0.00	0.00
hsa-miR-142-3p	-1.26	7.44	16.85	0.00	0.00
hsa-miR-17-5p	0.81	8.92	16.72	0.00	0.00
hsa-miR-409-3p	-1.11	11.54	16.61	0.00	0.00
hsa-miR-1180-3p	1.39	7.00	16.54	0.00	0.00
hsa-miR-1285-3p	1.86	5.77	16.44	0.00	0.00
hsa-miR-2392	-2.32	5.72	16.43	0.00	0.00
hsa-miR-654-3p	-1.36	6.82	16.14	0.00	0.00
hsa-miR-5010-5p	1.24	7.12	15.63	0.00	0.00
hsa-miR-323a-3p	-1.18	7.97	15.60	0.00	0.00
hsa-miR-382-5p	-0.98	10.79	15.57	0.00	0.00
hsa-miR-543	-1.29	8.24	15.39	0.00	0.00
hsa-miR-501-3p	0.94	8.49	15.31	0.00	0.00
hsa-let-7e-5p	-1.07	10.09	15.22	0.00	0.00
hsa-miR-941	1.01	7.87	15.15	0.00	0.00
hsa-miR-340-5p	-0.72	8.86	14.54	0.00	0.00
hsa-miR-378a-3p	0.57	11.69	14.42	0.00	0.00
hsa-miR-92b-5p	1.66	5.46	14.30	0.00	0.00
hsa-miR-146a-5p	-0.67	12.55	13.89	0.00	0.00
hsa-miR-224-5p	-1.00	9.61	13.86	0.00	0.00
hsa-miR-144-3p	1.22	7.75	13.81	0.00	0.00
hsa-miR-1224-5p	1.60	5.89	13.67	0.00	0.00
hsa-miR-199a-3p	-0.74	11.20	13.64	0.00	0.00
hsa-miR-215-5p	-1.17	7.31	13.61	0.00	0.00
hsa-miR-22-3p	0.54	11.83	13.57	0.00	0.00
hsa-miR-106a-5p	1.81	5.23	13.22	0.00	0.00
hsa-miR-142-5p	0.73	9.91	13.19	0.00	0.00
hsa-miR-1-3p	-1.53	6.83	13.04	0.00	0.00
hsa-miR-222-3p	-0.84	9.59	12.85	0.00	0.00
hsa-miR-27b-3p	-0.58	12.33	12.58	0.00	0.00
hsa-miR-10b-5p	-0.75	12.35	12.49	0.00	0.00
hsa-miR-4433b-3p	-1.38	8.64	12.46	0.00	0.00
hsa-miR-200a-3p	-1.31	6.46	12.23	0.00	0.00
hsa-miR-152-3p	-0.70	9.15	12.16	0.00	0.00
hsa-miR-4732-5p	1.73	7.66	12.13	0.00	0.00
hsa-miR-942-5p	1.42	5.70	11.92	0.00	0.00
hsa-miR-4446-3p	-1.35	7.09	11.60	0.00	0.00
hsa-miR-24-3p	-0.57	11.66	11.47	0.00	0.00
hsa-miR-3615	0.50	9.26	10.94	0.00	0.00
hsa-miR-432-5p	-0.99	10.66	10.88	0.00	0.00
hsa-miR-4732-3p	1.26	7.37	10.83	0.00	0.00
hsa-miR-379-5p	-1.09	8.28	10.66	0.00	0.00
hsa-miR-223-5p	-0.70	9.62	10.37	0.00	0.00
hsa-miR-483-3p	-1.99	7.41	10.36	0.00	0.00
hsa-miR-99a-5p	-0.56	10.85	10.31	0.00	0.00
hsa-miR-126-5p	-0.80	8.16	10.06	0.00	0.00
hsa-miR-889-3p	-1.28	6.64	10.03	0.00	0.00

hsa-miR-503-5p	0.82	7.71	10.01	0.00	0.00
hsa-miR-3158-3p	1.60	5.73	9.68	0.00	0.00
hsa-miR-25-5p	1.19	6.42	9.64	0.00	0.00
hsa-miR-30e-5p	0.58	11.34	9.57	0.00	0.01
hsa-miR-328-3p	-1.18	6.89	9.48	0.00	0.01
hsa-miR-27a-3p	-0.54	10.74	9.44	0.00	0.01
hsa-miR-127-3p	-0.87	10.74	9.37	0.00	0.01
hsa-miR-1260a	-1.47	5.22	9.28	0.00	0.01
hsa-miR-125b-5p	-0.75	8.29	9.08	0.00	0.01
hsa-miR-20b-5p	1.21	7.31	9.00	0.00	0.01
hsa-miR-760	-1.23	5.54	8.98	0.00	0.01
hsa-miR-1304-5p	1.13	6.14	8.96	0.00	0.01
hsa-miR-28-3p	-0.57	9.94	8.91	0.00	0.01
hsa-miR-1246	0.66	11.73	8.84	0.00	0.01
hsa-let-7a-5p	0.53	15.43	8.76	0.00	0.01
hsa-miR-433-3p	-0.87	8.03	8.70	0.00	0.01
hsa-miR-4485-3p	-2.38	8.02	8.57	0.00	0.01
hsa-miR-381-3p	-0.96	7.50	8.44	0.00	0.01
hsa-miR-11400	-1.94	5.23	8.27	0.00	0.01
hsa-miR-143-3p	-0.68	11.60	8.11	0.00	0.01
hsa-miR-651-5p	0.99	5.47	8.07	0.00	0.01
hsa-miR-21-5p	-0.45	15.06	8.07	0.00	0.01
hsa-miR-6773-5p	-1.13	5.99	7.99	0.00	0.01
hsa-miR-6852-5p	-1.70	5.08	7.94	0.00	0.01
hsa-miR-370-3p	-0.91	7.62	7.86	0.01	0.01
hsa-miR-130b-5p	-0.95	5.92	7.67	0.01	0.01
hsa-miR-375-3p	-0.71	10.75	7.34	0.01	0.01
hsa-miR-29a-3p	-0.52	9.06	7.29	0.01	0.01
hsa-miR-455-5p	-1.02	6.13	7.24	0.01	0.01
hsa-miR-106b-5p	1.05	7.03	7.05	0.01	0.02
hsa-miR-7706	0.96	5.96	7.05	0.01	0.02
hsa-miR-335-3p	-1.88	5.31	6.81	0.01	0.02
hsa-miR-487a-5p	-1.36	5.45	6.72	0.01	0.02
hsa-miR-30a-3p	-0.65	7.80	6.43	0.01	0.02
hsa-miR-654-5p	-1.06	6.38	6.29	0.01	0.02
hsa-miR-589-5p	0.97	5.59	6.10	0.01	0.03
hsa-miR-361-5p	-0.70	6.42	6.06	0.01	0.03
hsa-miR-206	-0.76	7.34	5.86	0.02	0.03
hsa-miR-361-3p	0.44	8.79	5.83	0.02	0.03
hsa-miR-1183	-0.72	10.13	5.81	0.02	0.03
hsa-miR-185-3p	0.80	5.40	5.58	0.02	0.03
hsa-let-7d-5p	0.42	11.08	5.56	0.02	0.03
hsa-let-7d-3p	-0.33	11.27	5.54	0.02	0.03
hsa-miR-452-5p	-1.32	5.58	5.43	0.02	0.04
hsa-miR-103a-3p	0.39	10.22	5.37	0.02	0.04
hsa-miR-193b-5p	-0.68	7.59	5.20	0.02	0.04
hsa-miR-1301-3p	-0.57	7.50	5.17	0.02	0.04

hsa-miR-425-5p	0.66	7.51	5.03	0.02	0.04
hsa-miR-485-5p	-0.69	7.90	5.03	0.02	0.04
hsa-miR-625-3p	-1.23	5.59	4.90	0.03	0.05
hsa-miR-139-5p	-0.98	6.43	4.89	0.03	0.05
hsa-miR-19b-3p	0.88	6.24	4.78	0.03	0.05
hsa-miR-342-5p	0.47	8.23	4.77	0.03	0.05

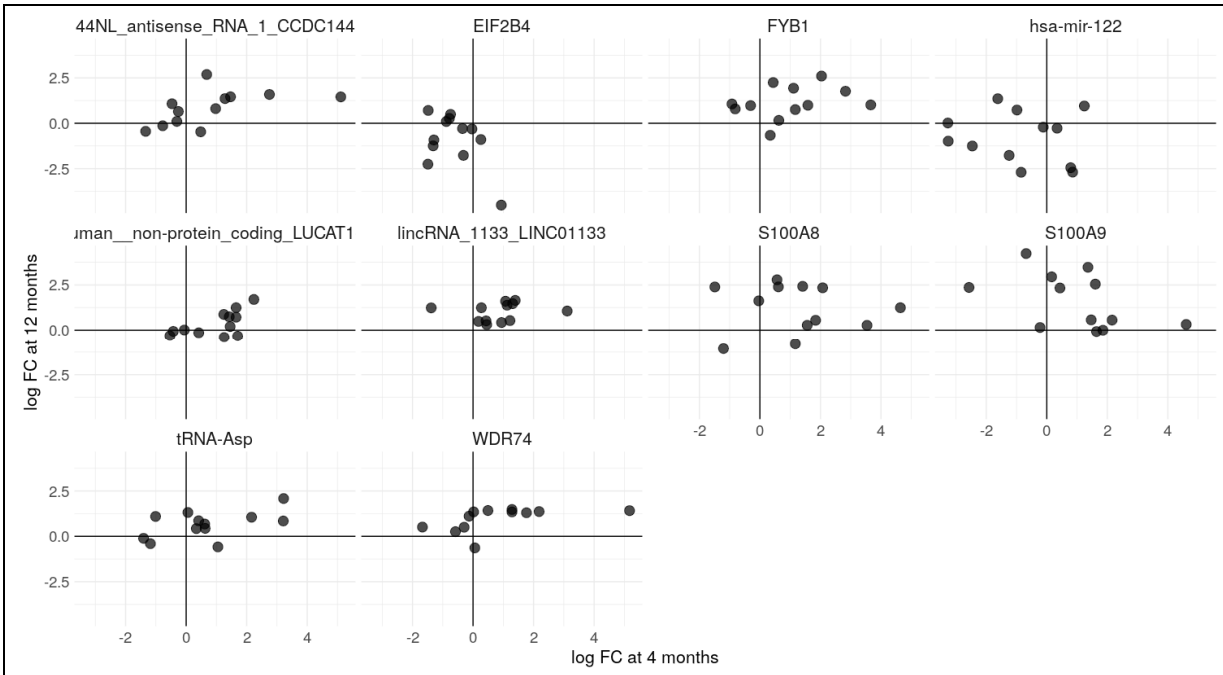
9.2.1 Correlations between duplicate samples across the two EUCLIDS sRNA runs



Appendix Figure 2 Sufficient space was available on the flow cell for two extra samples during the second sRNA-seq run (sRNA-seq run 2) when the 12-month-old Euclid samples were being run. I therefore sequenced two replicate samples that had been run on the 4-month cohort (sRNA-seq run 1). Both samples showed strong correlation in miRNA expression (TMM normalised log

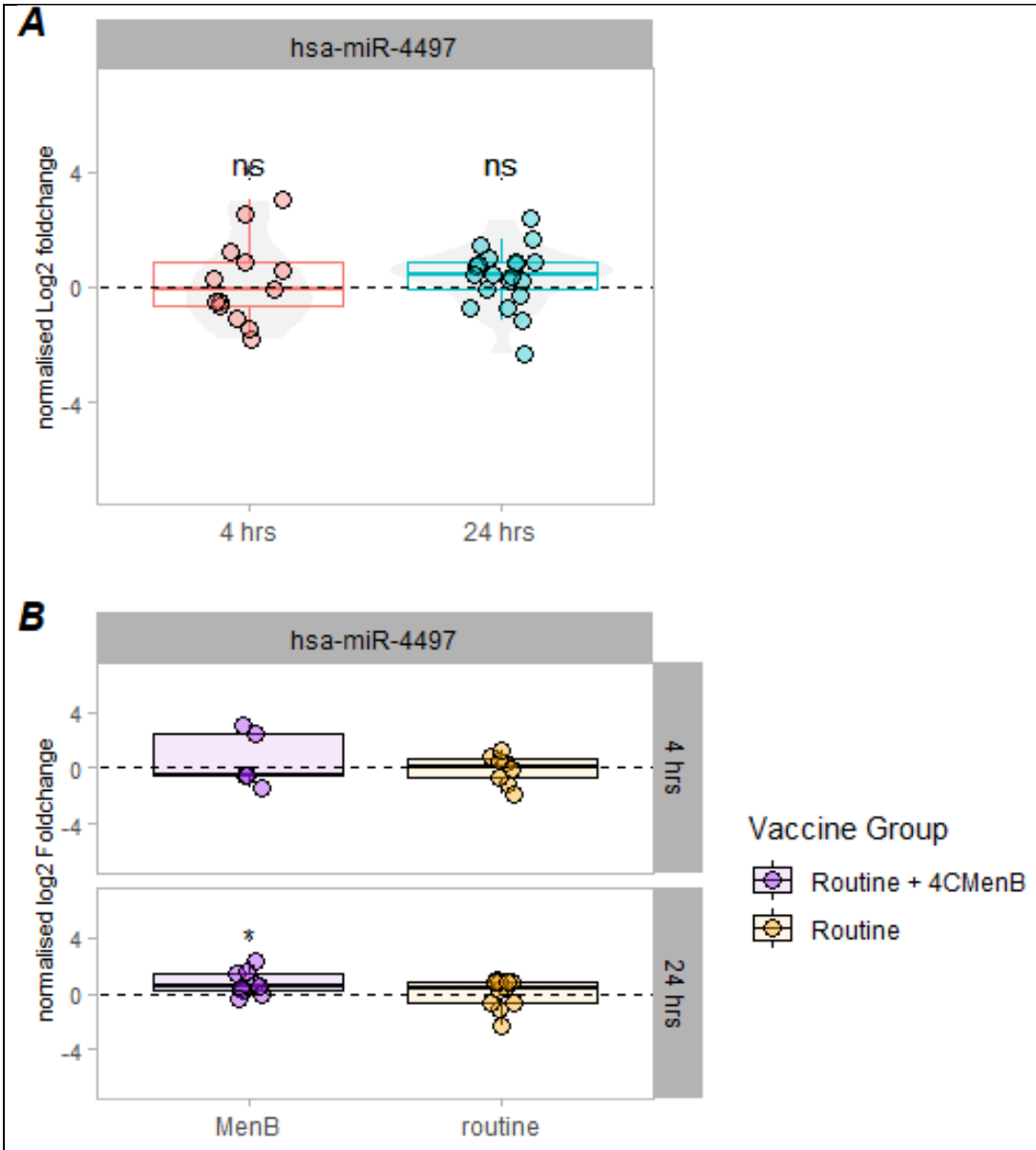
9.2.2 Comparing sRNA log_2 fold changes at 4 months and 12-months

The expression of sRNA seq features DE in the 4 month and 12-month cohort was correlated for each person.



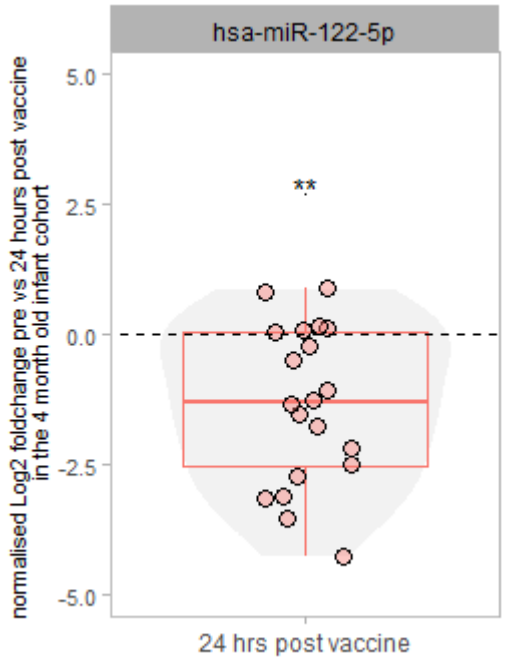
Appendix figure 3

Correlation plot of pre versus post-vaccination log_2 fold change for validated genes in participants sequenced at both 4 months (x-axis) and 12-months (y-axis) of age.

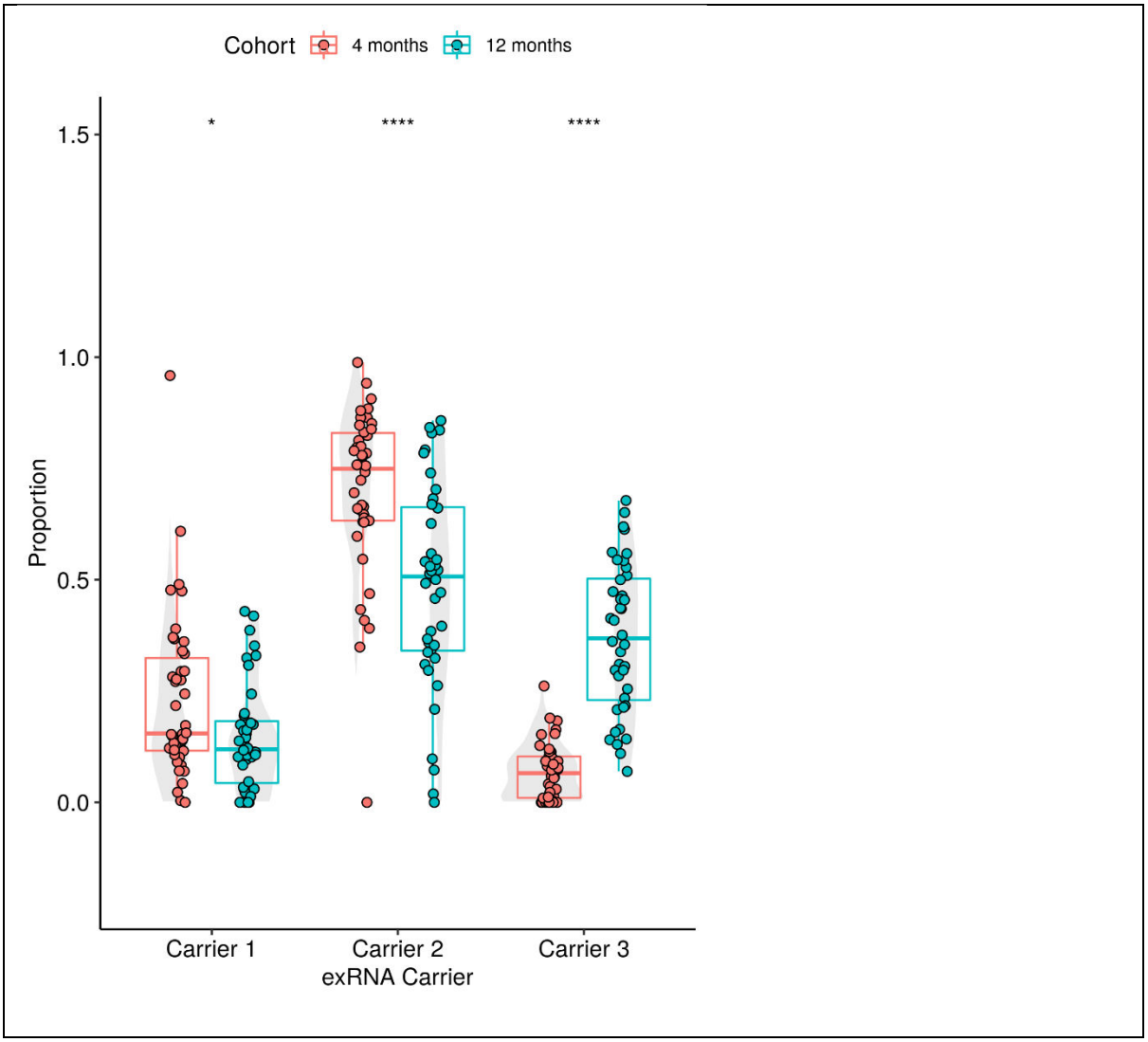


Appendix Figure 4 RT-qPCR results for hsa-miR-4497 in the 4-month-old EUCLIDS cohort. This miRNA was tested as initial analysis showed unilateral differential expression in the 4CMenB group. The RT-qPCR supports this finding. This miRNA was removed from the analysis as a later version of the alignment pipeline showed most sequences mapping to hsa-miR-4497 mapped to rRNA sequences (28S ribosomal N3 and 45S pre-ribosomal N2). The results of the RT-qPCR are included for transparency but do support the differential expression of this sequence. It is possible that the primers for hsa-miR-4497 bind to the rRNA sequences that were picked up in the

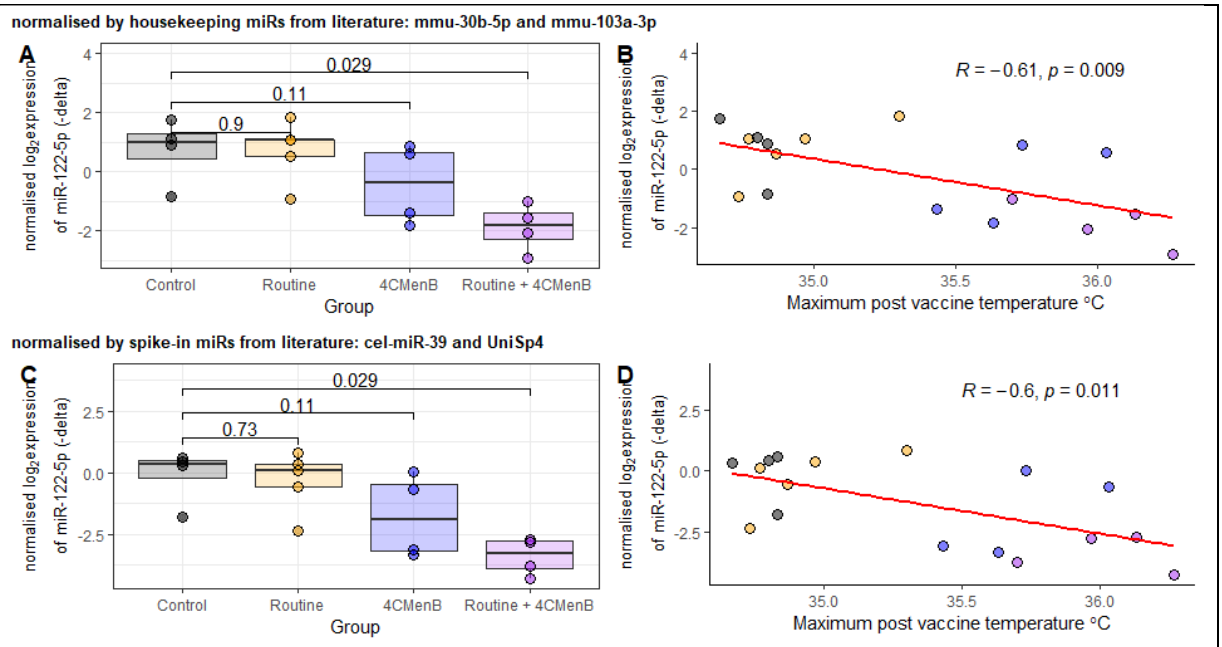
sRNA-seq as there is only one mismatch in base pairs between the canonical miR-4497 listed in miRbase and sequence contained in rRNA.



Appendix Figure 5 RT-qPCR of miR-122-5p in infant plasma samples per versus post-vaccination using LNA based primers. The LNA-based primers' result is similar to the Taqman advanced primer assays. This result acted as a bridging experiment for using LNA based primers for the mouse study PCRs.



*Appendix Figure 6 Relative proportion of each exRNA carrier at 4 and 12-months of ages. There was a greater relative estimated abundance of carrier 1 and carrier 2 at 4-months of ageand a corresponding reduction in carrier 3 compared with 12-monthts-old (*p<0.05. **p<0.001)*



Appendix Figure 7 Expression of miR-122-5p in mouse plasma 24 hours after vaccination using different normalisers.

Normalised expression plasma miR-122-5p expression, 24 hours after vaccination in each vaccine group using miRs suggested as a good mouse plasma housekeeping genes in the literature.*

*Scatterplot of the maximum temperature of a mouse in the 24 hours after vaccination versus the normalised expression of miR-122-5p using miRs suggested as a good mouse plasma housekeeping genes in the literature **.*

Normalised expression plasma miR-122-5p expression, 24 hours after vaccination in each vaccine group using spike-in miRs.*

*Scatterplot of the maximum temperature of a mouse 24 hours after vaccination versus the normalised expression of miR-122-5p using vaccine group using spike-in miRs **.*

** Significance values for comparisons tested using Wilcoxon rank-sum test*

*** Correlation was tested using the Pearson correlation co-efficient. Line of best-fit plot ted in red using the output of the linear model maximum temperature ~normalised miR-122-5p expression.*

Appendix Table 5 Results of MIEAA2 GSEA analysis using all miRNAs (ranked by their signed likelihood ratio based on their pre- versus post-vaccination expression.

Database	Cell Type	4-months of age analysis			12-month-old		Validated
		Enrichme	P	FDR	Enrichme	P value	
Cell-type	Hepatocy	depleted	<0.01	0.04	depleted	0.01	Yes
Cell-type	Muscle	depleted	<0.01	0.01	enriched	0.05	-
Immune	CD14	enriched	<0.01	<0.0	enriched	0.42	-
Immune	CD15	enriched	<0.01	<0.0	enriched	0.09	-
Immune	CD19	enriched	<0.01	<0.0	enriched	0.40	-
Immune	CD3	enriched	<0.01	<0.0	depleted	0.54	-
Immune	CD56	enriched	<0.01	<0.0	depleted	0.58	-
Organs	Trabecul	enriched	<0.01	0.01	enriched	0.22	-
exRNA	Ago2	enriched	<0.01	0.01	depleted	0.36	-
exRNA	HDL	enriched	0.02	0.04	enriched	0.28	-
exRNA	microvesi	enriched	<0.01	0.01	enriched	0.62	-

Pathway	**GO**	**GO Term**	**Enriche**	**P-**	**FDR**
coagulatio	GO:00301	positive regulation of blood	Depleted	0.000	0.029
immune	GO:00022	innate immune response in	Enriched	0.000	0.010
immune	GO:00022	astrocyte activation involved	Enriched	0.000	0.046
immune	GO:00083	signaling pattern recognition	Enriched	0.000	0.038
immune	GO:00108	regulation of tumor necrosis	Enriched	0.000	0.046
immune	GO:00303	osteoclast differentiation	Enriched	0.000	0.032
immune	GO:00303	high molecular weight B-cell	Enriched	0.000	0.032
immune	GO:00316	killing of cells of other	Enriched	0.000	0.046
immune	GO:00326	regulation of interferon-beta	Enriched	0.000	0.028
immune	GO:00327	negative regulation of	Enriched	0.000	0.046
immune	GO:00341	positive regulation of toll-	Enriched	0.000	0.043
immune	GO:00396	negative stranded viral RNA	Enriched	0.000	0.045
immune	GO:00450	positive regulation of	Enriched	0.000	0.044
immune	GO:00508	defense response to Gram-	Enriched	0.000	0.046
immune	GO:00618	antimicrobial humoral	Enriched	0.000	0.031
lipid	GO:00055	fatty acid binding	Enriched	0.000	0.032
lipid	GO:00066	phosphatidylserine	Enriched	0.000	0.045
lipid	GO:00086	lipid biosynthetic process	Enriched	0.000	0.029
lipid	GO:00108	negative regulation of	Enriched	0.000	0.020
lipid	GO:00108	regulation of cholesterol	Enriched	0.000	0.038
lipid	GO:00193	fatty acid oxidation	Enriched	0.000	0.043

lipid	GO:00335	fatty acid beta-oxidation	Enriched	0.000	0.047
lipid	GO:00361	alpha-linolenic acid	Enriched	0.000	0.027
lipid	GO:00550	lipid homeostasis	Enriched	0.000	0.043
lipid	GO:00550	phospholipid homeostasis	Enriched	0.000	0.010
lipid	GO:00606	regulation of cholesterol	Enriched	0.000	0.045
lipid	GO:00703	triglyceride homeostasis	Enriched	0.000	0.047
lipid	GO:19034	protein transport from ciliary	Enriched	0.000	0.040
lipid	GO:20003	regulation of ceramide	Enriched	0.000	0.046
lipid	GO:20006	negative regulation of	Enriched	0.000	0.032
lipid	GO:00476	acyl-CoA hydrolase activity	Enriched	0.000	0.043
protein	GO:00065	glutamine metabolic process	Enriched	0.000	0.010
protein	GO:00109	ubiquitin recycling	Enriched	0.000	0.046
protein	GO:00192	citrulline biosynthetic	Enriched	0.000	0.020
protein	GO:00197	ubiquitin-like protein	Enriched	0.000	0.046
protein	GO:00970	ubiquitin-protein transferase	Enriched	0.000	0.022
protein	GO:19901	protein linear	Enriched	0.000	0.018
gene	GO:00011	RNA polymerase I general	Enriched	0.000	0.046
gene	GO:00103	histone monoubiquitination	Enriched	0.000	0.043
gene	GO:00165	NuRD complex	Enriched	0.000	0.048
gene	GO:00352	mRNA cleavage involved in	Enriched	0.000	0.024
gene	GO:00705	RISC-loading complex	Enriched	0.000	0.046
gene	GO:00718	histone H2B conserved C-	Enriched	0.000	0.045
gene	GO:00906	mRNA cleavage involved in	Enriched	0.000	0.024
muscle	GO:00058	striated muscle thin filament	Enriched	0.000	0.040
muscle	GO:00426	muscle cell differentiation	Enriched	0.000	0.043
muscle	GO:00487	sarcomerogenesis	Enriched	0.000	0.046
muscle	GO:00550	cardiac muscle cell	Enriched	0.000	0.048
muscle	GO:19043	regulation of skeletal muscle	Enriched	0.000	0.045
muscle	GO:19043	negative regulation of	Enriched	0.000	0.045
muscle	GO:19051	positive regulation of	Enriched	0.000	0.046
nerve	GO:00328	neuronal cell body	Enriched	0.000	0.040
nerve	GO:00516	serotonin uptake	Enriched	0.000	0.029
nerve	GO:19029	positive regulation of	Enriched	0.000	0.045
nerve	GO:00053	serotoninsodium symporter	Enriched	0.000	0.029
nerve	GO:00074	axonal fasciculation	Enriched	0.000	0.029
vesicle	GO:00161	synaptic vesicle uncoating	Enriched	0.000	0.043
vesicle	GO:00300	vesicle transport along actin	Enriched	0.000	0.018
vesicle	GO:00723	clathrin coat disassembly	Enriched	0.000	0.029
cell/recept	GO:00019	regulation of receptor	Enriched	0.000	0.038
cell/recept	GO:00329	inositol trisphosphate	Enriched	0.000	0.046
cell/recept	GO:00350	positive regulation of Rho	Enriched	0.000	0.019
cell/recept	GO:00906	activation of GTPase activity	Enriched	0.000	0.020
cell/recept	GO:19027	GTPase activator complex	Enriched	0.000	0.040
ion	GO:00084	inositol-1,4,5-trisphosphate	Enriched	0.000	0.046
ion	GO:00153	potassiumchloride	Enriched	0.000	0.040
ion	GO:00550	chloride ion homeostasis	Enriched	0.000	0.031
ion	GO:00550	potassium ion homeostasis	Enriched	0.000	0.029

ion	GO:00902	positive regulation of	Enriched	0.000	0.045
kidney	GO:00702	renal water absorption	Enriched	0.000	0.046
kidney	GO:00722	metanephric glomerular	Enriched	0.000	0.046
kidney	GO:20005	positive regulation of	Enriched	0.000	0.046
heart	GO:00905	endocardial cushion to	Enriched	0.000	0.029
heart	GO:19018	regulation of cell	Enriched	0.000	0.043
other	GO:00001	microfilament motor activity	Enriched	0.000	0.045
other	GO:00019	uropod	Enriched	0.000	0.043
other	GO:00038	CDP-diacylglycerol-inositol	Enriched	0.000	0.038
other	GO:00039	GPI-anchor transamidase	Enriched	0.000	0.045
other	GO:00040	carbamoyl-phosphate	Enriched	0.000	0.018
other	GO:00043	glycerol-3-phosphate O-	Enriched	0.000	0.038
other	GO:00046	transmembrane receptor	Enriched	0.000	0.045
other	GO:00053	nucleoside transmembrane	Enriched	0.000	0.038
other	GO:00057	vacuole	Enriched	0.000	0.046
other	GO:00057	vacuolar membrane	Enriched	0.000	0.019
other	GO:00080	DNA-dependent ATPase	Enriched	0.000	0.043
other	GO:00081	prostaglandin-I synthase	Enriched	0.000	0.038
other	GO:00083	single-stranded DNA 3-5	Enriched	0.000	0.040
other	GO:00085	monoamine transmembrane	Enriched	0.000	0.029
other	GO:00163	activin receptor activity,	Enriched	0.000	0.029
other	GO:00192	kinase regulator activity	Enriched	0.000	0.045
other	GO:00304	formimidoyltransferase	Depleted	0.000	0.046
other	GO:00304	formimidoyltetrahydrofolate	Depleted	0.000	0.046
other	GO:00306	protein binding, bridging	Enriched	0.000	0.032
other	GO:00312	replication fork protection	Enriched	0.000	0.048
other	GO:00316	nuclear pericentric	Enriched	0.000	0.045
other	GO:00316	G-protein gamma-subunit	Enriched	0.000	0.040
other	GO:00331	CAF-1 complex	Enriched	0.000	0.029
other	GO:00340	pha	Enriched	0.000	0.034
other	GO:00347	inhibin binding	Enriched	0.000	0.046
other	GO:00380	collagen receptor activity	Enriched	0.000	0.032
other	GO:00427	GPI-anchor transamidase	Enriched	0.000	0.045
other	GO:00447	G1 DNA damage checkpoint	Enriched	0.000	0.045
other	GO:00473	acylglycerol lipase activity	Enriched	0.000	0.046
other	GO:00707	BMP receptor complex	Enriched	0.000	0.046
other	GO:00973	7SK snRNA binding	Enriched	0.000	0.038
other	GO:00977	positive regulation of blood	Enriched	0.000	0.043
other	GO:00000	urea cycle	Enriched	0.000	0.038
other	GO:00020	regulation of oxidative	Enriched	0.000	0.044
other	GO:00062	de novo pyrimidine	Enriched	0.000	0.046
other	GO:00066	protein retention in ER	Enriched	0.000	0.046
other	GO:00068	nitrogen compound	Enriched	0.000	0.018
other	GO:00068	cell volume homeostasis	Enriched	0.000	0.029
other	GO:00070	vacuolar acidification	Enriched	0.000	0.043
other	GO:00082	spermidine biosynthetic	Enriched	0.000	0.046
other	GO:00096	response to toxic substance	Enriched	0.000	0.045

other	GO:00158	monoamine transport	Enriched	0.000	0.029
other	GO:00193	cyclooxygenase pathway	Enriched	0.000	0.046
other	GO:00315	cell-substrate adhesion	Enriched	0.000	0.024
other	GO:00329	positive regulation of	Enriched	0.000	0.045
other	GO:00332	regulation of nuclear cell	Enriched	0.000	0.047
other	GO:00343	NAD biosynthesis via	Enriched	0.000	0.038
other	GO:00346	cellular macromolecule	Enriched	0.000	0.045
other	GO:00434	steroid hormone mediated	Enriched	0.000	0.045
other	GO:00483	retinoic acid receptor	Enriched	0.000	0.045
other	GO:00510	positive regulation of	Enriched	0.000	0.044
other	GO:00517	response to redox state	Enriched	0.000	0.032
other	GO:00519	regulation of sensory	Enriched	0.000	0.045
other	GO:00714	cellular response to fluid	Enriched	0.000	0.029
other	GO:00717	organic substance transport	Enriched	0.000	0.038
other	GO:00900	regulation of thalamus size	Enriched	0.000	0.029
other	GO:19041	protein localization to	Enriched	0.000	0.038
other	GO:19042	positive regulation of CDP-	Enriched	0.000	0.045
other	GO:19042	positive regulation of serine	Enriched	0.000	0.045
other	GO:19049	brush border assembly	Enriched	0.000	0.038
other	GO:19905	bone regeneration	Enriched	0.000	0.045
other	GO:20003	negative regulation of	Enriched	0.000	0.032

Appendix 1. QC plots EUCLIDS sRNA-seq

Appendix 1.1 EUCLIDS 4-month-old cohort

9.2.3 Euclids 4-month cohort Library size statistics

Average raw library size (accepted reads pre-alignment)

median = 5.5 million reads (range: 3.2—7.4 million reads)

miRNA library size:

Pre filtering : median = 0.29 million reads (range: 0.11—1.1 million reads)

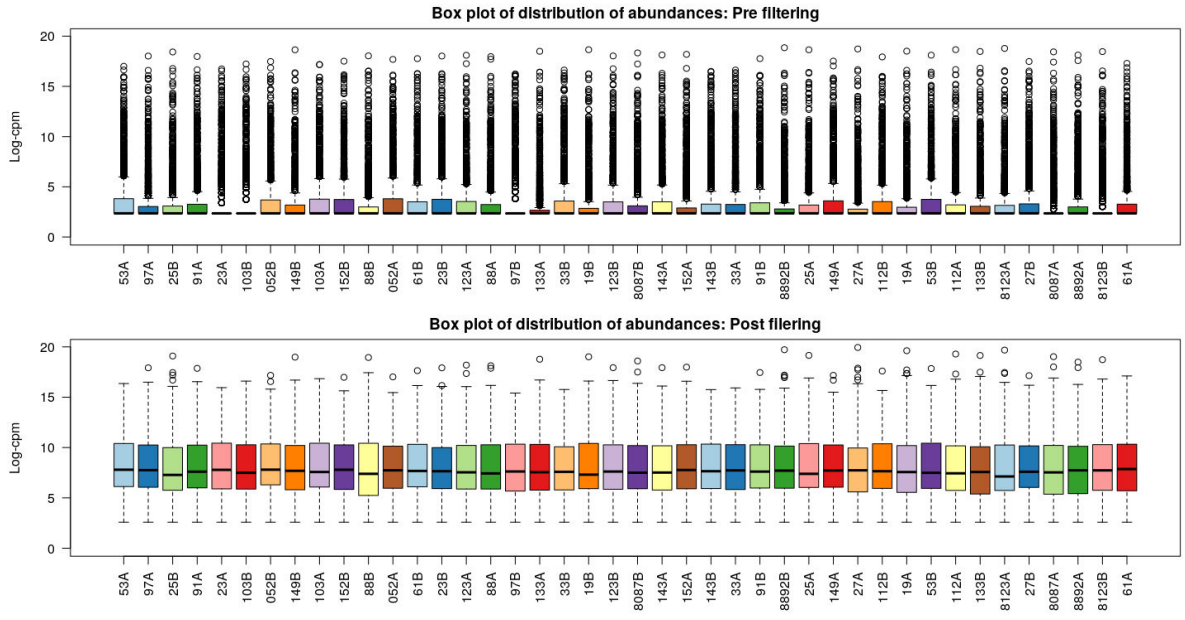
Post filtering: median = 0.29 million (range: 0.11 million—1.1 million)

sRNA library size:

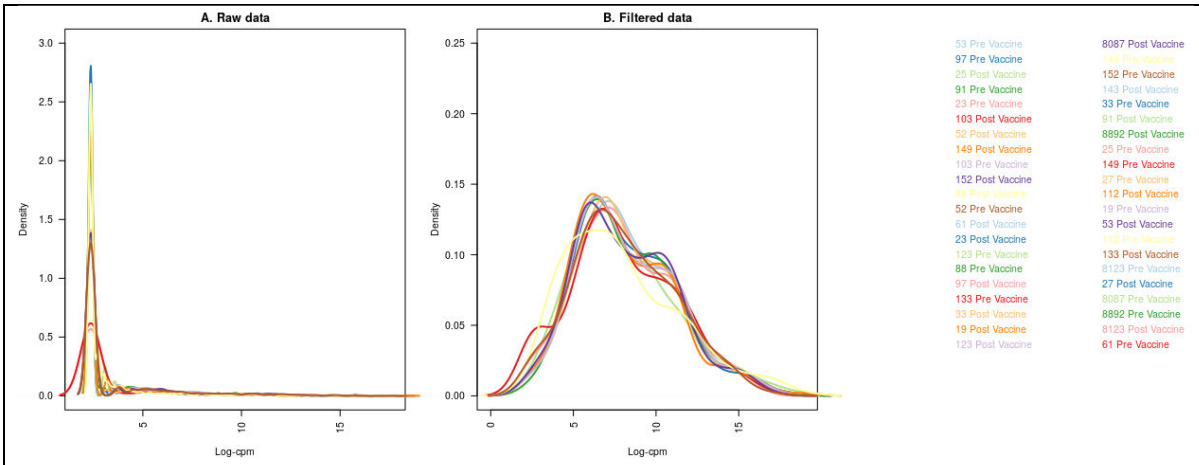
Pre filtering : median = 3.2 million reads (range: 2.1—5 million reads)

Post filtering: media = 2.8 million (range: 1.9 million—4 million)

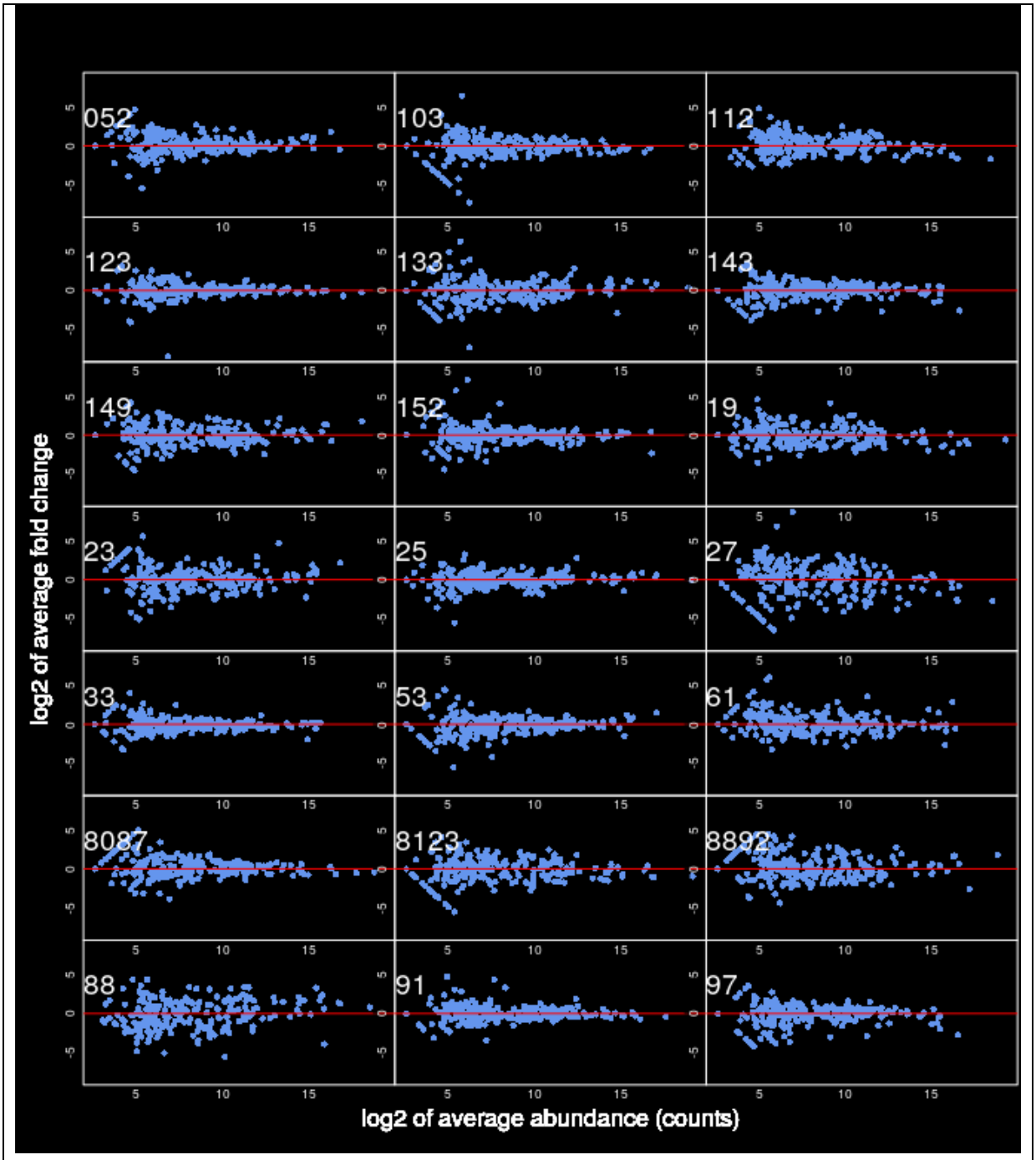
9.2.4 EUCLIDS 4-month-old cohort—miRNA reads



Appendix figure 8 EUCLIDS miRNA sRNA seq Box plot log CPM reads before (top) and after filtering out noise and application of TMM normalisation (bottom). The aim is for median values and the IQR to be similar across samples after noise removal and normalisation. This is the case, suggesting that the noise removal and normalisation worked well and that there are no outliers.



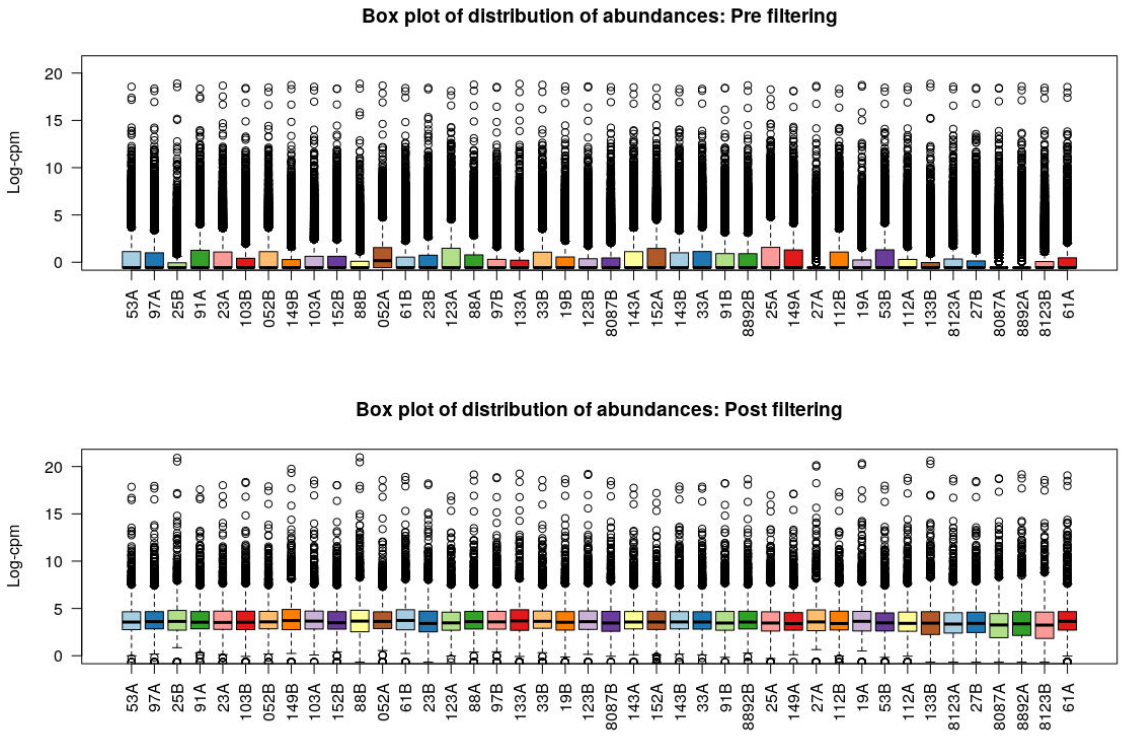
Appendix figure 9 EUCLIDS miRNA sRNA seq Density Distribution of log CPM reads before (A) and after filtering out noise, and application of TMM normalisation (B). The large peak at low count level in (A) is absent in B suggesting that the noise removal and normalisaiton has worked well and there are no outliers.



Appendix figure 10 M-A plots of pre versus post-vaccination miRNA expression for each participant. Each dot represents the log_2 fold change of an sRNA feature vs the average expression of the sRNA feature in the two samples. If the removal of noise and the normalisation process is adequate, then median sRNA feature fold changes should be 0 for each participant - the red horizontal line (log_2 fold change = 0) should bisect the cloud of dots in half (i.e., there should be approximately an equal number of dots above

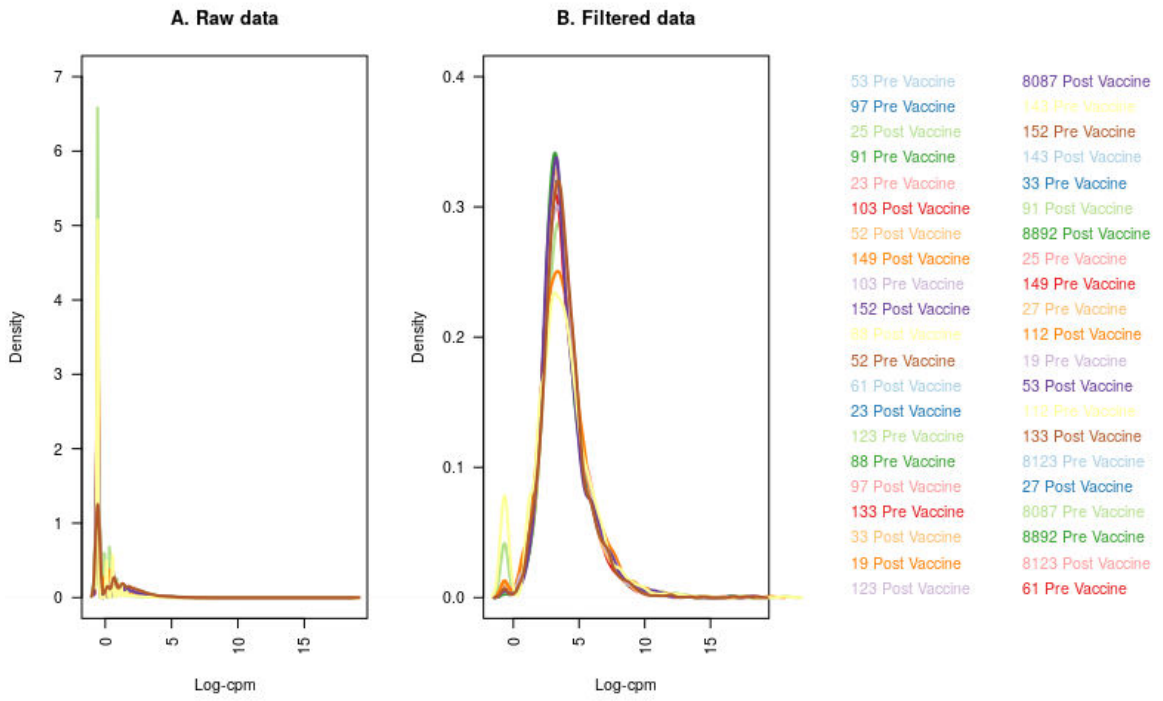
the red line as below it). This is the case, suggesting that the noise removal and normalisation worked well and that there are no outliers.

EUCLIDS 4-month-old cohort—sRNA reads

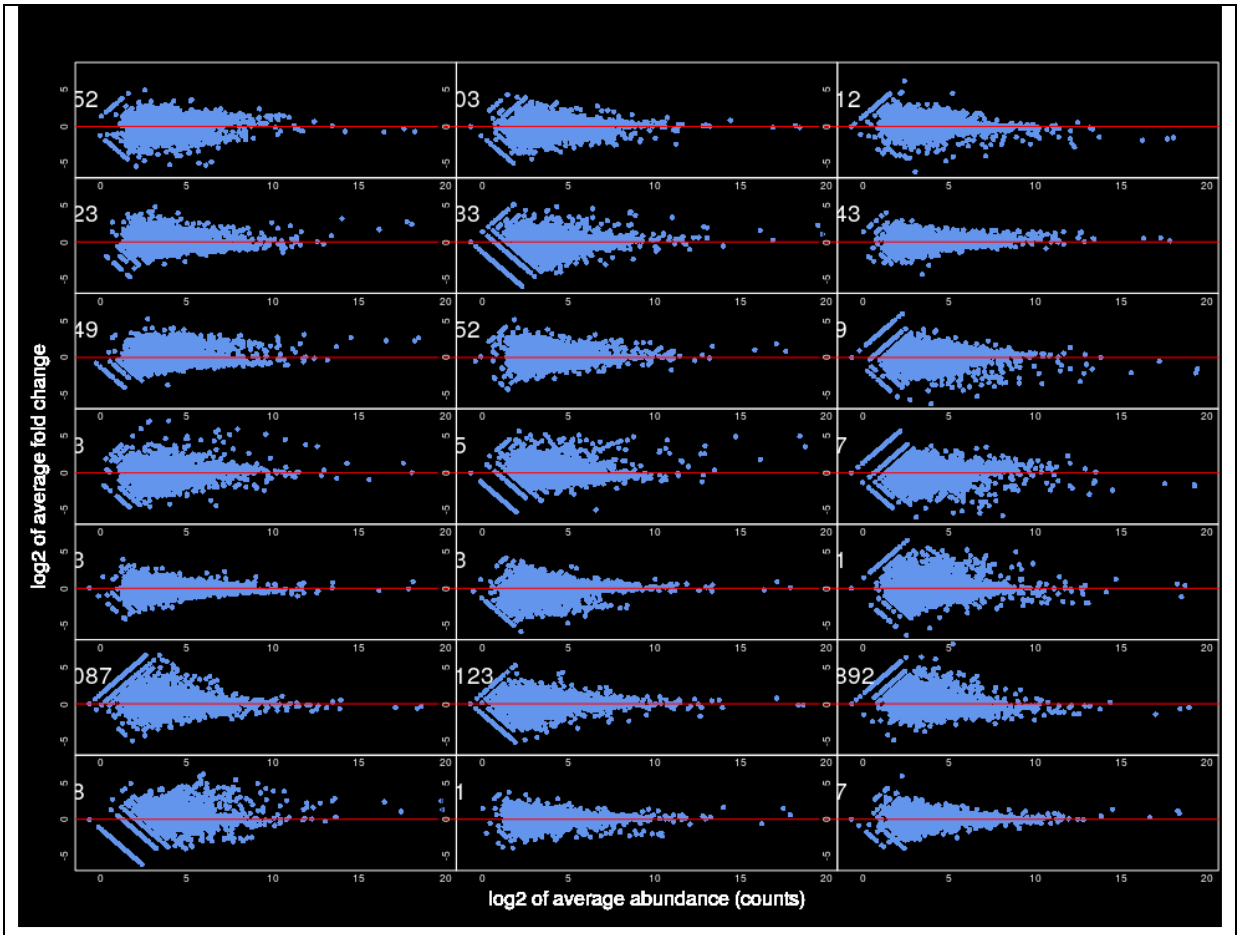


Appendix figure 11 EUCLIDS sRNAs: Each box plot shows the distribution of sRNA feature counts within each sample.

Box plot of log2 CPM before (top) and after filtering out noise and application of TMM normalisation (bottom). The aim is for median values and the IQR to be similar across samples after noise removal and normalisation. This is the case, suggesting that the noise removal and normalisation has worked well and there are no outliers.



Appendix figure 12 EUCLIDS miRNA sRNA seq Density Distribution of log CPM reads before (A) and after filtering out noise and TMM normalisation (B). The prominent peak at low count level in (A) is absent in B, suggesting that the noise removal and normalisation have worked well and there are no outliers.

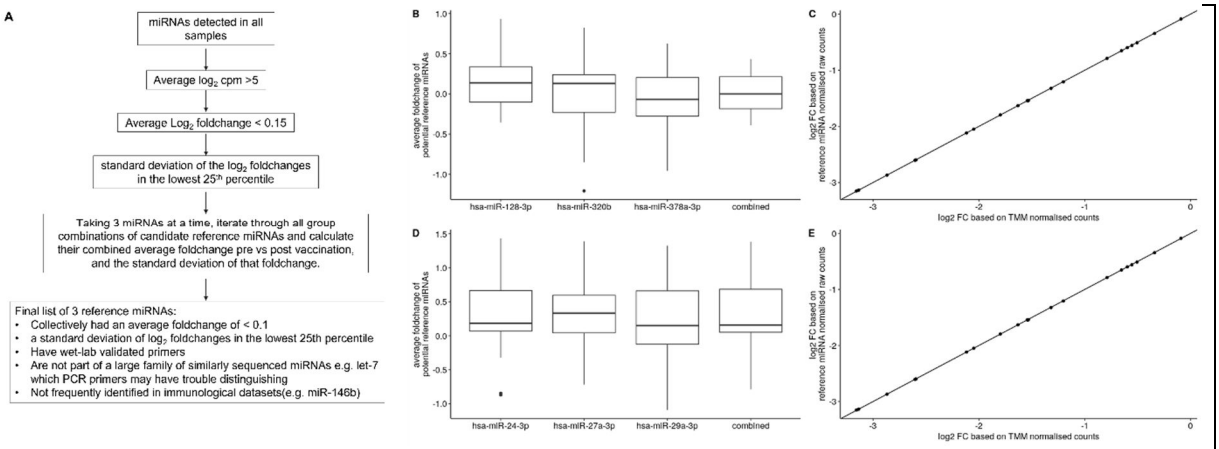


Appendix figure 13 M-A plots of pre versus post-vaccination sRNA feature expression for each participant. Each dot represents the log_2 fold change of an sRNA feature vs the average expression of the sRNA feature in the two samples. If the removal of noise and the normalisation process is adequate, then median sRNA feature fold changes should be 0 for each participant - the red horizontal line (log_2 fold change = 0) should bisect the cloud of dots in half (i.e., there should be approximately an equal number of dots above the red line as below it). This is the case, suggesting that the noise removal and normalisation worked well and there are no outliers.

9.2.5 Selection of miRNA for RT-qPCR normalisation

The NormFinder and BestKeeper algorithm did not work well for this data. Both algorithms selected miRNAs that tended towards differential expression [286,288]. Therefore, an algorithm was created, which selected miRNAs that were ubiquitously and abundantly present across samples and were stably, non-DE pre versus post-vaccination (Appendix Figure 14 A and methods section "Selection of appropriate endogenous control for RT-qPCR data normalisation"). Three candidate reference miRNAs were chosen: hsa-miR-320b, hsa-miR-128 and hsa-miR-378a (reference group 1, Appendix Figure 14B). The raw count data of DE miRNAs were re-analysed using the raw expression of these three miRNAs using a delta-delta approach analogous to RT-PCR data analysis. (i.e., $\log_2$ expression of DE miRNA—average $\log_2$ expression of all three reference miRNAs). Reassuringly, this recapitulated the results of the standard RNA-seq differential expression analysis (Appendix Figure 14C).

It is unclear whether miRNAs are globally downregulated or merely outcompeted by increased levels of tRNAs post-vaccination; therefore and additional three miRNAs were selected as potential normalisers from the set of miRNAs that were stably, non-DE pre versus post-vaccination when all sRNAs were analysed collectively (reference group 2, Appendix Figure 14 D:E). Comparing the stability in the expression of these two sets of miRNA normalisers could help delineate whether miRNAs are globally downregulated e.g., if reference group 2 are more stably expressed, then potentially miRNAs are globally downregulated and vice-versa.



Appendix Figure 14 Selecting candidate reference miRNAs for RT-PCR validation of sRNA-seq data

A. *Overview of selection algorithm for identification of candidate reference miRNA*
B. *Distribution of log_2 fold changes for candidate reference miRNAs using TMM normalised count data when restricting library size to miRNAs (reference group 1)*
C. *Log_2 fold changes of hsa-miR-122-5p (top DE miRNA per vs post-vaccination) based on TMM normalised count data versus log_2 fold changes using a delta-delta approach analogous to RT-PCR data analysis. (i.e., log_2 expression of hsa-miRNA-122-5p—average log_2 expression of hsa-miR128-3p, hsa-miR-320b, and hsa-miR-378a miRNAs)*
D. *Distribution of log_2 fold changes for candidate reference miRNAs using TMM normalised count data when all sRNAs are used (reference group 2)*
E. *Log_2 fold changes of hsa-miR-122-5p (top DE miRNA per vs post-vaccination) based on TMM normalised count data versus log_2 fold changes using a delta-delta approach analogous to RT-PCR data analysis. (i.e., log_2 expression of hsa-miRNA-122-5p—average log_2 expression of hsa-miR-24-3p, hsa-miR-27a-3p, and hsa-miR-29a-3p miRNAs)*

9.2.6 EUCLIDS library size statistics

Filtering threshold: 15 counts in the smallest library converted to a minimum count per million. A feature had to be present at or above this threshold in at least 25% of samples.

Average raw library size (accepted reads pre-alignment)

median = 4.1 million reads (range: 2.4—6 million reads)

miRNA library size:

Pre filtering : median = 0.11 million reads (range: 0.21—0.61 million reads)

Post filtering: median = 0.09 million (range: 0.17 million—0.51 million)

sRNA library size:

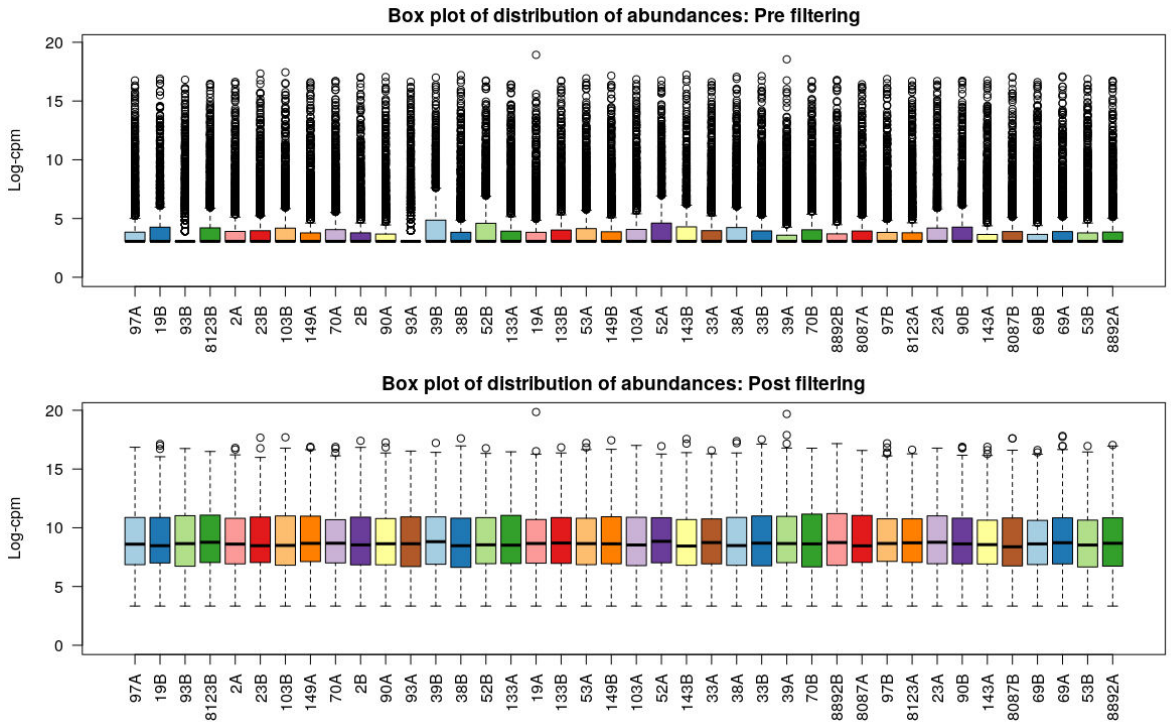
Pre filtering : median = 1.2 million reads (range: 1.2—4.5 million reads)

Post filtering: media = 0.8 million (range: 1.4 million—4.2 million)

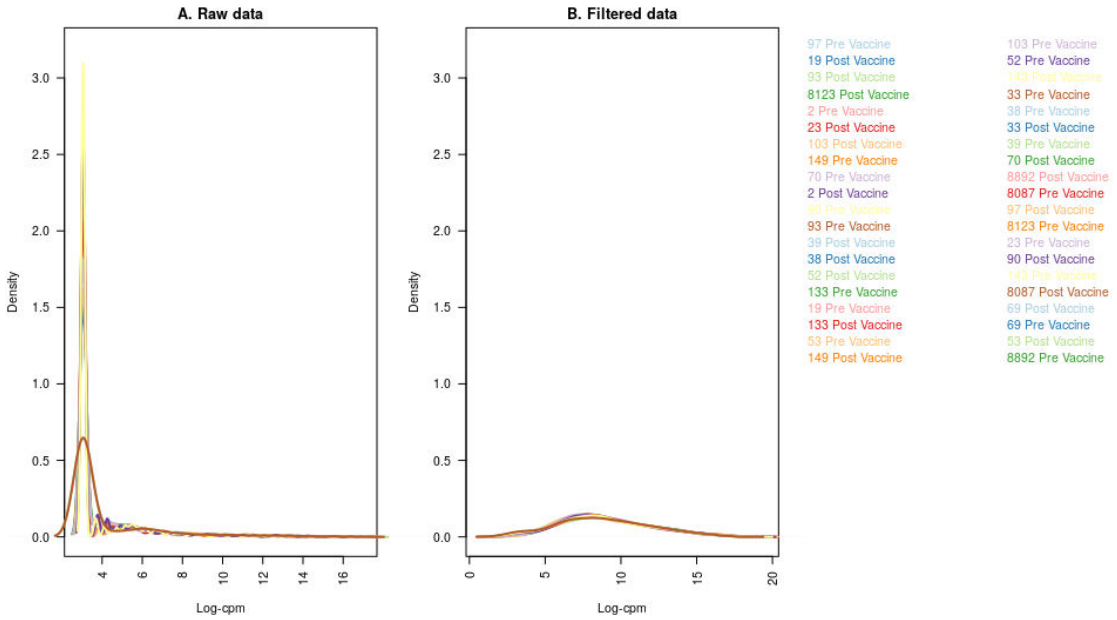
50,361 features were initially detected of which 2,796 (5.6%) passed QC filtering criteria and were retained in the final analyses.

9.2.7 EUCLIDS 12-month-old cohort—miRNA reads

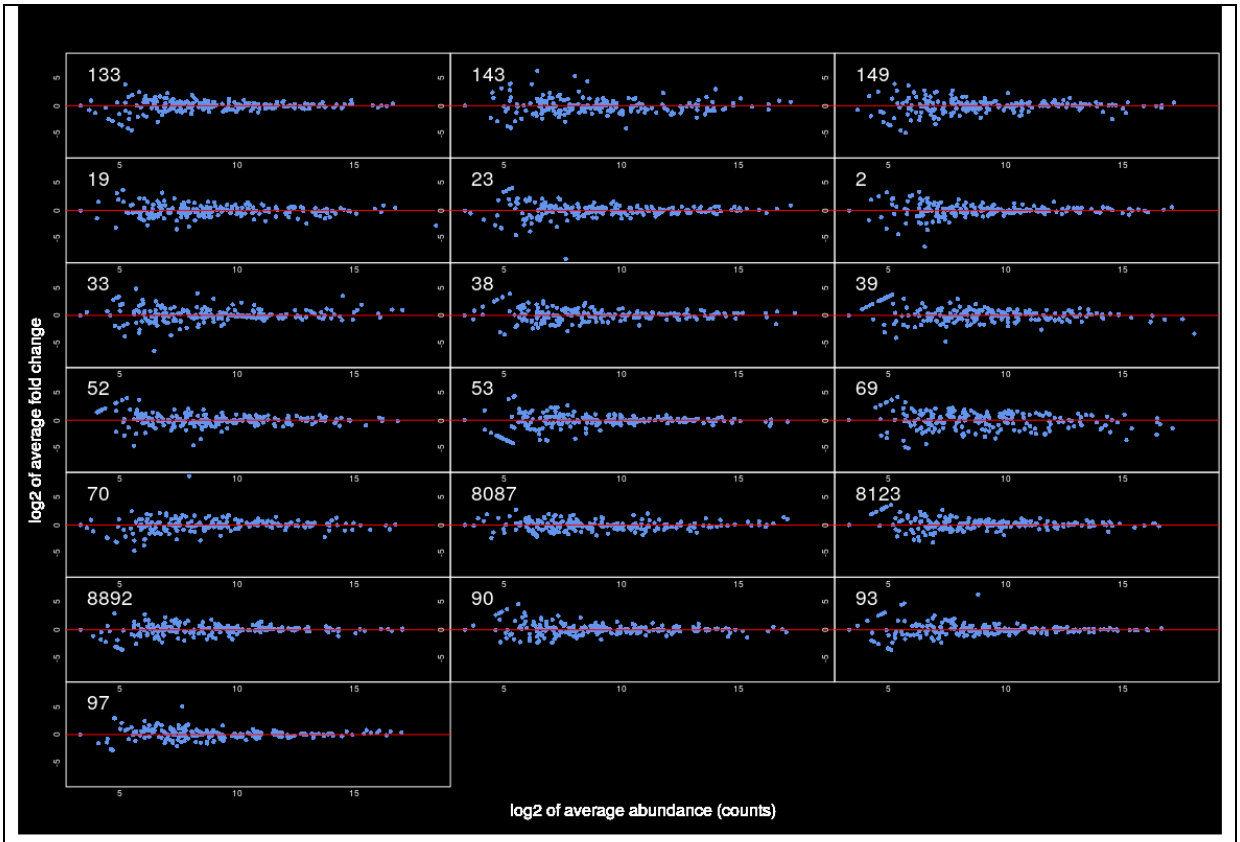
Filtering threshold: 15 counts in the smallest library converted to a minimum count per million. A feature had to be present at or above this threshold in at least 25% of samples.



Appendix figure 15 EUCLIDS miRNA sRNA seq Box plot log CPM reads before (top) and after filtering out noise and application of TMM normalisation (bottom). The aim is for median values and the IQR to be similar across samples after noise removal and normalisation. This is the case, suggesting that the noise removal and normalisation have worked well and there are no outliers.

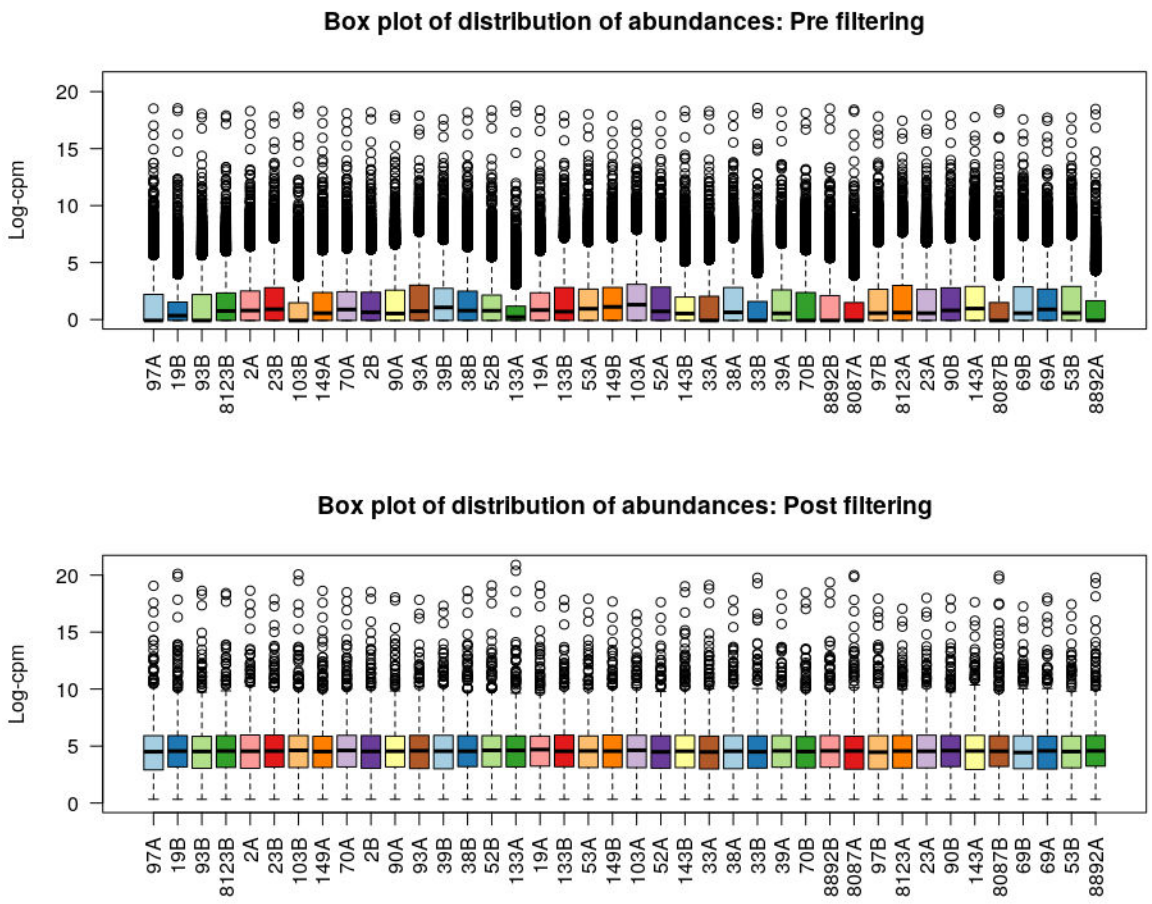


Appendix figure 16 EUCLIDS miRNA sRNA seq Density Distribution of log CPM reads before (A) and after filtering out the noise and TMM normalisation (B). The prominent peak at the low count level in (A) is absent in B, suggesting that the noise removal and normalisation have worked well and there are no outliers



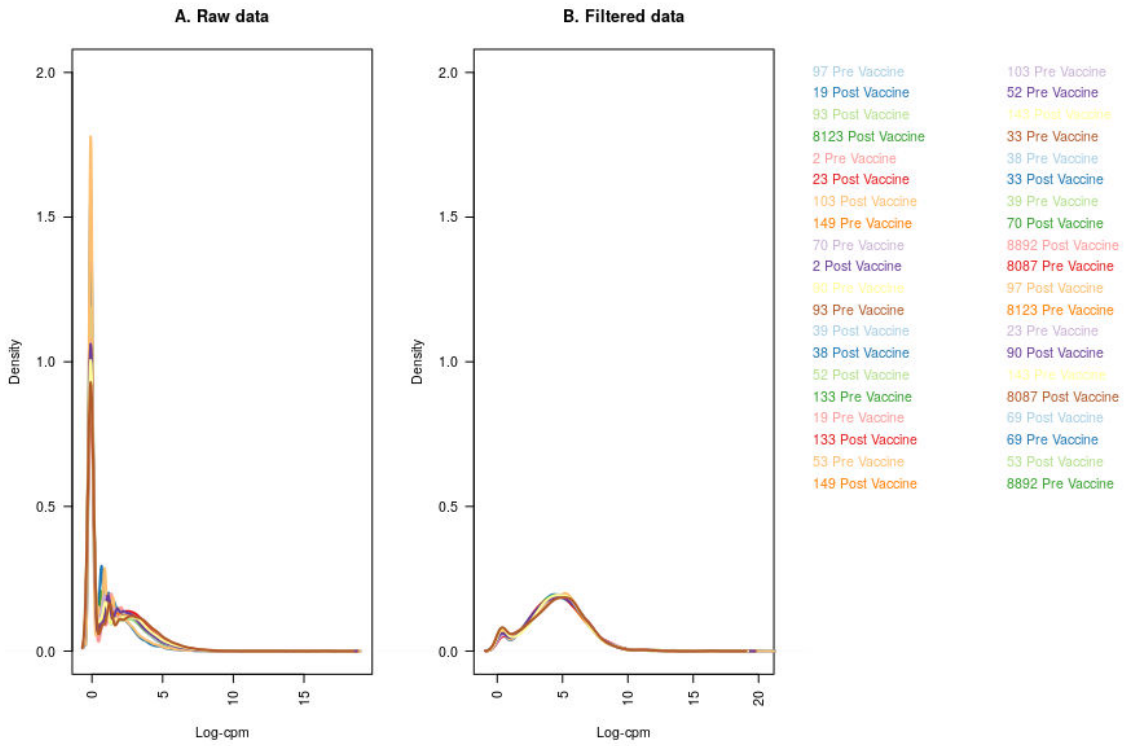
Appendix figure 17 M-A plots of pre versus post-vaccination miRNA expression for each participant. Each dot represents the log_2 fold change of an sRNA feature vs the average expression of the sRNA feature in the two samples. If the removal of noise and the normalisation process is adequate, then median sRNA feature fold changes should be 0 for each participant - the red horizontal line (log_2 fold change = 0) should bisect the cloud of dots in half (i.e., there should be approximately an equal number of dots above the red line as below it). This is the case, suggesting that the noise removal and normalisation worked well and that there are no outliers.

9.2.8 EUCLIDS 12-month-old cohort—sRNA reads

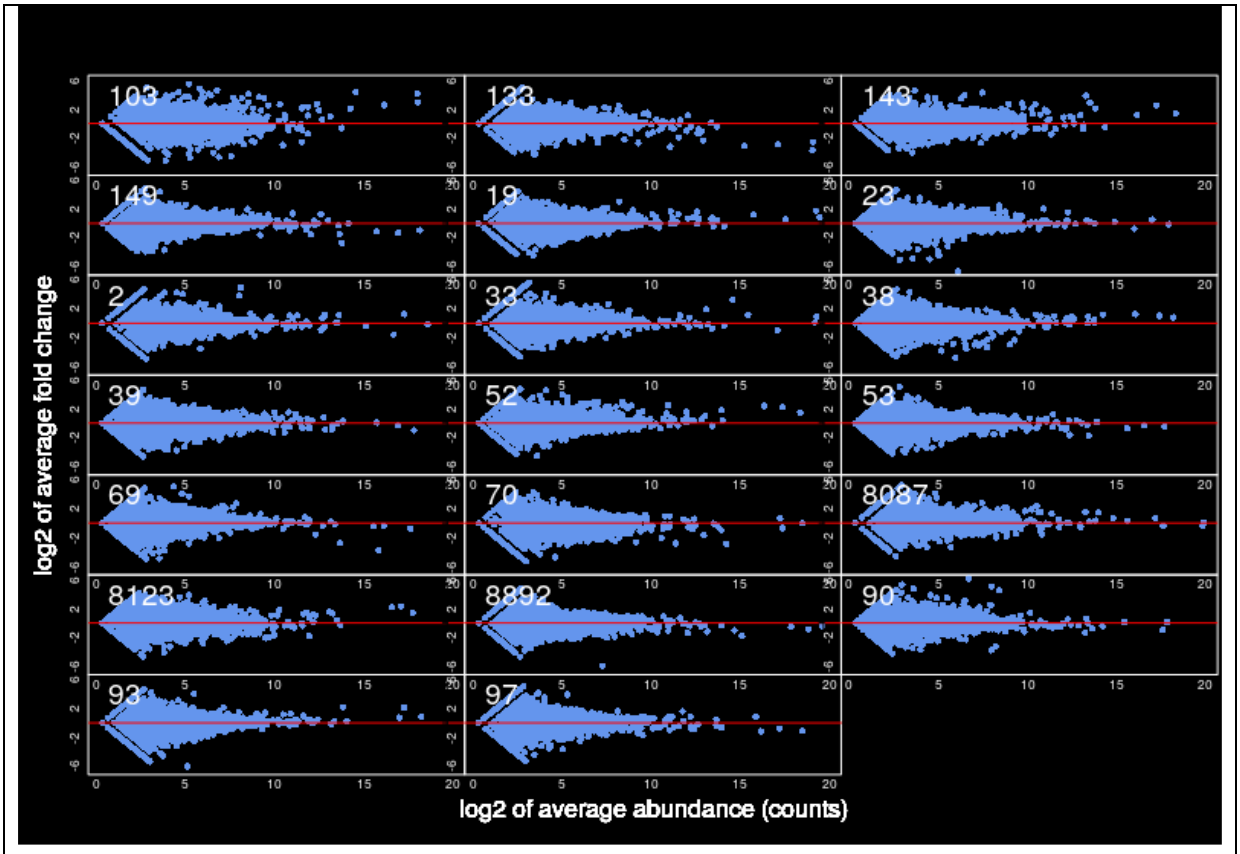


Appendix figure 18 EUCLIDS sRNAs: Each box plot shows the distribution of sRNA feature counts within each sample.

Box plot of log_2 CPM before (top) and after filtering out noise and application of TMM normalisation (bottom). The aim is for median values and the IQR to be similar across samples after noise removal and normalisation. This is the case, suggesting that the noise removal and normalisation have worked well and there are no outliers.



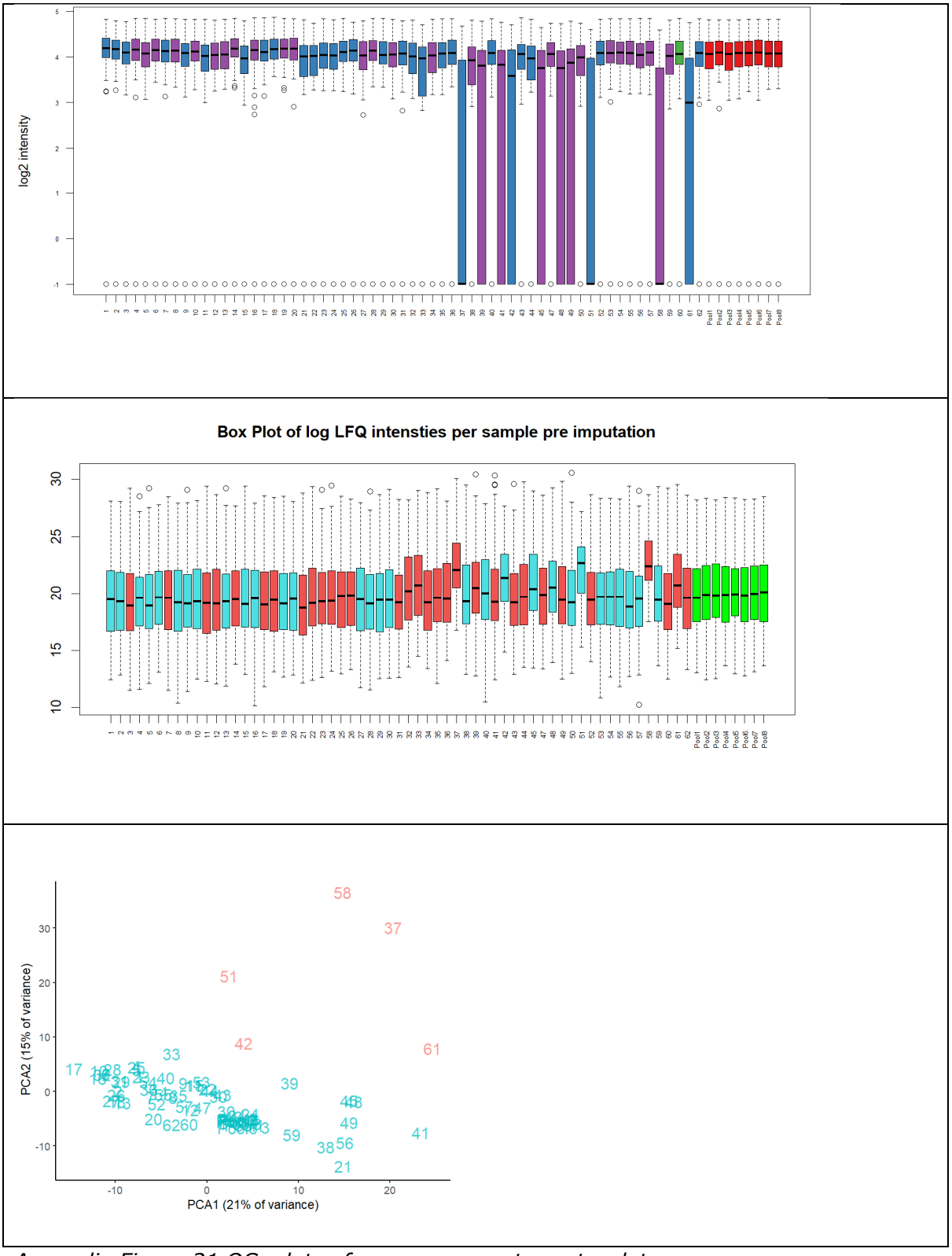
Appendix figure 19EUCLIDS miRNA sRNA seq Density Distribution of log CPM reads before (A) and after filtering out noise and TMM normalisation (B). The prominent peak at the low count level in (A) is absent in B, suggesting that the noise removal and normalisation have worked well and there are no outliers



Appendix figure 20 M-A plots of pre versus post-vaccination sRNA feature expression for each participant. Each dot represents the log_2 fold change of an sRNA feature vs the average expression of the sRNA feature in the two samples. If the removal of noise and the normalisation process is adequate, then median sRNA feature fold changes should be 0 for each participant - the red horizontal line (log_2 fold change = 0) should bisect the cloud of dots in half (i.e., there should be approximately an equal number of dots above the red line as below it). This is the case, suggesting that the noise removal and normalisation has worked well and there are no outliers.

9.3 Appendix for Chapter 4

9.3.1 Mass spectrometry Data QC

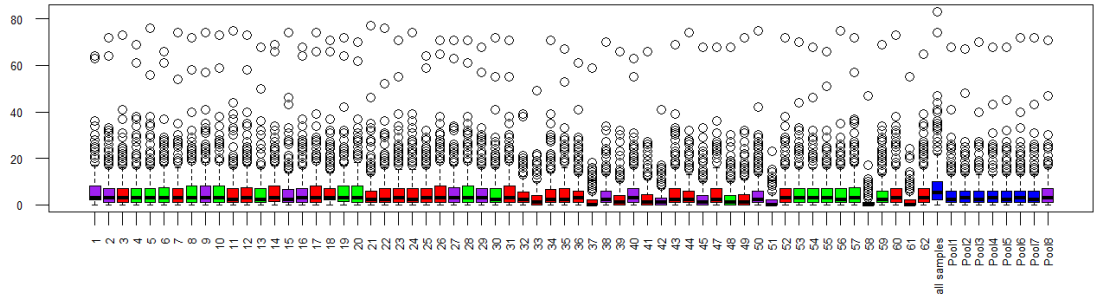


Appendix Figure 21 QC plots of raw mass spectrometry data

Panel A: Box plot of raw log_2 Intensities—at least 25% of proteins were missing in 8 samples (IQR extends to -1).

Panel B: Box plot of LFQ intensities. Samples that were missing many proteins have higher LFQ intensities due to over-compensation of the normalisation process in the LFQ algorithm.

Panel C: PCA of samples based on their LFQ intensity data before imputation is added. The 5 samples selected for removal based on plots A, B and C are highlighted in red.



Appendix Figure 22 box plot shows the distribution of the number of unique peptides per protein within each sample. The median number of peptides per protein across was 3.

9.3.2 Cytokines/chemokines and acute-phase proteins analysed by Luminex assays

Supplementary Table 2 Cytokines/chemokines and acute-phase proteins analysed by Luminex assays

Luminex Panel	Cytokine	Short name	Measured at 4 months	Measured at 12 months
HCVD3MAG-67K-03	C-reactive protein	CRP	Yes	Yes
	Haptoglobin	Haptoglobin	Yes	Yes
	L-selectin	L-selectin	Yes	Yes
HCYTOMAG-60K-24C	Epidermal growth factor	EGR	Yes	No
	Fraktelkine	FKN	Yes	No
	Granulocyte colony-stimulating factor	G-CSF	Yes	Yes
	Granulocyte-macrophage	GM-CSF	Yes	No

	colony-stimulating factor			
	GRO-alpha (CXCL1)	GRO-alpha	Yes	Yes
	Interferon-gamma	IFNy	Yes	No
	Interleukin-1 receptor antagonist	IL1RA	Yes	Yes
	Interleukin-1 alpha	IL1-A	Yes	Yes
	Interleukin-1 beta	IL1-B	Yes	Yes
	Interleukin-17 alpha	IL17-A	Yes	Yes
	Interleukin-2	IL-2	Yes	Yes
	Interleukin-3	IL-3	Yes	Yes
	Interleukin-4	IL-4	Yes	Yes
	Interleukin-5	IL-5	Yes	Yes
	Interleukin-6	IL-6	Yes	Yes
	Interleukin-7	IL-7	Yes	No
	Interleukin-8	IL-8	Yes	Yes
	Interleukin-10	IL10	Yes	Yes
	Interleukin-13	IL13	Yes	No
	Interferon gamma-induced protein 10 (CXLC10)	IP10	Yes	Yes
	Soluble CD 40 ligand	sCD40L	Yes	Yes
	Transforming growth factor-alpha	TGF-alpha	Yes	Yes
	Tumour necrosis factor-alpha	TNF-alpha	Yes	Yes

Supplementary Table 3 DE proteins measured by LC/MS in the 4-month-old EUCLIDS cohort. Proteins that are upregulated are in red. Down-regulated proteins are in blue

DE proteins measured by LC/MS in the 4-month-old EUCLIDS cohort. Proteins that are upregulated are in red. Down-regulated proteins are in blue

	Protein	*Log_2 FC*	*Average Expression*	*p-value*	*FDR*
Both groups combined Baseline vs 24 hours post-vaccination	C-reactive protein	4.61	16.10	8.90E-	1.56E-
	Serum amyloid A-2	8.13	14.08	1.44E-	1.26E-
	Serum amyloid A-1	7.49	18.15	1.38E-	8.07E-
	Leucine-rich alpha-2-	1.48	22.94	2.90E-	1.27E-
	Lactotransferrin	3.22	15.49	6.66E-	2.33E-
	Lipopolysaccharide-	1.41	13.30	6.03E-	1.76E-
	Complement component	1.42	19.58	4.03E-	1.01E-
	Protein Z-dependent	1.03	16.09	3.53E-	7.71E-
	Alpha-1-antichymotrypsin	0.81	23.07	7.50E-	1.39E-
	Alpha-1-acid glycoprotein	2.34	20.12	7.92E-	1.39E-
	Apolipoprotein C-III	-0.88	23.81	4.62E-	7.35E-
	Retinol-binding protein 4	-0.58	23.56	5.29E-	7.71E-

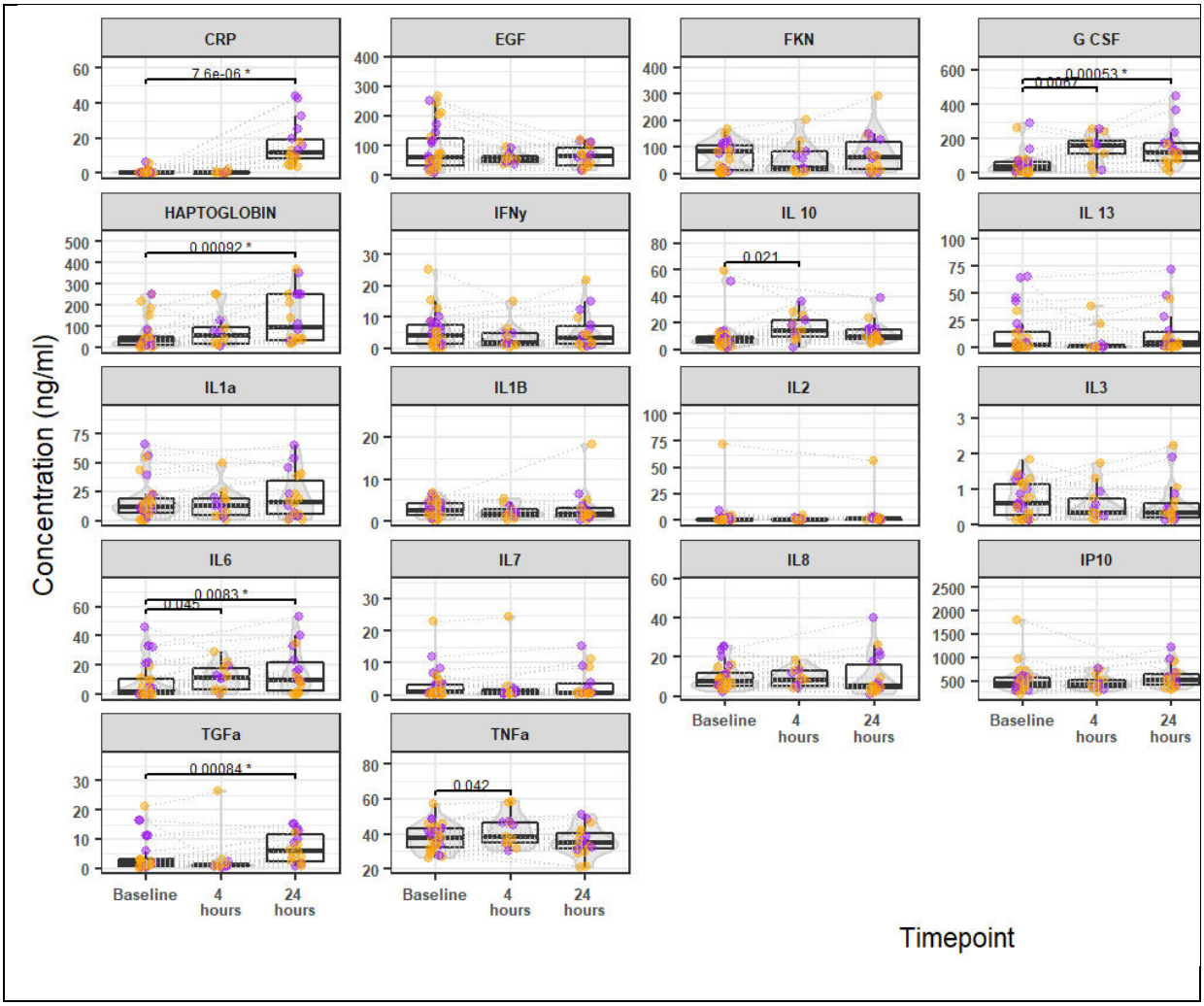
	Complement factor B	0.45	23.52	1.04E-	1.41E-
	Gelsolin	-0.35	23.30	2.37E-	2.80E-
	Apolipoprotein E	-0.44	23.88	2.40E-	2.80E-
	Serum amyloid A-4	0.80	22.31	3.26E-	3.57E-
	Alpha-1-acid glycoprotein	1.71	19.07	3.74E-	3.85E-
	Prostaglandin-H2 D-	-0.57	18.30	4.81E-	4.64E-
	Serum amyloid P-	0.72	20.83	5.04E-	4.64E-
	Haptoglobin	1.70	20.29	5.37E-	4.70E-
	Vitamin K-dependent	-0.48	17.52	5.87E-	4.88E-
	Alpha-1-antitrypsin_Short	0.57	27.07	6.14E-	4.88E-
4CMenB	C-reactive protein	5.80	16.10	5.73E-	1.00E-
	Serum amyloid A-2	10.09	14.08	5.10E-	4.46E-
Group	Serum amyloid A-1	8.91	18.15	1.30E-	7.56E-
	Leucine-rich alpha-2-	1.69	22.94	1.00E-	4.39E-
Baseline vs	Lactotransferrin	3.69	15.49	1.24E-	4.33E-
	Lipopolysaccharide-	1.62	13.30	9.21E-	2.69E-
24 hours	Complement component	1.58	19.58	4.20E-	0.001
	Alpha-1-antichymotrypsin	0.96	23.07	1.41E-	0.003
post-	Protein Z-dependent	1.18	16.09	2.28E-	0.004
	Vitamin K-dependent	-0.73	17.52	1.18E-	0.019
vaccination	Gelsolin	-0.48	23.30	1.25E-	0.019
	Alpha-1-acid glycoprotein	2.34	20.12	1.31E-	0.019
	Apolipoprotein C-III	-0.99	23.81	1.55E-	0.021
	Serum amyloid A-4	1.00	22.31	4.13E-	0.049
	Fibrinogen gamma chain	2.13	17.59	4.23E-	0.049
Routine	C-reactive protein	3.40	16.10	1.46E-	2.55E-
	Serum amyloid A-2	6.11	14.08	3.92E-	3.43E-
Group	Serum amyloid A-1	6.08	18.15	1.89E-	1.10E-
	Leucine-rich alpha-2-	1.25	22.94	1.66E-	7.24E-
Baseline vs	Lactotransferrin	2.68	15.49	5.38E-	0.002
	Lipopolysaccharide-	1.19	13.30	1.69E-	0.005
24 hours	Retinol-binding protein 4	-0.70	23.56	8.84E-	0.022
post-	Complement component C9	1.20	19.58	1.46E-03	0.032
vaccination					
4CMenB vs	Apolipoprotein C-I	-1.62	23.83	1.09E-	0.018
	Apolipoprotein A-I	-1.21	28.71	2.01E-	0.018
Routine at 4	Apolipoprotein A-II	-1.22	24.76	7.82E-	0.046
4CMenB vs	C-reactive protein	-2.40	16.10	6.49E-	1.14E-
Routine at	Serum amyloid A-2 protein	-3.98	14.08	4.53E-05	0.004
24 hours					
post-					
vaccination					

Supplementary Table 4 DE proteins measured by LC/MS in the 12-month-old EUCLIDS cohort. Proteins that are upregulated are in red. Down-regulated proteins are in blue

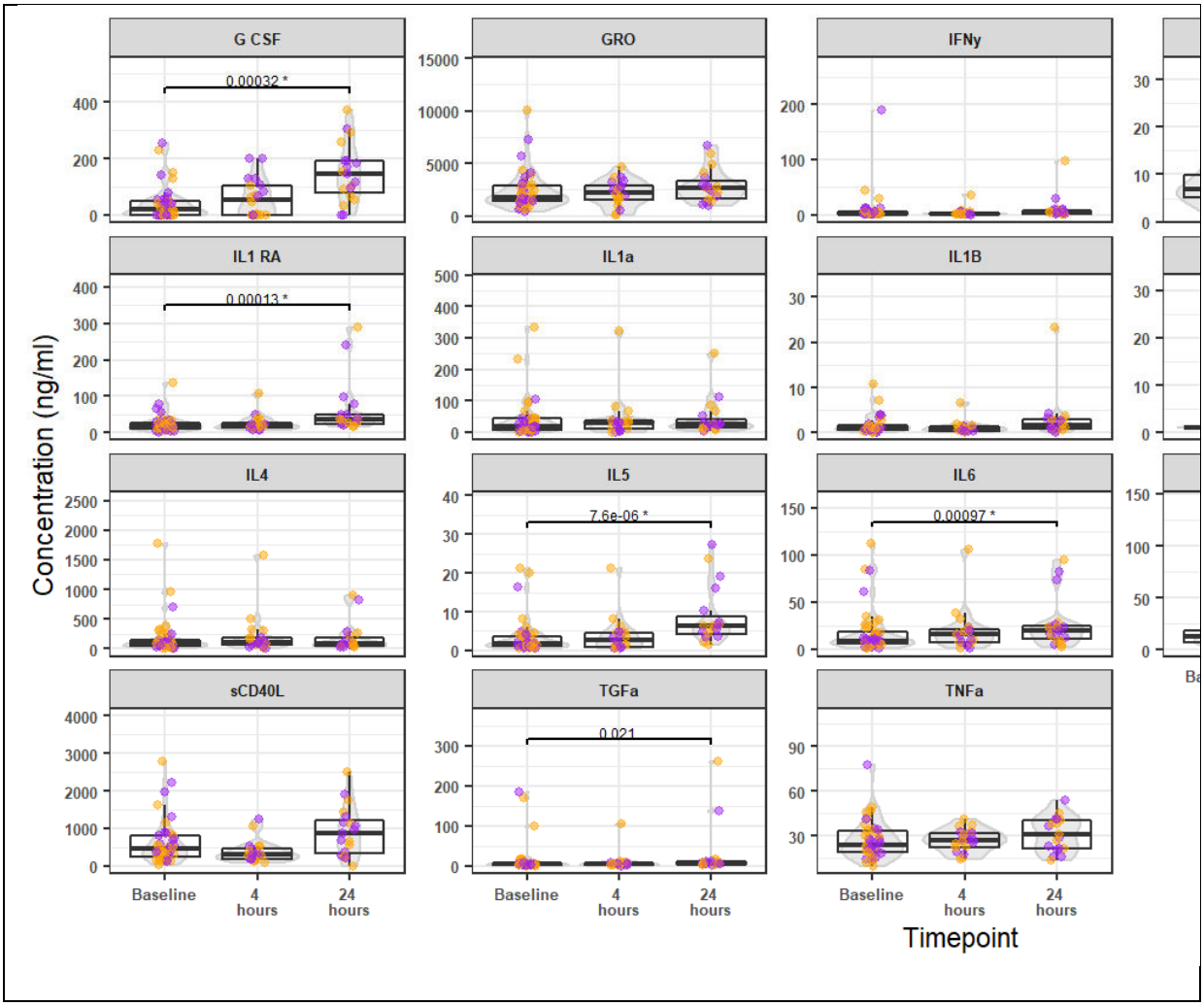
	Protein	*Log2 FC*	*Average abundance (normalised intensity)*	*p-value*	*FDR*	*Validation of 4 month result? (i.e. FDR < 0.05 @ 4 months, p<0.05 @ 12 months*	*Validation of 4 months results when restricting to truly independent cohort?*
Both groups combined Baseline vs 24 hours post-vaccination	Serum amyloid A-1 protein	4.04	1.09	2.3E-	**5.4E**	validated	Yes
	Serum amyloid A-2 protein	4.75	-2.98	3.6E-	**4.1E**	validated	Yes
	Lipopolysaccharide-binding protein	1.20	-0.31	5.0E-10	**3.8E-08**	validated	Yes
	Leucine-rich alpha-2-glycoprotein	0.85	4.62	6.5E-10	**3.8E-08**	validated	Yes
	C-reactive protein	1.64	-5.53	4.2E-	**1.9E**	validated	Yes
	Fibrinogen-like protein 1	1.33	-2.29	8.4E-	**3.3E**		
	Protein Z-dependent protease inhibitor	0.55	-0.10	1.6E-05	**5.5E-04**	validated	Yes
	Complement C1r subcomponent	0.62	-0.29	1.4E-	**4.0E**		
	Complement component C9	0.53	1.31	1.0E-	**0.03**	validated	Yes
	Coagulation factor VII	-0.57	-1.83	1.6E-	**0.03**		
	Apolipoprotein B-100	-0.29	5.45	1.6E-	**0.03**		
	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	0.49	-1.08	1.9E-03	**0.04**		
	Alpha-1-antichymotrypsin	0.36	0.15	0.01	0.14	validated	No
	Prostaglandin-H2 D-isomerase	-0.38	-1.96	0.03	0.22	validated	Yes
MenB vs D1	Serum amyloid A-1 protein	5.88	1.09	1.8E-	**4.2E**	Validated	Yes
	Serum amyloid A-2 protein	6.70	-2.98	3.0E-	**3.5E**	Validated	Yes
	C-reactive protein	2.58	-5.53	1.7E-	**1.3E**	Validated	Yes

	Lipopolysaccharide-binding protein	1.55	-0.31	2.7E-10	**1.6E-08**	Validated	Yes
	Leucine-rich alpha-2-glycoprotein	1.10	4.62	4.6E-10	**2.1E-08**	Validated	Yes
	Fibrinogen-like protein 1	1.89	-2.29	3.3E-	**1.3E**		Yes
	Protein Z-dependent protease inhibitor	0.68	-0.10	5.0E-05	**1.7E-03**	Validated	Yes
	Immunoglobulin lambda variable 1-36	-0.98	-2.65	1.9E-04	**0.01**		
	Apolipoprotein B-100	-0.43	5.45	3.8E-	**0.01**		
	Complement component C9	0.73	1.31	5.6E-	**0.01**	Validated	Yes
	Complement C1r subcomponent	0.67	-0.29	1.4E-03	**0.03**		
	Collagen alpha-1(I) chain	-1.11	-4.08	1.5E-	**0.03**		
	Selenoprotein P	-0.35	1.71	2.6E-	**0.05**		
	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	0.62	-1.08	2.7E-03	**0.05**		
	Alpha-1-antichymotrypsin	0.45	0.15	0.01	0.12	Validated	Yes
	Vitamin K-dependent protein C	-0.55	-4.27	0.04	0.27	Validated	
Routine baseline vs D1	Serum amyloid A-1 protein	2.05	1.09	2.7E-	**0.04**	validated	Yes
	Serum amyloid A-2 protein	2.63	-2.98	3.5E-	**0.04**	validated	Yes
	Lipopolysaccharide-binding protein	0.82	-0.31	5.4E-04	**0.04**	validated	Yes
	Leucine-rich alpha-2-glycoprotein	0.58	4.62	7.2E-04	**0.04**	validated	Yes
	Protein Z-dependent protease inhibitor	0.42	-0.10	0.01	0.35	validated	
MenB vs	Serum amyloid A-1 protein	-3.84	1.09	3.3E-	**7.7E**		
	C-reactive protein	-1.96	-5.53	5.1E-	**5.9E**	validated	Yes

Routine 1 day	Serum amyloid A-2 protein	-4.07	-2.98	2.5E-05	**1.9E-03**	validated



*Supplementary figure 1 Dot plots of all proteins assayed by Luminex in 4-month-old infants. Samples from infants that received concomitant 4CMenB = purple. Samples from infants that received routine vaccine = orange. Numbers represent raw p-values. Only p-values < 0.05 are shown. * denotes fdr < 0.05*



*Supplementary figure 2 Dot plots of all proteins assayed by Luminex in 12-month-old infants. Samples from infants that received concomitant 4CMenB = purple. Samples from infants that received routine vaccine = orange. Numbers represent raw p-values. Only p-values < 0.05 are shown. * denotes fdr < 0.05*

Appendix 1.1 The function of DE proteins

Supplementary Table 5 DE proteins, their leading production site and functions. Red upregulated between baseline and 24 hours or upregulated in 4CMenB group versus Routine group.

Protein (Alias) Gene symbol	Description/Main site of production	Function in the immune response	Significant associations at 4 months	Validated @ 12 months? (i.e. $p<0.05$)
C-reactive protein Gene Symbol: CRP	A short pentraxin / Liver	Acute-phase reactant. - Binds to phosphocholine on bacterial cell walls to promote agglutination, bacterial capsular swelling, phagocytosis and complement fixation. - Interacts with DNA and histones. May scavenge nuclear material released from damaged circulating cells.	Upregulated at 24 hours	Yes
			Higher in the 4CMenB group than routine vaccines at 24 hours	Yes
Serum amyloid A-2 protein Gene Symbol: SAA2	Apolipoprotein Liver	- Acute-phase reactant. - Apolipoprotein of the HDL complex- associates with HDL during inflammation - Soluble pattern recognition receptor. - Binds bacterial outer membrane protein A (OmpA). - Opsonises gram-negative bacteria and blocks virus entry into cells - Recruits immune	Upregulated at 24 hours	Yes
			Higher in the 4CMenB group than routine vaccines at 24 hours	Yes

		cells to inflammatory sites. Induces enzymes that degrade the extracellular matrix. - May promote survival of myeloid-derived suppressor cells, which attenuate the inflammatory response and induce cytokines and chemokines in immune and non-immune cells[532]		
Serum amyloid A-1 protein Gene Symbol: SAA1	As per serum amyloid A-2 protein Liver	As per serum amyloid A-2 protein	Upregulated at 24 hours	yes
Leucine-rich alpha-2-glycoprotein Gene Symbol: LRG1	Secreted glycoprotein Liver neutrophils macrophages endothelial cells—induced by TNFa	- Acute-phase reactant. - Function is not well known expressed during granulocyte differentiation involved in neutrophil differentiation promotes angiogenesis	Upregulated at 24 hours	yes
Lactotransferrin (Lactoferrin) Gene Symbol: LTF	Transferrin Activated Neutrophils Epithelial cells	- Present in body fluids - Present in granulocyte granules—released in significant quantities by activated neutrophils and contain in NETS.[533] - Antimicrobial properties[533] - Scavengers free	Upregulated at 24 hours	

		iron to prevent access by pathogens - Binds to LPS on gram-negative bacteria - Attenuation of host inflammatory response[533] - Mitigates oxidative stress - Competes with LPS for cell surface PRR receptors, e.g. TLR4 and CD14, to prevent their activation - Augments induction of the adaptive immune response[533]		
Lipopolysaccharide -binding protein Gene Symbol: LBP	Liver Adipocytes Intestinal cells	Acute-phase protein - Binds to LPS[534] - Enhances affinity of CD14 (coreceptor to TLR4) to LPS[534]	Upregulated at 24 hours	yes
Complement component C9 Gene Symbol: C9	Complement protein Liver Macrophages	- Pore-forming subunit of the membrane attack complex	Upregulated at 24 hours	yes
Protein Z-dependent protease inhibitor Gene symbol: SERPINA10	Member of the serpin superfamily of serine proteinase inhibitors Liver	Acute-phase protein - Inhibits coagulation by inhibiting factors Xa and XIa in the clotting cascade[534] - May inhibit IL6 and TNF-alpha production by monocytes[535]	Upregulated at 24 hours	yes
Alpha-1-antichymotrypsin Alias: AACT, ACT, GIG24, GIG25 Gene symbol: SERPINA3	Member of the serpin superfamily of serine proteinase inhibitors Liver	The physiological function is unclear - inhibits neutrophil cathepsin G and mast cell chymase, which convert angiotensin-1 to the active angiotensin-2	Upregulated at 24 hours	yes

		- likely protects tissues during the inflammatory response by inhibiting proteolytic enzymes		
Alpha-1-acid glycoprotein 1 Alias: Orosomucoid Gene symbol: ORM1	Member of the calycin superfamily, Lipocalin family Liver	Acute-phase protein - Protein transporter - binds basic and neutrally charged lipophilic compounds to transport them in circulation - Modulates the activity of the immune system during the acute-phase reaction	Upregulated at 24 hours	no
Apolipoprotein C-III Gene symbol: APOC3	Apolipoprotein Liver Small intestine	- Component of chylomicrons, high-density lipoproteins and very low-density lipoprotein - Promotes hypertriglyceridemia by inhibiting hydrolysis and clearance and hepatic uptake of triglyceride-rich lipoproteins	Downregulated at 24 hours	no
Retinol-binding protein 4 Gene symbol: RBP4	Member of the calycin superfamily, Lipocalin family Liver	- Transports retinol from the liver to the peripheral tissues.	Downregulated at 24 hours	no
Complement factor B Gene symbol: CFB	Member of the complement pathway Liver	Part of the alternate pathway of the complement cleaved into Bb and Ba. - Bb combines with complement factor 3b to generate the C3 or C5 convertase - Also implicated in proliferation and	Upregulated at 24 hours	no

		differentiation of preactivated B-lymphocytes, rapid spreading of peripheral blood monocytes, stimulation of lymphocyte blastogenesis and lysis of erythrocytes. - Ba inhibits the proliferation of preactivated B-lymphocytes		
Gelsolin Gene Symbol: GSN	lipid-binding actin regulatory protein Muscle	- Severs filamentous actin released by dying cells, which otherwise causes a hyper viscous environment that can cause microthrombosis and impair access by immune cells - Binds to LPS and lipoteichoic acid, which prevents activation of their respective TLRs - Binds to host lysophosphatidic acid, platelet-activating factor, and sphingosine 1-phosphate - falls during inflammation due to it being sequestered by scavenging of actin.	Downregulated at 24 hours	no
Apolipoprotein E Gene Symbol: APOE	Apolipoprotein Liver Macrophages	- Contained in chylomicrons, VLDL and HDL - Mediates the binding, internalisation, and catabolism of lipoprotein particles.	Downregulated at 24 hours	no

		- Inhibits neutrophil and macrophage activation - Facilitates lipid antigen presentation by CD1 molecules to natural killer T-cells (NKT) - Influences cholesterol in the cell membrane of dendritic cells—a reduction in APOE enhances MHC-II-dependent antigen presentation and $CD4^+$ T-cell activation[536]		
Serum amyloid A-4 protein Gene symbol: SAA4	Serum amyloid family Liver. Vitamin A is required for its expression.[537]	- Constitutively expressed not thought to be induced in the acute-phase response - see SAA1 and SAA2 above	Upregulated at 24 hours	no
Alpha-1-acid glycoprotein 2 Alias: Orosomucoid 1 Gene symbol: ORM1	Member of the calycin superfamily, Lipocalin family Liver	See Alpha-1-acid glycoprotein 2 above	Upregulated at 24 hours	no
Prostaglandin-H2 D-isomerase Gene symbol: PTGDS	Brain Dendritic cells Male reproductive tract	- catalyses prostaglandin H2 conversion to prostaglandin D2, which promotes mast cell maturation and degranulation, smooth muscle contraction/relaxation and inhibits platelet aggregation[533].	Downregulated at 24 hours	no
Serum amyloid P-component	Pentraxin	- soluble pattern recognition receptor	Upregulated at 24 hours	no

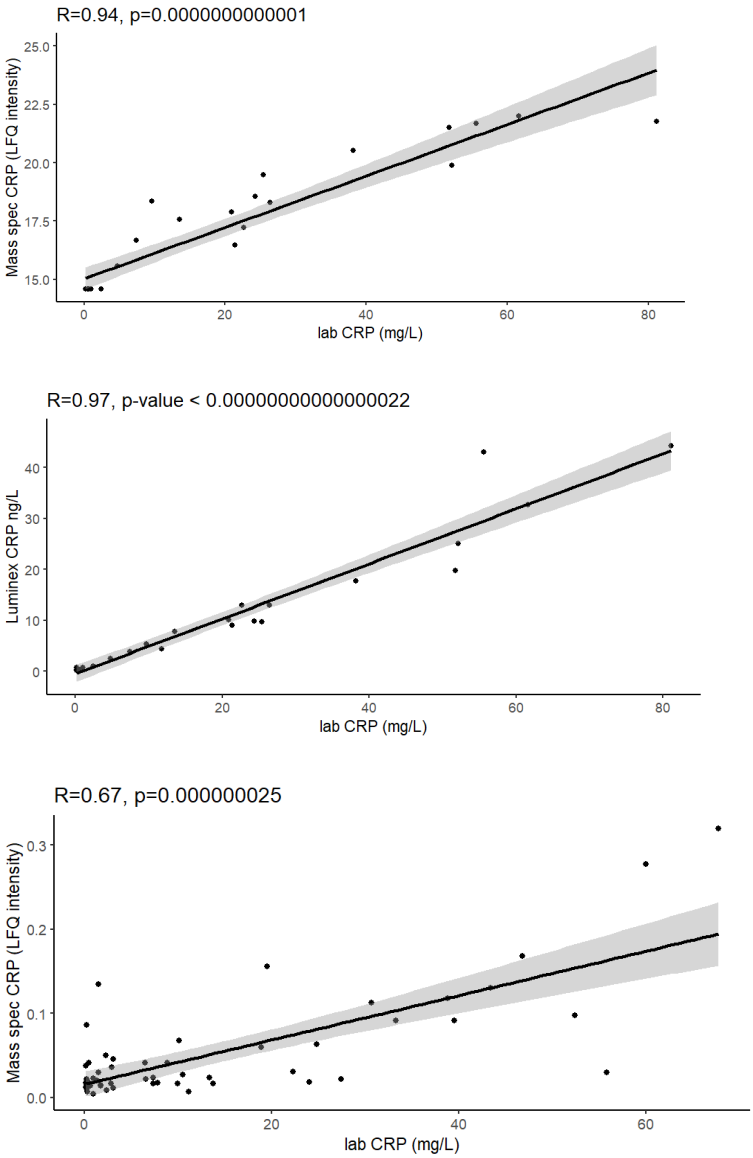
Gene symbol: APCS	Liver	- may scavenge nuclear material released from damaged circulating cells. - May function as a calcium-dependent lectin - activate complement7 - interact with Fcγ receptors7		
Haptoglobin Gene symbol: HP	Liver	- APP - Binds free haemoglobin in circulation released during haemolysis - regulates the immune response[538] - may have anti-inflammatory functions - inhibition of extracellular proteinases neutralisation and clearance of harmful components such as released haemoglobin, heme, metal ions, and reactive oxygen and nitrogen species[539] - modulate immune cell activities. - may facilitate antigen presentation - may act as a coactivator of T-cells but also shown to TH2 cytokine release in-vitro[540] - knock out mice suggest that it augments lymphocyte development	Upregulated at 24 hours	no
Vitamin K-	Serine protease	- anticoagulant—	Downregulate	Yes, in

dependent protein C Alias Protein C, autoprothrombin IIA and blood coagulation factor XIX Gene symbol: PROC	liver	regulates clotting by inactivating factors Va and VIIIa - Anti-inflammatory	d at 24 hours	the 4CMenB group, only
Alpha-1-antitrypsin Gene symbol SERPINA1	Serine protease liver	Inhibits trypsin, chymotrypsin, plasminogen activator, elastase, plasmin and thrombin - protects against damage by enzymes produced in the immune response	Upregulated at 24 hours	no
Fibrinogen gamma chain Gene symbol: FGG	Coagulation factor liver	- polymerises with fibrinogen alpha and fibrinogen beta to form a fibrin matrix during clotting. - found in inflammatory foci Protects against interferon gamma-mediated haemorrhage during infection. - facilitates migration of leukocytes across endothelium[541] - May potentiate leukocyte activation and cytokine secretion[541]	Upregulated at 24 hours	no
Apolipoprotein C-I Gene symbol: APOC1	Apolipoprotein Liver macrophages	- a constituent of HDL, VLDL and chylomicrons - binds free fatty acids - Inhibits lipoprotein binding to the LDL receptor, LDL receptor-related	Higher in the 4CMenB group than routine vaccines at 4 hours	no

		protein, and VLDL receptor that occurs via apolipoprotein E - major plasma inhibitor of cholesteryl ester transfer protein (CETP). - shown to bind to LPS *in-vitro*		
Apolipoprotein A-I Gene symbol: APOA1	Apolipoprotein Liver intestine	- Constituent of HDL - triggers esterification of cholesterol so that it can be loaded into HDL for transport to the liver - and by acting as a cofactor for the lecithin cholesterol acyltransferase (LCAT).	Higher in the 4CMenB group than routine vaccines at 4 hours	no
Apolipoprotein A-II Gene symbol: APOA2	Apolipoprotein Liver	- Abundant in HDL - May stabilise HDL (high-density lipoprotein) structure by its association with lipid	Higher in the 4CMenB group than routine vaccines at 4 hours	no
Granulocyte colony-stimulating factor Gene symbol: CSF3	fibroblasts, endothelial cells, monocytes, macrophages and bone marrow stromal cells production induced by IL-1B, TNFa, and LPS	- stimulates the production, differentiation, release and function of granulocytes		
GRO-alpha Alias: Growth-regulated alpha protein; CXCL1	Type of alpha chemokine Macrophages, neutrophils and epithelial cells. Induced by NF-	- Chemoattractant - Recruits and activate neutrophils		

	κB and C/EBPβ signalling pathways receptor: CXCR2			
IL1 RA Gene symbol: IL1RN	Most cells of the body Including epithelial cells, adipocytes, leukocytes, hepatocytes In response to IL-1B Receptor: IL1R1	- Inhibits the activity of interleukin-1 (a potent pro-inflammatory cytokine) by non-productively binding to the receptor IL1R.		
IL5 Gene symbol: IL5	T-helper 2 cytokine Th2 lymphocytes, Mast cells Receptor: IL5R	- stimulates B-cell growth and terminal differentiation - increases immunoglobulin secretion - primarily IgA.		
IL6 Gene symbol: IL6	Macrophages Liver Induced by TNFa Receptor: IL6R	- Acts on B-cells, T-cells, hepatocytes, hematopoietic progenitor cells and cells of the CNS - Pro-inflammatory cytokine - stimulates an acute-phase response in the liver - encourages the production of neutrophils - inhibits T-regulatory cells - promotes differentiation of B-cells into Ig-secreting cells - promotes differentiation of Th17 cells. endogenous pyrogen - Anti-inflammatory		

		myokine		
TGFa, Transforming growth factor-alpha Gene symbol: TGFA	Macrophages, brain cells, and keratinocytes Receptor: epidermal growth factor receptor	Growth factor—promotes cell proliferation		
L selectin Alias: CD62L, SELL Gene symbol: SELL	Leukocytes	- cell adhesion molecule found on leukocytes. - acts as a homing marker for lymphocytes to enter secondary lymphoid tissue. - Shed from leukocyte surface during cell activation.		



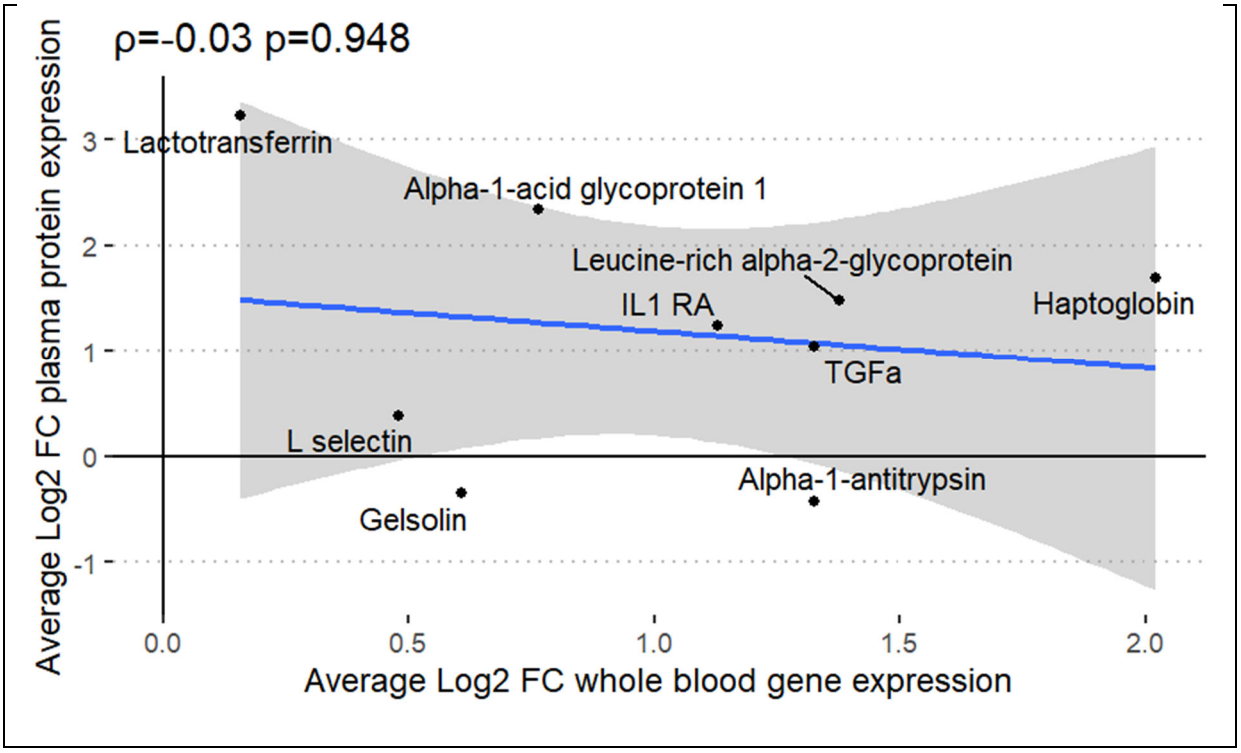
Supplementary figure 3 correlations between CRP measurements

Upper panel: correlation between clinical lab measured CRP vs mass spec intensity in the 4-month-old infant

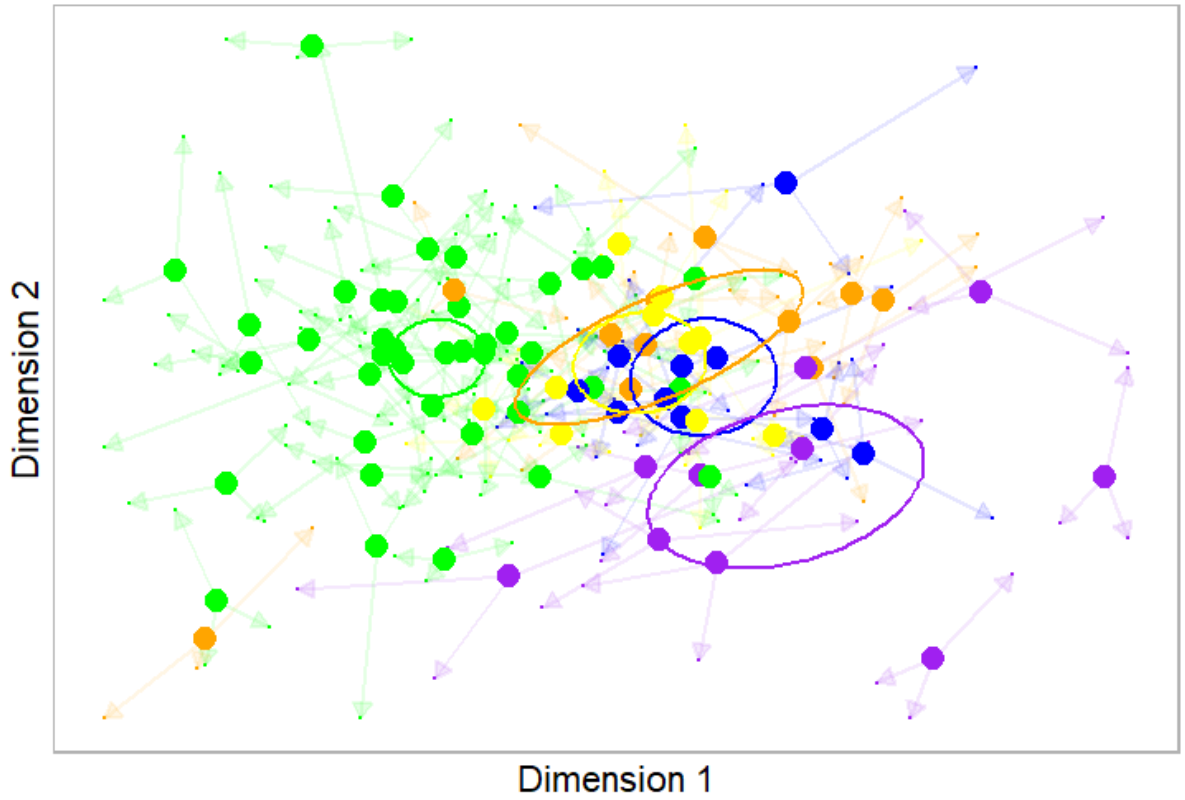
Middle panel: correlation between clinical lab measured CRP vs Luminex concentrations in the 4-month-old infants

Lower panel: correlation between clinical lab measured CRP vs mass spec intensity in the 12-month-old infants

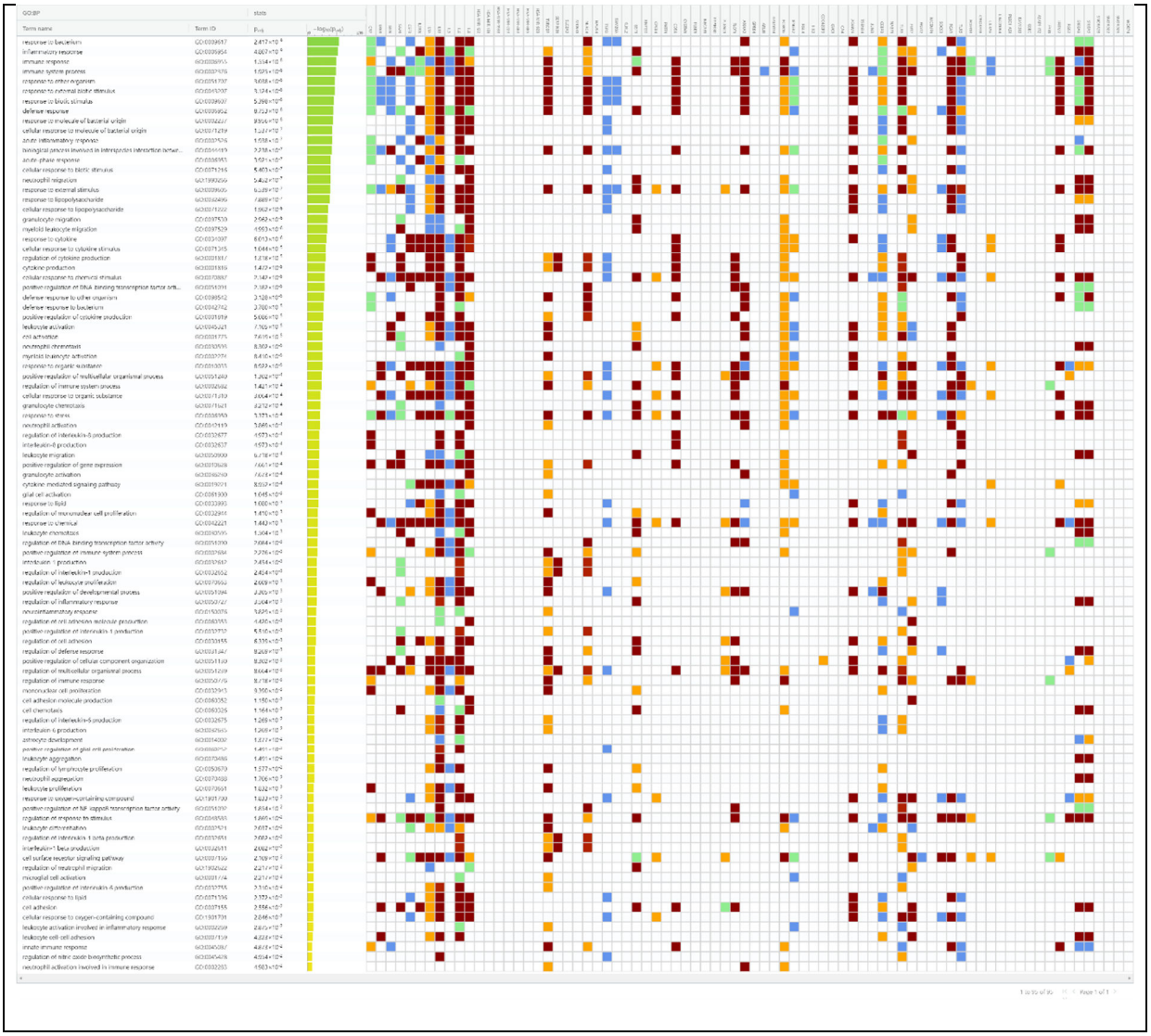
9.4 Appendix for Chapter 5



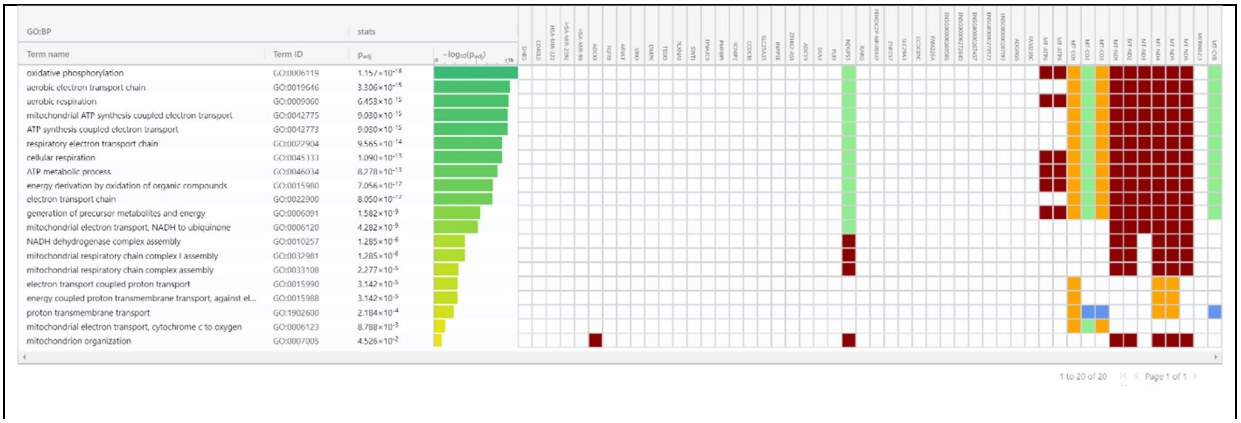
Appendix Figure 23 Agreement plot of log_2 foldchanges in the mRNA and protein data for the baseline vs 1 day post vaccination comparison in 4-month-old participants of EUCLIDS.



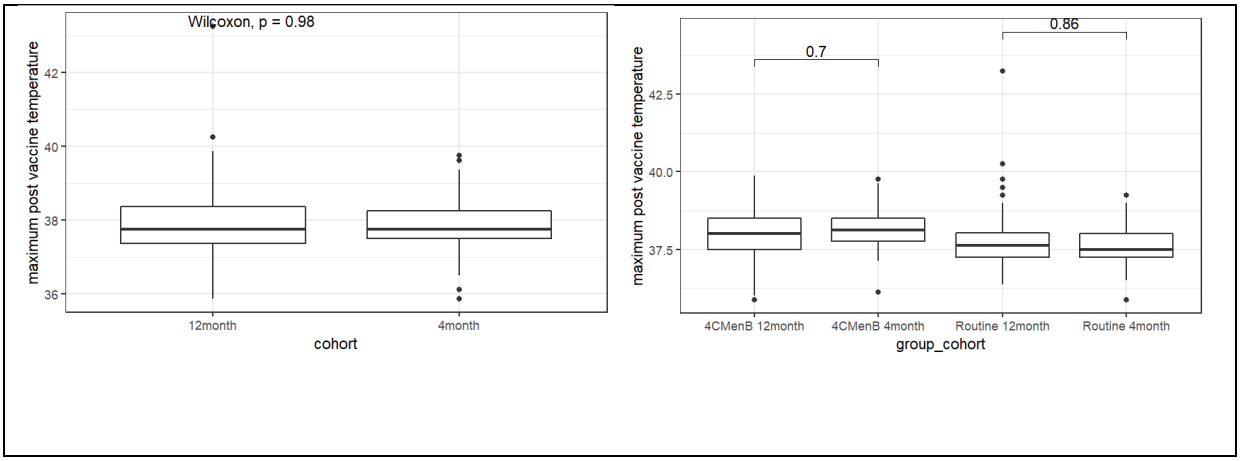
Appendix Figure 24 CCA based on mRNA, miRNA and sRNA alone. It shows that the separation of vaccine groups is not just due to proteins.



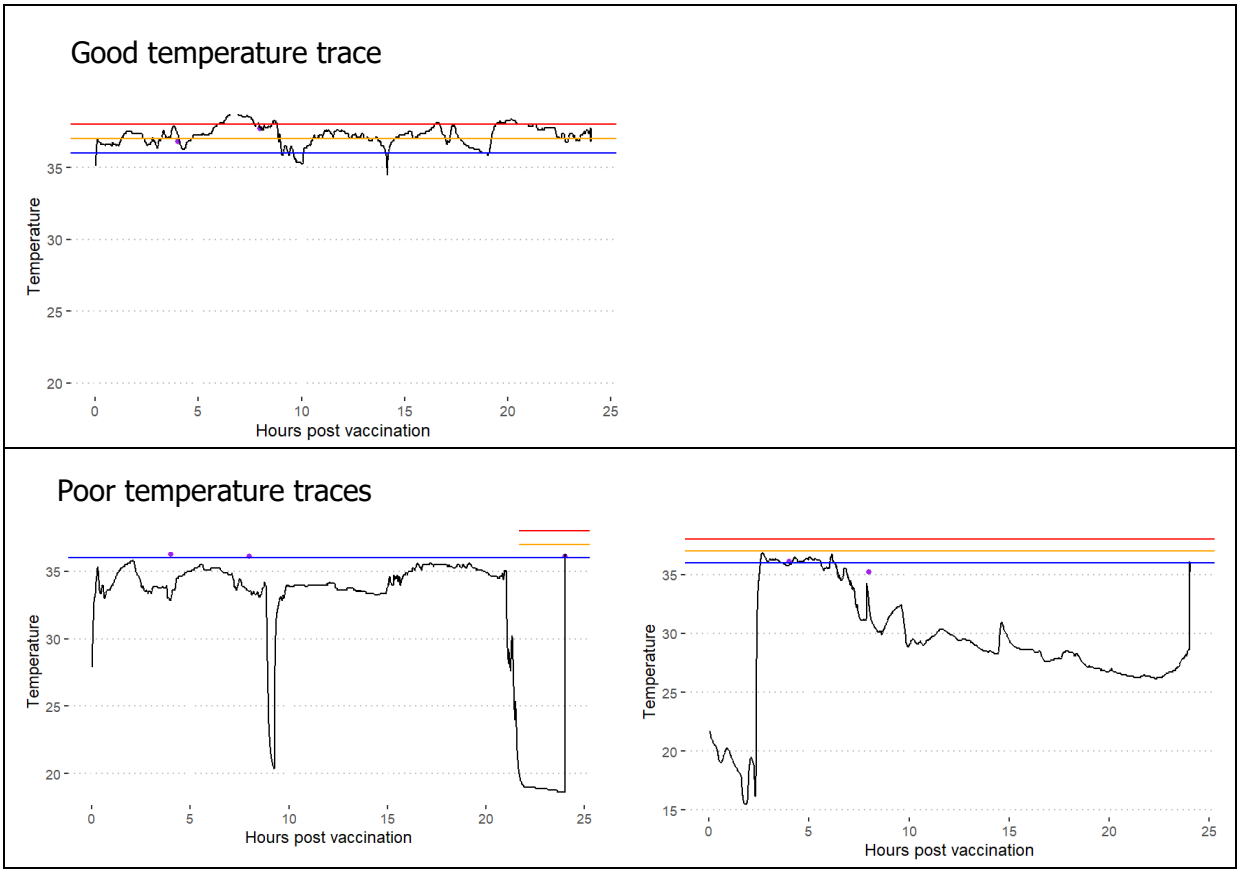
Appendix Figure 25 GSEA using gprofiler on cluster 1 of the relevance network built on the CCA output of pre and post-vaccine samples in both cohorts



Appendix Figure 26 GSEA using gprofiler on cluster 2 of the relevance network built on the CCA output of pre and post-vaccine samples in both cohorts

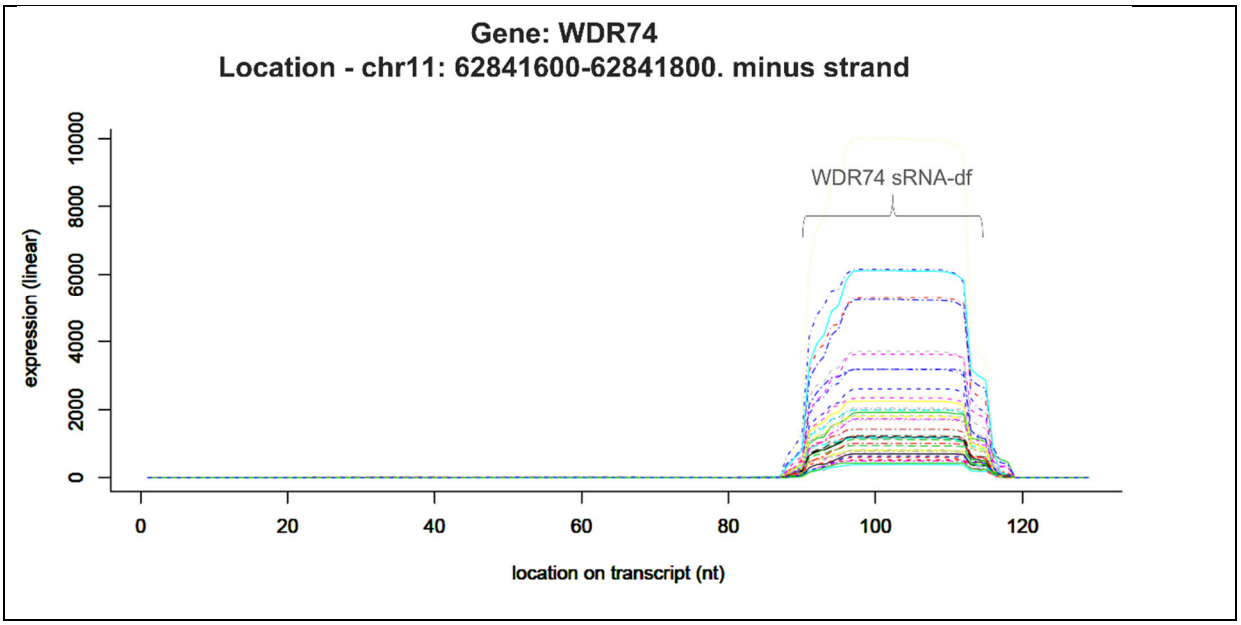


Appendix Figure 27 Maximum temperature post-vaccination in each cohort and each cohort-group



Appendix Figure 28. Examples of temperature of ibutton temperature traces. The upper graph is an example of a reliable looking trace. The lower two graphs are examples of unreliable traces with temperatures outside of the normal range of body temperature for

a healthy child (lowest range for normal temperature is around 36°C). Blue line 36°C, yellow line 37°C, red line 38°C. Purple dots show corresponding axillary temperatures.



Appendix Figure 29 Expression profile along the WDR74 gene covering the fragment that is identified in the sRNA-seq data. Lines represent the expression profile of this locus in different samples.

This sRNA-df was upregulated in plasma 24 hours after vaccination. All reads come from a single locus within the WDR74 gene and show a similar profile, which only differs by abundance. This is different to the pattern of random uniform expression profiles characteristic of degradation products. This suggests systematic processing of the WDR74 RNA into single fragments that span ~30 nucleotides of the gene. This suggests it could be a biologically functional sRNA.

9.5 Appendix for Chapter 6

Appendix 2.1

9.5.1 Additional background information on *S.* Typhi and typhoid fever

9.5.1.1 Clinical features

Typhoid is characterised by a non-specific, febrile, flu-like illness. Features, which are specific but not always present include a rose spot rash, typhoid tongue (a yellowish coating on the centre of the tongue with erythema around the edges), and relative bradycardia (absence of tachycardia in febrile patients)[444]. Common but non-universal laboratory findings include leukopenia, thrombocytopenia, raised inflammatory markers and elevated liver transaminases[444].

9.5.1.2 Outcomes in typhoid: recovery, death or chronic carriage

Though typhoid is primarily a self-limiting illness, it is fatal in up to 10 to 20% of untreated cases, predominantly from encephalopathy, nephritis, hepatitis and intestinal perforation[441]. Two to five per cent of people who survive typhoid become long-term carriers, which can last decades. Carriers maintain typhoid transmission within communities by shedding *S.* Typhi in stool and, to a lesser extent, urine [442,443] [442,443].

Infection with *S.* Typhi provides incomplete immunity, which wanes over time. A human challenge model of infection showed that prior infection with *S.* Typhi conferred only 23% protection against challenge; these findings are somewhat supported by another study, which found that a previous challenge with *S.* Typhi resulted in a relative risk of typhoid of 0.70 after a second challenge though this was non-significant (95% CI 0.41–1.21; p = 0.24), possibly due to inadequate study power[542,543].

9.5.1.3 Diagnosis

The non-specific signs, symptoms and lab results of typhoid make diagnosing enteric fever difficult. Consequently, diagnostic criteria based on clinical features have limited specificity and sensitivity[544]. The gold standard diagnostic method is bone marrow aspiration, though, for obvious reasons, this is rarely used. Blood culture is the most sensitive and specific commonly used diagnostic method but can take at 48 hours to reveal a positive result, are not widely available in endemic settings, and sensitivity is dependent upon adequate blood volumes and appropriate sample handling; moreover, prior antibiotic use may render them sterile. A systematic review and meta-analysis estimated that the sensitivity of blood culture is around 0.51 for a 2-mL blood sample (typically a paediatric blood sample), improving to 0.65 for a 10-mL blood sample[450].

Point-of-care tests for typhoid are based on detecting serum antibodies against *S.* Typhi antigens, usually H (the flagellar), O (the outer oligosaccharide domain

of LPS) or the Vi antigens. These tests are fast, cheap and easy to use but show moderate diagnostic accuracy due to the non-specific cross-reactivity of antibodies generated against *S.* Typhi, inability to differentiate acute from past infection, and negativity in the early stages of infection [545546].

9.5.1.4 Typhoid pathogenesis and the host response.

S. Typhi is a gram-negative, flagellated, facultative, anaerobic bacteria characterised by the -O (the outer oligosaccharide domain of LPS), -H (flagella) and Vi antigens[547]. Most understanding of the pathogenesis of typhoid is extrapolated from *in-vitro* work and some *in-vivo* work in mice infected with of *S.* Typhimurium because it creates a typhoid-like illness in rodents.

After ingestion, *S.* Typhi enters the stomach, where most bacteria will be killed by gastric acid. The viable bacteria that make it to the terminal ileum must navigate past a protective mucous layer and the glycocalyx—a layer of proteoglycans, glycoproteins, and glycolipids bound to the plasma membrane of intestinal epithelial cells[548]. *S.* Typhi then attaches to the gut epithelium via the fimH protein, a mannose-binding adhesin found at the tip of the fimbria. The specificity of *S.* Typhi fimH binding is thought to make *S.* Typhi human-estricted[549]. It is generally thought that *S.* Typhi traverses the gut wall by crossing through M cells, specialised antigen-sampling epithelial cells that lie over the Peyer's patch in the terminal ileum. Some recent co-culture work and ex-vivo biopsies suggest that take-up may also occur via enterocytes[517,550]. Adherence to the epithelial cell membrane is facilitated by *S.* Typhi's SPI9 protein and its STIV outer membrane protein, which bind to ligands of the host

cell membrane[447]. Invasion into the cell is mediated by *S.* Typhi's type three secretion system 1 (T3SS1)—this is a needle-like structure, which injects effector proteins (SipA, SipC) through the epithelial cell membrane. The effector proteins cause restructuring of the cytoskeleton so that *S.* Typhi becomes engulfed, facilitated by motility from its flagella. Once inside the cell, *S.* Typhi downregulates T3SS1 and flagellin (to avoid TLR5 recognition) and expresses the Vi polysaccharide capsule. The Vi polysaccharide is thought to be a crucial immune evasion strategy that acts as a physical barrier between its LPS and detection by host cell TLR4 receptors[447]. The Vi polysaccharide capsule also helps *S.* Typhi to avoid complement activation and killing because it contains no free hydroxyl groups to, which C3b can bind[448]. Avoidance of complement-mediated killing may partly explain why serum bactericidal antibodies are not associated with protection in a human challenge model of infection[551]. The ability of *S.* Typhi to bypass detection by the immune response separates it from the non-typhoidal serovars of *Salmonella* that usually induce an associated neutrophilic secretory diarrhoea. As an aside, *S.* Paratyphi is a bit of an enigma—these serovars cause enteric fever despite lacking a Vi capsule suggesting Paratyphi can suppress TLR4 mediated proinflammatory pathways after its LPS activates them.

After passing through the epithelial cell, *S.* Typhi is taken up by monocytes. Once inside the vacuole of monocytes, *S.* Typhi expresses anti-oxidative proteins, which protect it from hydrogen peroxide produced by the oxidative burst. This strategy allows *S.* Typhi to hijack monocytes and become disseminated through the body. It is unclear whether infection results from the breakthrough of a single bacterium into the bloodstream (bottleneck hypothesis)

or several bacteria. Interestingly, primary bacteraemia in human challenge models does not inevitably lead to acute typhoid, indicating that the immune response can neutralise *S.* Typhi typhoid after systemic spread[449].

For a long time, the ability of *S.* Typhi (and *S.* Paratyphi) to cause enteric fever was thought to be due to the typhoid toxin. This toxin is released into cells and secreted into the extracellular environment. It was thought to be a virulence factor that enables Typhi and paratyphi to subvert the immune response. A recent human challenge model failed to show any difference in infection rates between people challenged with wild-type *S.* Typhi versus those challenged with a knock-out toxin strain[552]. These findings indicate that the toxin is not integral to pathogenesis. The toxin may be important for establishing long-term carriage rather than acute disease[552].

Once *S.* Typhi reaches the liver and gall bladder, it re-enters the GI system and is shed into stool. Gallstones provide an ideal substrate for forming biofilms, which shield the bacteria from the immune response, setting up the conditions for chronic carriage and long-term shedding of *S.* Typhi into the gastrointestinal tract[553].

9.5.1.5 Typhoid vaccines

The first typhoid vaccines were heat- and phenol-inactivated and acetone-killed whole-cell vaccines and emerged in the late 1800s[554]. They were widely used in the British and American armies and in Thailand and Sri Lanka[554–556]. However, despite their efficacy (estimated at 51%-88%), their use was withdrawn due to an unacceptable reactogenicity profile[555]. Second-generation typhoid vaccines comprise the oral live attenuated vaccine—Ty21a and the injectable Vi polysaccharide vaccine (Vi-PS).

TY21a, tradename Vivotif® (PaxVax), protects for 5 to 7 years with a three-year efficacy rate of around 51% (36–62%)[555]. Vi-PS (tradenames: Typhim®, Sanofi Pasteur, and Typherix ®, GSK) is a subunit vaccine containing purified Vi capsular polysaccharide of *S.* Typhi strain Ty2. Estimated efficacy is 55% (30–70%) and protects for around 3 years[555]. Both Ty21a and Vi-PS are restricted to the over 2s. For the TY21a this is partly due to a formulation issue, for the Vi-PS vaccine it is because the Vi polysaccharide is a T-cell independent antigen and thus poorly immunogenic in children under two (see 1.2.1.2.3.2). This is problematic as under 2s represent a substantial proportion of typhoid morbidity and mortality.

The Vi conjugate vaccine (Vi-TCV) is a recent solution to this issue. VI-TCV comprises vi-polysaccharide conjugated to tetanus toxoid, so it can elicit T-cell-dependent immune responses and therefore is immunogenic in younger infants. It was WHO pre-qualified in 2018 as a single dose schedule for children over 6 months off the back of the VAS study above[40]. Phase three field trials of Vi-TCV

in endemic settings show excellent efficacy. Preliminary data from a Nepalese paediatric trial found 1-year protective efficacy of 81·6% (95% CI 58·8–91·8)[469]. A phase three trial of Vi-TCV in districts within Hyderabad, Pakistan, found an efficacy of 95% (95% CI 93%–96%) against blood-culture-confirmed *S.* Typhi and 55% (95% CI 52%–57%) against suspected typhoid fever[470]. The success of Vi-TCV in clinical trials has led to its inclusion in childhood vaccine schedules funded by the global alliance against vaccines (GAVI) for low and middle-income countries.

9.5.1.6 Challenges in Typhoid

9.5.1.6.1 Diagnosis

As discussed above, diagnosing typhoid is difficult due to its (mostly) non-specific signs and symptoms, the insensitivity of blood culture under field conditions and the poor performance of lateral flow/point-of-care tests.

9.5.2 Key findings in other datasets from VAST

Appendix 3.1

Key findings from VAST are in Appendix Table 6 below. The results raise several discussion points. Firstly an increase in NK cell subsets 7 days after vaccination negatively correlated with TfH and memory B-cells in the Vi-TCV group, suggesting that NK cells inhibit TfH cells and thus B-memory cells (see introduction for information on TfH and B-cell interactions). Secondly, an increase in immunological markers of mucosal immunity predominates Vi-vaccine responses (increase in IgA antibodies and gut homing markers on plasma cells). This is unexpected since the vaccination site usually dictates where subsequent immunity is targeted to—IgG is the predominant antibody

class induced after intramuscular vaccination, whereas IgA responses are minimal, and lymphocytes do not show mucosal migratory homing markers. It is postulated that the increase in mucosal-associated immunity may be due to Vi-antigen activating mucosal memory B-cells induced by gut flora that express Vi[557].

Higher IgG and IgA Vi-antibody titres were associated with protection after challenge. However, average Vi IgA titres in protected Vi-PS vaccinees were similar to average titres in diagnosed Vi-TCV vaccinees, i.e., for reasons not yet understood, levels of IgA that protected Vi-PS vaccinees from typhoid did not protect Vi-TCV vaccinees. As a result, a threshold for protection could not be identified. The discrepancy in protective Vi IgA titres between vaccine groups potentially indicates an aspect of Vi-antibody mediated protection differs between vaccines (*e.g.,* antibody Fc mediated function). This is a working hypothesis and is somewhat supported by identifying different correlates of protection in Vi-TCV and Vi-PS vaccinees using machine learning. A definitive explanation remains elusive, however.

An alternative consideration is that Vi antibody titres are not mechanistic correlates of protection, but their fold-rises correlate with a mechanistic response not yet identified. If so, it’s difficult to fathom what that actual mechanistic correlation could be—T-cell responses, for example, are not expected against Vi. Another possibility is that an aspect of Vi-TCV responses counteracts the protective effects of Vi-antibody, e.g., regulatory cells. The final explanation is that Vi-IgA is a mechanistic correlate of protection (hence Vi

vaccines protect against typhoid) and is also a marker of immuno-competence, i.e. the ability of an individual to respond to a vaccine and thus reflects an immunological capacity to respond to a pathogen.

Appendix 4.1

Appendix Table 6 Key findings in VAST in datasets generated by others

Output	Finding	Reference if published
Vaccine effectiveness	Vi-TCV 54·6% (95% CI 26·8-71·8) effective Vi-PS 52·0% (23·2-70·0) effective Vi-TCV associated with milder disease in terms of symptoms, quantity of *S.* Typhi in blood culture and number of days of stool shedding compared with control	558,559
Serological readings		559
Effect of vaccine on Vi antibody titres	Increase in Vi- IgG1, IgG2, and IgG3 subclass, IgM and IgA titres 28 days after vaccination in both vaccines, but increases were higher in Vi-TCV vaccinees All antibody titres significantly positively correlated with each other All antibody titres correlated all plasma cells frequencies, with few exceptions—linking B-cell numbers with antibody responses.	
Antibody functions	Increases in Vi-specific antibody functional responses: ADCD (antibody-dependent complement deposition), ADCP (antibody-dependent complement deposition) and ADNP (antibody-dependent neutrophil phagocytosis) at 28 days post-vaccination - All were increased in the Vi-TCV group relative to Vi-PS, but this only persisted after correcting for antibody titres in the ADNP groups ADNP correlated with Vi-IgG1 and Vi-IgG2	
Serological correlates of protection	Associations between IgA and IgG Vi titres and challenge outcome but no predictive threshold for protection established Machine learning identified slightly different correlates for Vi-PS and Vi-TCV based on measurements taken 28 days post-vaccination.	455

	Absolute titre levels and foldchanges in IgA and IgG associated with protection when both Vi-vaccine groups are combined. - For foldchange, these associations only persists for IgA (Vi-PS group only) - For absolute titres, these associations persist for IgG and IgA (Vi-PS group only) IgA associated protection in Vi-vaccinees - Protection after Vi-PS was positively associated with IgG, IgA, IgM, polyfunctional antibodies and antibody dependent cell phagocytosis (ADCP) and negatively associated measures of antibody dependent NK cell activity (ADNKA), and antibody binding to the immunoglobulin Fc receptor FCYR2B. - Protection after Vi-TCV was positively correlated with IgA quantity and avidity, and IgG1 avidity, and negatively associated with ADNKA, ADCP, and IgG1 titres - Serological correlates of protection were associated with PBMC cell profiles 24 hours after challenge	
Whole blood gene expression studies		
Whole blood gene expression correlates of protection after vaccination	B-cell receptor sequences in whole blood 7 days vaccination after, which correlated with Vi titres and protection (unpublished).	unpublished
Effect of *S.* Typhi exposure on whole blood gene expression	Validation of DE genes 12 hours after challenge. Increase in neutrophil related genes. No analysis of TD/D7 samples yet	
Plasma cytokines		

Effect of *S.* Typhi challenge on plasma cytokines	CXCL1(GROa), CX3CL1 (Fractalkine) and sCD40L increased 12 hours after challenge	[557]
Flow Cytometry		40,51
Effect of vaccination on lymphocytes	Vi specific plasms cells: Vi-specific IgG, IgM plasma cell responses seen at 7 and 10 post-vaccination. - Day of peak response varied amongst participants. - Frequencies were higher in Vi-TCV vaccinees. Vi-specific memory cells: Vi-IgG memory B-cells increased at 10 days post-vaccination. Not measured at 7 days post-vaccination. (Vi-TCV group only) - Significant heterogeneity in memory cell responses amongst Vi-TCV recipients: 1 in 5 participants produced no detectable B-cells Tetanus specific B-cells - IgG plasma cells peak 7 days but still present at 10 days (Vi-TCV only) - Tetanus specific IgG memory cells seen at 10 days post-vaccination and still elevated at 28 days post-vaccination. T follicular heleper cells: Increase in TfH cells at day 7 (Vi-TCV group only). Also seen in CyTOF data. - No relationship between Vi memory response and Vi-specific plasma cells	
Vi-specific, cellular correlates of protection	Association between Vi-specific plasma cells 7 days after vaccination associated with protection - But does not persist at the subgroup level.	
PMBC CyToF		51
Changes in PBMC cellular composition after	Plasma cells - Increase in plasma cells expressing IgA or IgG and mucosal migratory markers (α4β7+ or CCR10) 7 days after vaccination (Vi-TCV group only)	

vaccination	- Plasma cells with mucosal migratory makers were strongly correlated with ADNP and Vi-IgA1 titres but not Vi-IgA2 titres (VI-TCV group only) - Plasma cells with mucosal migratory makers were strongly correlated with Vi-IgA2 (Vi-PS vaccinees only). - Positive correlation between fold rise in IgG+ plasma cells expressing mucosal migratory markers and frequencies of Vi-IgG ASCs (measured by flow) suggesting the increased numbers of plasma cells are Vi-specific T-follicular helper cells - Increase in T-follicular helper cells 7 days after vaccination (Vi-TCV group only). T-follicular cells with a Tfh-1 and Tfh-17 were both expanded. - Tfh responses associated with IgA+ plasma cells expressing mucosal migratory makers. But this was not the case for IgG+ plasma cells. - Tfh response was associated with Vi-specific memory response Dendritic cells - Increase in dendritic cells subsets - Increase in activated dendritic cells. NK cells - Increase in NK cell subsets - Negative association between induction of TfH and several CD56bright NK cell subsets, and this was associated with poorer Vi-specific memory B-cell responses. Interpreted to mean NKs targeting of Tfh supresses development of memory B-cells
Changes in PBMC cellular composition early after challenge	Modest association between vaccine status and PBMC profiles PBMC composition was perturbed within 24 hours of challenge regardless of outcome. - reduction in CD62L+ $CD56^{dim}$ NK cells - reduction in plasmacytoid dendritic cells - increase in lymph node homing markers on T-cells - increase in TH_{17} and Cytotoxic T_{17} cells - increase in memory B-cells homing to lymphnodes

	PBMC composition was associated with challenge outcome: - PLS-DA analysis identified a correlation between PBMC cell profile and challenge outcome. These profiles was correlated with serological correlates of protection. - The protective profile was characterised by activated monocytes and naïve B-cells, and the susceptible profile was characterised by higher IgA and IgG+ B-cells that had a antigen presenting phenotype (as opposed to an antibody secreting phenotype). These features also correlated with bacteraemia at diagnosis, suggesting a spectrum of vaccine mediated protection from immune to severe infection.	
Cellular correlates of outcome	- Higher baseline frequencies of naïve B-cells and expansion of CD16+ monocytes associated with protection from typhoid - Higher baseline frequencies of terminally differentiated T-cells and a higher relative proportion of CD62L+ memory IgA+ and IgG+ B-cells associated with development of typhoid (CD62L is a homing marker for lymph nodes) - This was interpreted to mean that, paradoxically, a PBMC profile suggestive of greater antigen experience was associated with infection	
Changes in PBMC cellular composition at typhoid diagnosis	- neutrophil-to-lymphocyte ratio (NLR) and eosinophil-to-lymphocyte ratio (ELR) significantly perturbed at diagnosis and were associated with symptom severity - Vi-TCV vaccination, but not Vi-PS, significantly lowered NLR and number of self-reported symptoms. - NLR significantly negatively correlated with Vi-IgG (but not Vi-IgA or Vi-IgM) titres suggesting Vi-IgG titres may play a significant role in reducing disease severity.	
Full blood count data at typhoid diagnosis	- Increase in neutrophils and monocytes (particularly activated classical monocytes) - Downregulation of eosinophils and lymphocytes - Upregulation in CD4 T-lymphocytes and naïve B-cells. - Downregulation in memory B-cells, NK cells and T-cells.	51

Appendix Table 7 Participant-level demographic information

Unique identifier	number of samples sequenced	Vaccine	outcome	late shedder	number of day9 onwards stool samples	Typhoid clinically likely	Peak elispot day	Max symptom score	IgA FC baseline D0	IgA LFC baseline D0
1	5	Vi-TCV	diagnosed	no	4	yes	non.responder	0	14.1155	1.149696266
2	6	Vi-PS	non diagnosed	no	5	no	D7.responder	1	#N/A	#N/A
3	4	Vi-PS	non diagnosed	no	5	no	NA	0	#N/A	#N/A
4	6	Vi-TCV	non diagnosed	no	6	no	D10.responder	1	#N/A	#N/A
5	6	Vi-TCV	non diagnosed	no	3	yes	D7.responder	3	#N/A	#N/A
6	3	Control	diagnosed	no	2	yes	N/A	6	1	0
7	7	Vi-PS	diagnosed	no	2	yes	non.responder	12	#N/A	#N/A
8	3	Vi-PS	diagnosed	no	2	yes	non.responder	5	#N/A	#N/A
9	6	Vi-PS	non diagnosed	no	5	no	non.responder	2	#N/A	#N/A
10	4	control	non diagnosed	no	2	no	no	4		
11	7	Vi-TCV	diagnosed	late_shedder	6	yes	D10.responder	6	#N/A	#N/A
12	6	Vi-PS	non diagnosed	no	5	no	D7.responder	13	31.38102442	1.496667117
13	7	Vi-TCV	non diagnosed	late_shedder	5	yes	D7.responder	2	#N/A	#N/A
14	6	Vi-TCV	non diagnosed	no	5	no	D7.responder	1	210.9175	2.324112615
15	7	Vi-PS	non diagnosed	no	2	no	D7.responder	2	#N/A	#N/A
16	5	Vi-PS	diagnosed	no	3	yes	NA	5	#N/A	#N/A
17	3	control	diagnosed	no	4	yes	no	1		
18	3	control	diagnosed	no	#N/A	yes	no	8		
19	4	control	diagnosed	no	#N/A	yes	no	5		
20	4	control	non diagnosed	no	6	no	no	2		
21	5	Vi-TCV	non diagnosed	no	6	yes	D10.responder	3	#N/A	#N/A
22	4	control	non diagnosed	no	1	no	no	6		
23	5	Vi-TCV	diagnosed	no	5	yes	D7.responder	7	#N/A	#N/A
24	3	Vi-TCV	diagnosed	no	4	yes	D10.responder	0	#N/A	#N/A
25	6	Vi-TCV	diagnosed	late_shedder	4	yes	D10.responder	19	#N/A	#N/A
26	3	control	diagnosed	no	2	yes	no	18		
27	7	Vi-PS	diagnosed	no	2	yes	D7.responder	8	#N/A	#N/A
28	2	**control**	diagnosed	no	#N/A	yes	no	14		
29	4	control	diagnosed	no	3	yes	no	7		
30	6	Vi-TCV	diagnosed	no	2	yes	D10.responder	11	5.944253859	0.774097348
31	6	Vi-PS	non diagnosed	no	6	no	D7.responder	2	40.809	1.610755953
32	3	Vi-TCV	diagnosed	no	1	yes	NA	9	201.685	2.304673599
33	7	Vi-PS	non diagnosed	late_shedder	6	yes	D7.responder	9	#N/A	#N/A
34	5	Vi-PS	non diagnosed	late_shedder	5	no	non.responder	6	#N/A	#N/A
35	3	Vi-PS	non diagnosed	no	6	yes	D7.responder	7	2.3145	0.364457185
36	5	Vi-TCV	diagnosed	no	4	yes	non.responder	5	9.974	0.998869364
37	3	control	diagnosed	no	#N/A	yes	no	11		
38	2	control	diagnosed	no	2	yes	no	5		
39	6	Vi-PS	non diagnosed	no	5	no	NA	2	#N/A	#N/A
40	1	Vi-TCV	non diagnosed	no	2	no	D7.responder	6	#N/A	#N/A
41	5	Vi-TCV	diagnosed	no	2	yes	NA	9	29.587	1.471100932
42	3	control	diagnosed	no	1	yes	no	19		
43	3	control	diagnosed	no	#N/A	yes	no	10		
44	3	control	non diagnosed	yes	4	yes	no	8		
45	5	Vi-PS	non diagnosed	no	4	no	non.responder	1	#N/A	#N/A
46	5	Vi-PS	diagnosed	no	0	yes	NA	8	#N/A	#N/A
47	6	Vi-TCV	non diagnosed	no	4	no	D7.responder	2	17.75375375	1.249290192
48	4	Vi-TCV	non diagnosed	no	5	no	D10.responder	3	#N/A	#N/A
49	3	Vi-PS	non diagnosed	no	4	no	non.responder	11	31.861	1.503259403
50	5	Vi-PS	non diagnosed	late_shedder	6	yes	D10.responder	1	#N/A	#N/A
51	6	Vi-TCV	non diagnosed	no	5	no	D7.responder	4	#N/A	#N/A
52	5	Vi-TCV	non diagnosed	no	3	no	D7.responder	3	#N/A	#N/A
53	3	control	diagnosed	no	1	yes	no	21		
54	6	Vi-TCV	non diagnosed	no	6	no	D10.responder	7	#N/A	#N/A
55	5	Vi-TCV	diagnosed	no	3	yes	D10.responder	16	#N/A	#N/A
56	6	Vi-PS	non diagnosed	no	5	no	D7.responder	1	#N/A	#N/A
57	5	Vi-PS	diagnosed	late_shedder	4	yes	D7.responder	16	74.78	1.873785461
58	3	control	diagnosed	yes	1	yes	no	#N/A		
59	6	Vi-TCV	non diagnosed	no	3	yes	D7.responder	5	19.668	1.2937602
60	2	Vi-TCV	non diagnosed	late_shedder	6	yes	NA	7	194.5	2.288919606
61	4	control	non diagnosed	yes	6	no	no	0		
62	6	Vi-TCV	diagnosed	no	3	yes	D7.responder	8	#N/A	#N/A
63	6	Vi-PS	diagnosed	no	1	yes	non.responder	12	17.751	1.249222824
64	3	Vi-PS	non diagnosed	no	6	no	non.responder	3	#N/A	#N/A
65	5	Vi-TCV	non diagnosed	no	5	yes	D7.responder	3	#N/A	#N/A
66	6	Vi-PS	diagnosed	no	3	yes	non.responder	9	7.037	0.847387551
67	2	Vi-TCV	non diagnosed	no	6	yes	NA	3		
68	6	Vi-TCV	non diagnosed	no	5	no	D7.responder	3	15.768	1.197776611
69	6	Vi-PS	non diagnosed	no	6	no	D7.responder	1	72.3425	1.859393513
70	3	control	diagnosed	no	#N/A	yes	no	6		
71	4	control	diagnosed	no	2	yes	no	8		
72	7	Vi-TCV	non diagnosed	no	4	yes	D10.responder	7	#N/A	#N/A
73	3	control	diagnosed	no	3	yes	no	8		
74	4	control	diagnosed	no	#N/A	yes	no	8		
75	5	Vi-TCV	non diagnosed	no	5	no	D7.responder	4	#N/A	#N/A
76	6	Vi-PS	non diagnosed	late_shedder	6	yes	D7.responder	2	31.4475	1.497586126
77	4	Vi-TCV	non diagnosed	no	4	no	D7.responder	0	#N/A	#N/A
78	5	Vi-PS	non diagnosed	no	4	no	non.responder	3	2	0.301029996
79	3	Control	diagnosed	no	4	yes	N/A	5	1	0
80	6	Vi-TCV	diagnosed	late_shedder	4	yes	D7.responder	8	74.3125	1.871061872
81	4	Vi-TCV	non diagnosed	no	3	no	NA	1	#N/A	#N/A
82	2	control	non diagnosed	no	6	no	no	2		
83	5	Vi-PS	non diagnosed	no	4	no	D10.responder	0	160.075	2.20432351
84	6	Vi-TCV	non diagnosed	no	6	no	D7.responder	5	89.54	1.95201709
85	6	Vi-PS	non diagnosed	late_shedder	4	yes	D7.responder	1	64.052	1.806532695
86	3	control	diagnosed	no	4	yes	no	15		
87	2	Vi-PS	non diagnosed	no	5	no	D7.responder	14		
88	6	Vi-TCV	non diagnosed	no	4	no	D10.responder	4	76.1975	1.881940723
89	6	Vi-PS	diagnosed	no	1	yes	non.responder	6	#N/A	#N/A
90	6	Vi-TCV	non diagnosed	late_shedder	3	yes	D7.responder	4	56.85355392	1.754757618
91	6	Vi-TCV	diagnosed	no	4	yes	D7.responder	1	#N/A	#N/A
92	6	Vi-PS	diagnosed	late_shedder	6	yes	D10.responder	3	#N/A	#N/A
93	7	Vi-PS	diagnosed	late_shedder	2	yes	non.responder	6	#N/A	#N/A
94	6	Vi-PS	non diagnosed	no	4	no	D7.responder	3	#N/A	#N/A
95	4	control	diagnosed	no	2	yes	no	9		
96	6	Vi-PS	diagnosed	no	0	yes	non.responder	15	#N/A	#N/A
97	3	control	diagnosed	no	6	yes	no	12		
98	7	Vi-PS	diagnosed	no	2	yes	non.responder	2	#N/A	#N/A
99	1	control	diagnosed	yes	5	yes	no	4		
100	2	control	diagnosed	no	1	yes	no	15		

9.5.2.1 New frontiers in the diagnosis of S. Typhi infection—host-based markers of infection

Novel biomarkers of typhoid infection are being sought using omics technologies. There is evidence to suggest this as a viable approach in principle. Näsström et al. used a host metabolomics approach to identify signatures that could differentiate blood culture-confirmed typhoid fever patients from healthy controls and a febrile non-typhoidal group in patient cohorts from Asia and Africa[451]. Unfortunately, the authors did not report the findings in a way that allowed evaluation of clinical utility–sensitivity, specificity and misclassification statistics were missing, for example.

Blohmke et al. used a transcriptomic approach to distinguish Typhoid from non-typhoidal illnesses. They integrated publicly available data and their typhoid datasets to identify a blood transcriptomic profile that differentiated typhoid from three other causes of febrile illness (malaria, dengue and TB)[452]. A supervised machine learning algorithm (random forest) identified a 5-gene transcriptomic signature that differentiated typhoid from the other febrile illnesses. The signature comprising of *STAT1*, *SLAMF8*, *PSME2*, *WARS* and *ALDH1A1* worked well on the validation cohort and correctly identified enteric fever with a sensitivity and specificity of 97.1 and 88.0%, respectively (area under the receiver operating characteristic curve, AUROC: 96.7%).

9.5.2.1.1 Catch 22—the inherent limitations of typhoid diagnostic field studies and the promise of challenge studies

Blohmke et al. and Näsström et al. used their signatures to infer the typhoid status of clinically suspected, culture-negative typhoid cases. The clinically suspected, culture-negative typhoid group is a target group for novel diagnostics, but it is an inherently problematic group, too. No satisfactory reference standard diagnostic test exists, so it is challenging to know how well the signatures worked in these studies. Patients with clinically suspected, culture-negative typhoid are, by their very nature, missing from the classification models, so how can we be sure that the models will perform well in this group? They could, for example, represent an immunologically distinct cohort. Challenge studies offer an ideal way to circumvent this dilemma because regardless of the blood culture results, a person with a fever exposed to *S.* Typhi 7 days earlier almost certainly has typhoid regardless of their blood parameters, and clinical picture and blood culture results. This means clinically suspected, culture-negative typhoid group can be identified as having typhoid with certainty (which is not possible in field studies), and thus, signatures can be derived that work for this group—the same group a signature is most needed for. Further work on host biomarkers of typhoid infection is needed—identifying cases of typhoid that conventional diagnostics would miss could allow more accurate epidemiological data, guide the appropriate antibiotic prescription and provide a more accurate outcome measure in vaccine effectiveness studies.

9.5.2.1.2 Preventing typhoid

Typhoid persists in water systems making it difficult to eradicate[560]. *S.* Typhi is transmitted via the faeco-oral route by consuming contaminated water and food or through inadequate hygiene and sanitation. Thus, the mainstay of prevention is WASH interventions—safe water, sanitation, and hygiene. Unfortunately, these simple-sounding interventions can take a long time to implement. Efforts to improve sanitation infrastructure are undermined by limited funding, political instability, conflict, poverty and natural disasters. Typhoid vaccinations offer an adjunct to WASH approaches and have the advantage that the protection they confer persists within vaccinated populations in times of displacement, i.e. it is a portable prevention method. Furthermore, vaccines protect overseas travellers; helping countries reduce the chance of importing typhoid when they return.

9.5.2.1.3 Typhoid vaccine development

Despite the success of Vi-TCV and another Vi conjugate vaccine (Vi-CRM197) in gaining WHO pre-qualification, more typhoid vaccines are needed to ensure the global demand for typhoid vaccines can be met. In addition, a plurality of vaccine options creates a competitive pricing environment that ensures good value for purchasers and protects security in the procurement system—if the supply chain of one vaccine breaks down other options are available.

Two issues make typhoid trials difficult. One, as discussed above, the difficulty in diagnosing typhoid creates noise in the data, which means larger sample sizes are needed, and it lengthens the time to completion because investigators have to wait to accrue sufficient blood culture positive diagnoses in order to meet the trial endpoint. A novel method of inferring typhoid vaccine efficacy has been

developed based on seroefficacy data[561]. This method uses Vi titres to infer infection with a boost in antibodies over time, indicating typhoid infection. It has shown some success but is at odds with the findings of controlled human challenge models of typhoid, which found typhoid infection does not increase Vi-titres[562,563]. New methods of typhoid diagnosis are required.

Typhoid vaccine development is also hindered by the lack of a correlate of protection and biomarkers of an effective vaccine response (see introduction section 1.2.1.4.1 on correlates of protection in vaccinology). As previously noted, many vaccines fail in development, sometimes after costly phase 2 trials, due to a lack of informed knowledge about the best parameters for an effective and safe vaccine. The lack of correlates slows the licensure of new typhoid vaccines because efficacy cannot be gauged from relatively small short immunogenicity trials. Instead, intensive human challenge studies are needed, or large, expensive field trials are needed involving thousands of participants and take years to complete.

Correlates are also helpful because the duration of protection can be assessed, which can inform decisions regarding the requirement/timing of boosters. The most obvious candidate correlate of protection is anti-Vi IgG titres. Suggestions for Vi-IgG threshold have been proposed by some vaccine studies but have not been widely adopted partly because these thresholds have not translated into a globally accredited assay [564].

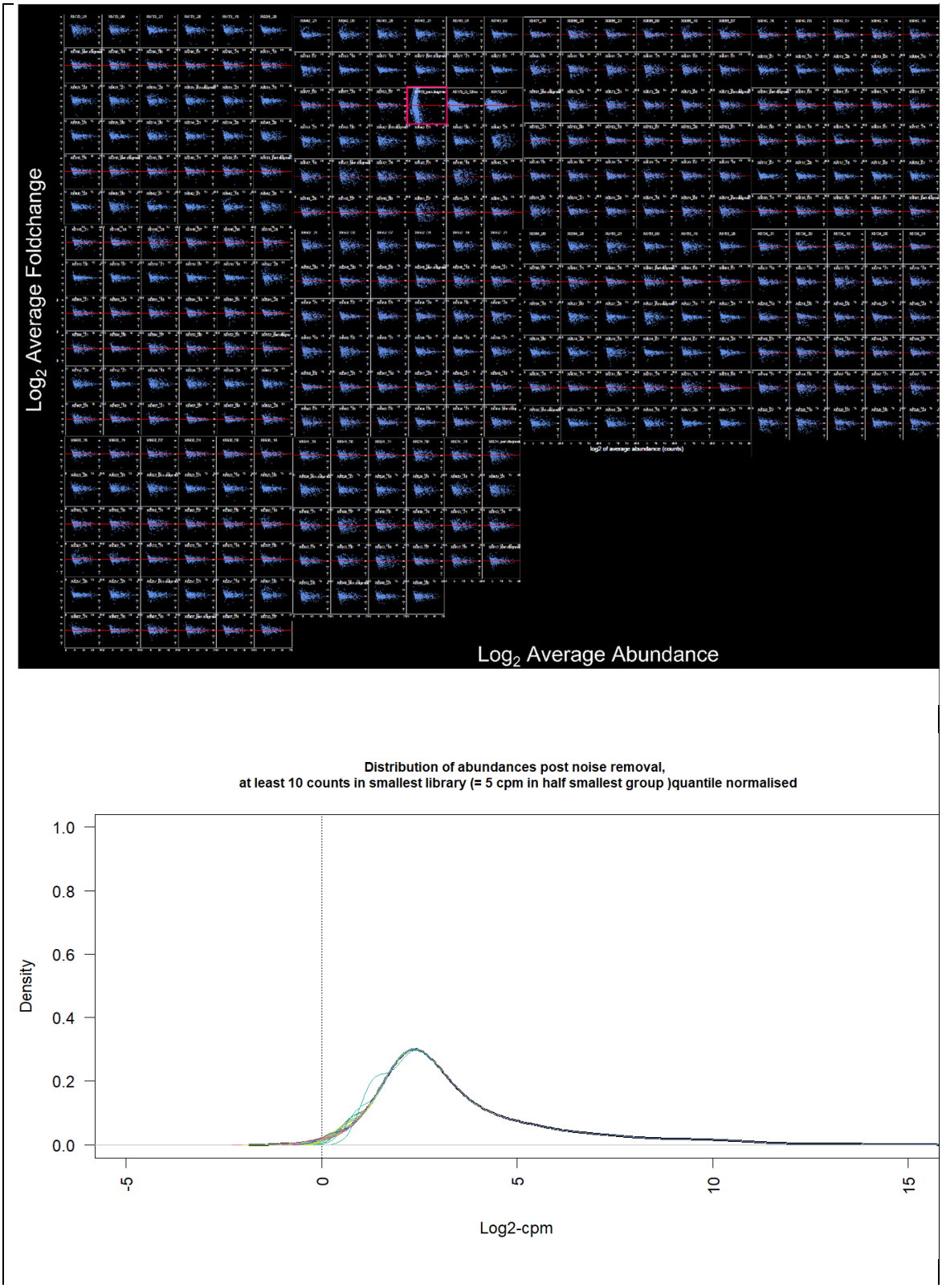
9.5.2.1.4 Emerging antibiotic resistance

Antibiotic resistance is a major concern, particularly with the spread of the H58 lineage of *S.* Typhi, which is associated with multi-drug resistance (MDR). In 2016, there was an outbreak of extensively drug-resistant (XDR) typhoid in Pakistan that was resistant to ciprofloxacin, ceftriaxone, amoxicillin, chloramphenicol, and trimethoprim-sulphamethoxazole[454]. Overall, the greatest hope for overcoming the threat posed by XDR *S.* Typhi is to reduce disease burden through hygiene interventions and vaccination. In addition, novel treatments that switch the host response towards a more favourable outcome would be beneficial, as they may limit disease morbidity and mortality without exerting the selective pressure conferred by antibiotics and be an option for treatment in XDR *S.* Typhi. However, a better understanding of the host response to S. Typhi is needed to achieve this

.

Appendix 5.1

9.5.3 QC and normalisation of sRNA data

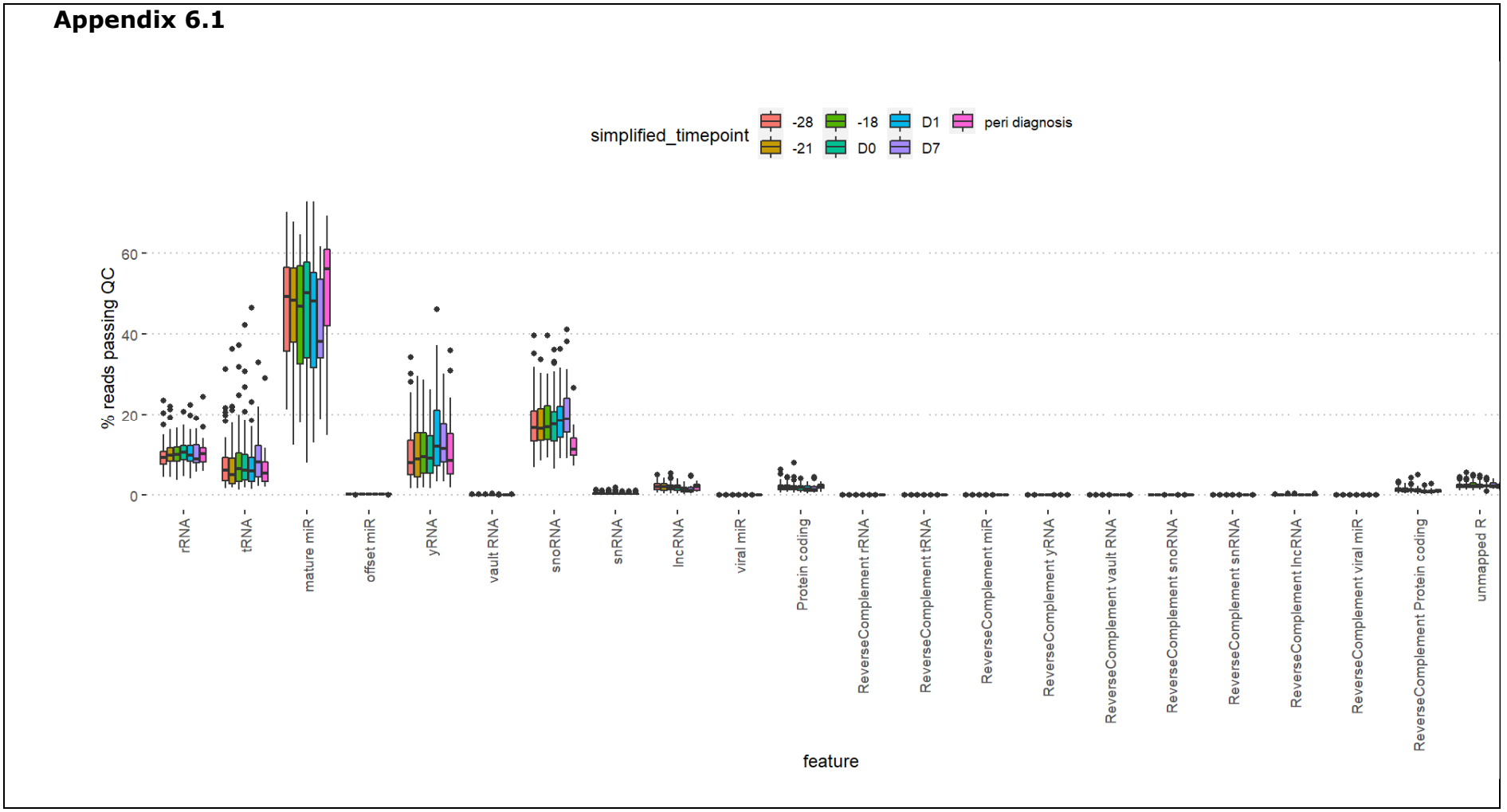


Appendix Figure 30 QC plots of sRNA-seq in VAST.

Upper panel: MA plots. Highlighted in pink is an example of an MA plot of an outlier sample. This sample was removed in the final analysis.

Lower panel: Density plot showing distribution plot of counts across abundances after removal of noise, outlier samples and normalisation. Plot shows good alignment of feature count distributions across the samples, with small and acceptable deviations.

Appendix 6.1



Appendix Figure 31 Box plots of sRNA class distribution after noise removal and normalisation (all sRNA classes together) and broken down by time point. There was a significant reduction in sRNA-dfs from snoRNAs at typhoid diagnosis. This mirrors an increase in miRNAs. miRNAs were therefore *normalised and analysed separately to other sRNA classes.*

9.5.1 Differentially expressed sRNA classes after *Vaccination*

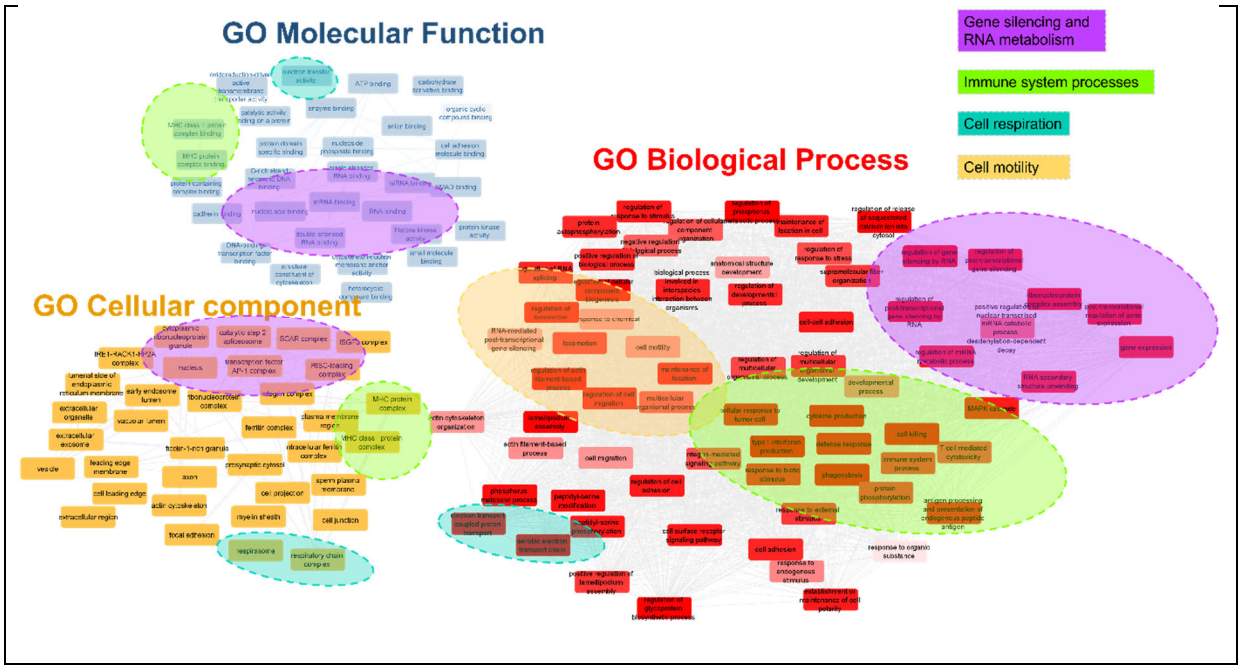
Supplementary Table 6 Number of DE features per sRNA class for each analysis post vaccination. Colour intensity relates to number of DE features. tRNAs were collapsed to anti-codon level. Numbers in brackets represents number of unique features analysed.

Theme	analysis	RNA class							
		mature miRNA (310)	lncRNA (858)	protein coding (723)	offset miRNA (66)	Vault (3)	snoRNA (327)	tRNA (81)	antisense to protein coding (331)
Q1. Does Vaccination alter miRNA expression? (i.e. timepoint vs timepoint)	baseline vs 7 days post vaccine	0	1	4	0	0	0	0	0
	baseline vs 10 days post vaccine	0	0	0	0	0	0	0	0
	baseline vs 28 days post vaccine	4	1	2	0	0	0	0	0
	7 days post vaccine vs 10 days post vaccine	0	0	0	0	0	0	0	0
	7 days post vaccine vs 28 days post vaccine	0	0	3	1	0	0	0	0
	10 days post vaccine vs 28 days post vaccine	0	3	3	0	0	0	0	0
	baseline (remained well) vs 7 days post vaccines (remained well)	0	0	0	0	0	0	0	0
	baseline (remained well) vs 10 days post vaccine (remained well)	0	0	0	0	0	0	0	0
	baseline (remained well) vs 28 days post vaccine (remained well)	1	0	0	0	0	0	0	0
	7 days post vaccine (remained well) vs 10 days post vaccine (remained well)	0	0	0	0	0	0	0	0
	7 days post vaccine (remained well) vs 28 days post vaccine (remained well)	0	0	0	0	0	0	0	0
	10 days post vaccine (remained well) vs 28 days post vaccine (remained well)	1	0	0	0	0	0	0	0
	baseline (developed typhoid) vs 7 days post vaccines (developed typhoid)	0	0	0	0	0	0	0	0
	baseline (developed typhoid) vs 10 days post vaccine (developed typhoid)	0	0	1	0	0	0	0	0
	baseline (developed typhoid) vs 28 days post vaccine (developed typhoid)	1	0	0	0	0	0	0	0
	7 days post vaccine (developed typhoid) vs 10 days post vaccine (developed typhoid)	0	0	0	0	0	0	0	0
	7 days post vaccine (developed typhoid) vs 28 days post vaccine (developed typhoid)	0	0	0	0	0	0	0	0
	10 days post vaccine (developed typhoid) vs 28 days post vaccine (developed typhoid)	0	0	1	0	0	0	0	0
	baseline Vi-TCV vs 7 days post vaccines Vi-TCV	0	1	5	0	1	0	0	0
	baseline Vi-TCV vs 10 days post vaccine Vi-TCV	0	0	0	0	0	0	0	0
	baseline Vi-TCV vs 28 days post vaccine Vi-TCV	1	40	10	3	0	12	2	4
	7 days post vaccine Vi-TCV vs 10 days post vaccine Vi-TCV	0	0	1	0	0	0	0	0
	7 days post vaccine Vi-TCV vs 28 days post vaccine Vi-TCV	1	0	4	0	0	0	0	0
	10 days post vaccine Vi-TCV vs 28 days post vaccine Vi-TCV	0	0	1	0	0	0	0	0
	baseline Vi-PS vs 7 days post vaccines Vi-PS	0	0	0	0	0	0	0	0

	baseline Vi-PS vs 10 days post vaccine Vi-PS	0	0	0	0	0	0	0	0
	baseline Vi-PS vs 28 days post vaccine Vi-PS	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS vs 10 days post vaccine Vi-PS	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS vs 28 days post vaccine Vi-PS	0	0	0	0	0	0	0	0
	10 days post vaccine Vi-PS vs 28 days post vaccine Vi-PS	0	0	0	0	0	0	0	0
	baseline Vi-PS (developed typhoid) vs 7 days post vaccine Vi-PS (developed typhoid)	0	0	0	0	0	0	0	0
	baseline Vi-PS (developed typhoid) vs 10 days post vaccine Vi-PS (developed typhoid)	0	0	0	0	0	0	0	0
	baseline Vi-PS (developed typhoid) vs 28 days post vaccine Vi-PS (developed typhoid)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS (developed typhoid) vs 10 days post vaccine Vi-PS (developed typhoid)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS (developed typhoid) vs 28 days post vaccine Vi-PS (developed typhoid)	0	0	0	0	0	0	0	0
	10 days post vaccine Vi-PS (developed typhoid) vs 28 days post vaccine Vi-PS (developed typhoid)	0	0	0	0	0	0	0	0
	baseline Vi-PS (remained well) vs 7 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	baseline Vi-PS (remained well) vs 10 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	baseline Vi-PS (remained well) vs 28 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS (remained well) vs 10 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS (remained well) vs 28 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	10 days post vaccine Vi-PS (remained well) vs 28 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	baseline Vi-TCV (remained well) vs 7 days post vaccine Vi-TCV (remained well)	0	2	5	0	1	0	0	0
	baseline Vi-TCV (remained well) vs 10 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	baseline Vi-TCV (remained well) vs 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-TCV (remained well) vs 10 days post vaccine Vi-TCV (remained well)	0	0	1	0	0	0	0	0
	7 days post vaccine Vi-TCV (remained well) vs 28 days post vaccine Vi-TCV (remained well)	1	0	1	0	0	0	0	0
	10 days post vaccine Vi-TCV (remained well) vs 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	baseline Vi-TCV (developed typhoid) vs 7 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	baseline Vi-TCV (developed typhoid) vs 10 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	baseline Vi-TCV (developed typhoid) vs 28 days post vaccine Vi-TCV (developed typhoid)	0	0	0	1	0	0	0	0
	7 days post vaccine Vi-TCV (developed typhoid) vs 10 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-TCV (developed typhoid) vs 28 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	10 days post vaccine Vi-TCV (developed typhoid) vs 28 days post vaccine Vi-TCV (developed typhoid)	0	0	2	0	0	0	0	0
Q2. Do Vi-PS and Vi-TCV induce different miRNAs?	baseline Vi-PS vs baseline Vi-TCV	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS vs 7 days post vaccine Vi-TCV	0	1	4	0	1	0	0	0
	10 days post vaccine Vi-PS vs 10 days post vaccine Vi-TCV	0	0	0	0	0	0	0	0
	28 days post vaccine Vi-PS vs 28 days post vaccine Vi-TCV	0	0	0	0	0	0	0	0
	baseline Vi-PS (developed typhoid) vs baseline Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0

	10 days post vaccine Vi-PS (developed typhoid) vs 10 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	28 days post vaccine Vi-PS (developed typhoid) vs 28 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS (developed typhoid) vs 7 days post vaccine Vi-TCV (developed typhoid)	1	0	0	0	0	0	0	0
	baseline Vi-PS (remained well) vs baseline Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS (remained well) vs 7 days post vaccine Vi-TCV (remained well)	1	1	6	1	1	0	0	0
	10 days post vaccine Vi-PS (remained well) vs 10 days post vaccine Vi-TCV (remained well)	0	0	1	0	0	0	0	0
	28 days post vaccine Vi-PS (remained well) vs 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs Vi-TCV baseline 7 days post vaccine vs Vi-PS baseline 7 days post vaccine	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs Vi-TCV baseline 10 days post vaccine vs Vi-PS baseline 10 days post vaccine	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs Vi-TCV baseline 28 days post vaccine vs Vi-PS baseline 28 days post vaccine	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs Vi-TCV 7 days post vaccine 10 days post vaccine vs Vi-PS 7 days post vaccines 10 days post vaccine	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs Vi-TCV 7 days post vaccine 28 days post vaccine vs Vi-PS 7 days post vaccine 28 days post vaccine	0	0	2	0	0	0	0	0
	difference of the $\log_2$ FCs Vi-TCV 10 days post vaccine 28 days post vaccine vs Vi-PS 10 days post vaccine 28 days post vaccine	0	0	1	0	0	0	0	0
	difference of the $\log_2$ FCs baseline 7 days post vaccine Vi-PS (developed typhoid) vs baseline 7 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs baseline 7 days post vaccine Vi-PS (remained well) vs baseline 7 days post vaccine Vi-TCV (remained well)	0	0	1	0	0	0	0	0
	difference of the $\log_2$ FCs baseline 10 days post vaccine Vi-PS (developed typhoid) vs baseline 10 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs baseline 10 days post vaccine Vi-PS (remained well) vs baseline 10 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs baseline 28 days post vaccine Vi-PS (developed typhoid) vs baseline 28 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs baseline 28 days post vaccine Vi-PS (remained well) vs baseline 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs 7 days post vaccine 10 days post vaccine Vi-PS (developed typhoid) vs 7 days post vaccine 10 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs 7 days post vaccine 10 days post vaccine Vi-PS (remained well) vs 7 days post vaccine 10 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs 7 days post vaccine 28 days post vaccine Vi-PS (developed typhoid) vs 7 days post vaccine 28 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs 7 days post vaccine 28 days post vaccine Vi-PS (remained well) vs 7 days post vaccine 28 days post vaccine Vi-TCV (remained well)	0	0	1	0	0	0	0	0
	difference of the $\log_2$ FCs 10 days post vaccine 28 days post vaccine Vi-PS (developed typhoid) vs 10 days post vaccine 28 days post vaccine Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs 10 days post vaccine 28 days post vaccine Vi-PS (remained well) vs 10 days post vaccine 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
Q3. Are there any correlates of protection? (i.e. Diagnosed vs non diagnosed)	baseline (developed typhoid) vs baseline (remained well)	0	0	0	0	0	0	0	0
	7 days post vaccine (developed typhoid) vs 7 days post vaccine (remained well)	0	0	0	0	0	0	0	0
	10 days post vaccine (developed typhoid) vs 10 days post vaccine (remained well)	40	2	2	0	0	1	4	2
	28 days post vaccine (developed typhoid) vs 28 days post vaccine (remained well)	0	0	0	0	0	0	0	0
	difference of the $\log_2$ FCs (remained well) baseline 7 days post vaccine vs (developed typhoid) baseline 7 days post vaccine	0	0	0	0	0	0	0	0

	difference of the log_2 FCs (remained well) baseline 10 days post vaccine vs (developed typhoid) baseline 10 days post vaccine	0	0	0	0	0	0	0	0
	difference of the log_2 FCs (remained well) baseline 28 days post vaccine vs (developed typhoid) baseline 28 days post vaccine	0	0	0	0	0	0	0	0
	difference of the log_2 FCs (remained well) 7 days post vaccine 10 days post vaccine vs (developed typhoid) 7 days post vaccine 18	0	0	0	0	0	0	0	0
	difference of the log_2 FCs (remained well) 7 days post vaccine 28 days post vaccine vs (developed typhoid) 7 days post vaccines 28 days post vaccine	0	0	0	0	0	0	0	0
	difference of the log_2 FCs (remained well) 10 days post vaccine 28 days post vaccine vs (developed typhoid) 10 days post vaccine 28 days post vaccine	0	0	0	0	0	0	0	0
	baseline Vi-PS (developed typhoid) vs baseline Vi-PS (remained well)	0	0	0	0	0	0	0	0
	baseline Vi-TCV (developed typhoid) vs baseline Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-PS (developed typhoid) vs 7 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	7 days post vaccine Vi-TCV (developed typhoid) vs 7 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	10 days post vaccine Vi-PS (developed typhoid) vs 10 days post vaccine Vi-PS (remained well)	0	1	0	0	0	4	1	1
	10 days post vaccine Vi-TCV (developed typhoid) vs 10 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	28 days post vaccine Vi-PS (developed typhoid) vs 28 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	28 days post vaccine Vi-TCV (developed typhoid) vs 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs baseline 7 days post vaccine Vi-PS (developed typhoid) vs baseline 7 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs baseline 7 days post vaccine Vi-TCV (developed typhoid) vs baseline 7 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs baseline 10 days post vaccine Vi-PS (developed typhoid) vs baseline 10 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs baseline 10 days post vaccine Vi-TCV (developed typhoid) vs baseline 10 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs baseline 28 days post vaccine Vi-PS (developed typhoid) vs baseline 28 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs baseline 28 days post vaccine Vi-TCV (developed typhoid) vs baseline 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs 7 days post vaccine 10 days post vaccine Vi-PS (developed typhoid) vs 7 days post vaccine 10 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	1
	difference of the log_2 FCs 7 days post vaccine 10 days post vaccine Vi-TCV (developed typhoid) vs 7 days post vaccine 10 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs 7 days post vaccine 28 days post vaccine Vi-PS (developed typhoid) vs 7 days post vaccine 28 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs 7 days post vaccine 28 days post vaccine Vi-TCV (developed typhoid) vs 7 days post vaccine 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs 10 days post vaccine 28 days post vaccine Vi-PS (developed typhoid) vs 10 days post vaccine 28 days post vaccine Vi-PS (remained well)	0	0	0	0	0	0	0	0
	difference of the log_2 FCs 10 days post vaccine 28 days post vaccine Vi-TCV (developed typhoid) vs 10 days post vaccine 28 days post vaccine Vi-TCV (remained well)	0	0	0	0	0	0	0	0



Appendix Figure 32 Enrichment of sRNA-dfs, using mRNAs detected in whole blood in VAST as background. Some themes have been picked out in coloured ellipses

9.5.2 Post-vaccination miRNA foldchanges in a publicly available dataset do not validate specific differential expression results but do correlate with post-Vi-vaccine foldchanges.

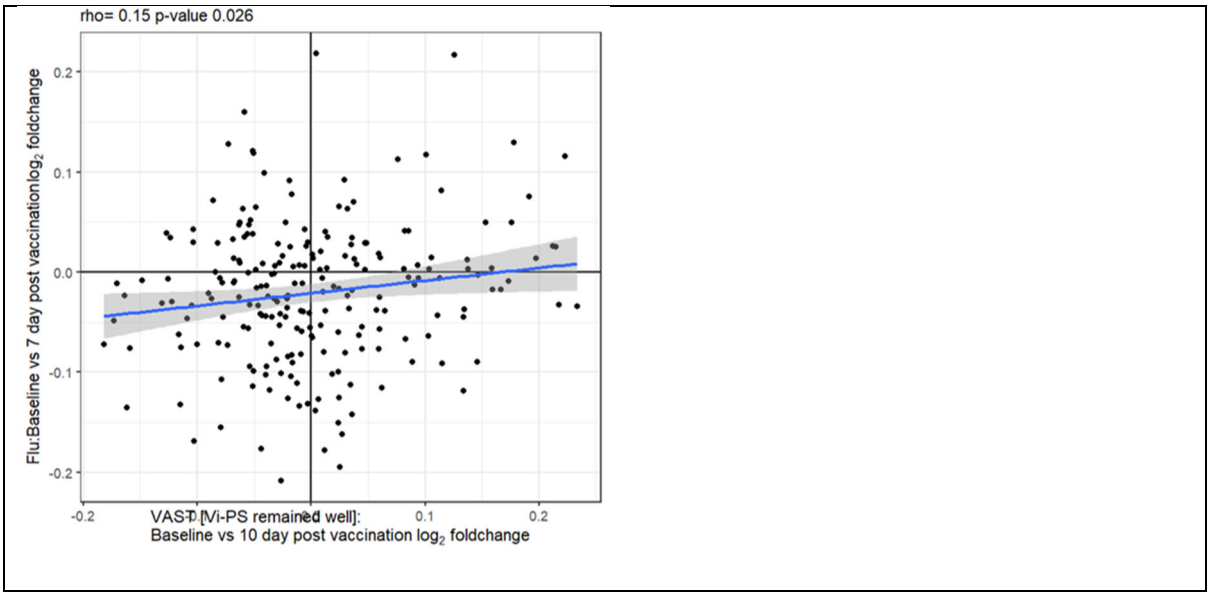
The handful of miRNAs DE longitudinally between time points is listed in Table 23. DE miRNAs were most apparent when comparing baseline and 28 days post-vaccination timepoints amongst all participants—this identified differential expression of four miRNAs: hsa-miR-92a-3p, hsa-miR-423-5p, hsa-miR-320a-3p and hsa-miR-1301-3p. There was no independent cohort in VAST to confirm

these findings; therefore, data from a published vaccine study was used to determine identify supporting evidence for these results.

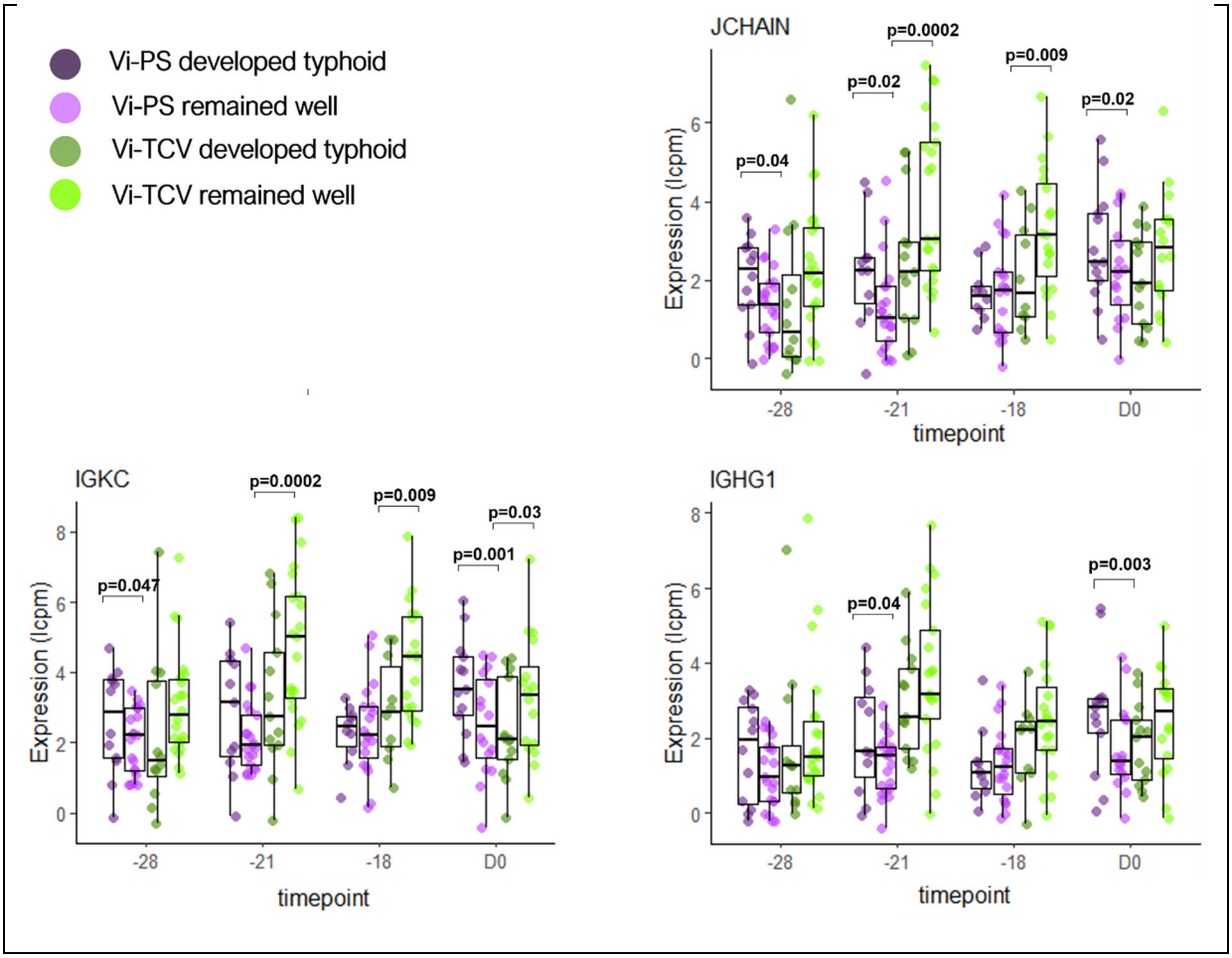
This data was generated in a flu vaccine study at baseline, 7 and 14 days after vaccination using a probe-based technology (HTG EdgeSeq miR technology). There are issues with translating results across different technologies, vaccines, and time points; nevertheless, it presented the best opportunity to interrogate DE miRNAs in an independent dataset.

Only baseline and 7 days post-vaccination time points were present in both the flu and VAST study. Thus, results from VAST were compared with results from the flu study using the nearest available time point.

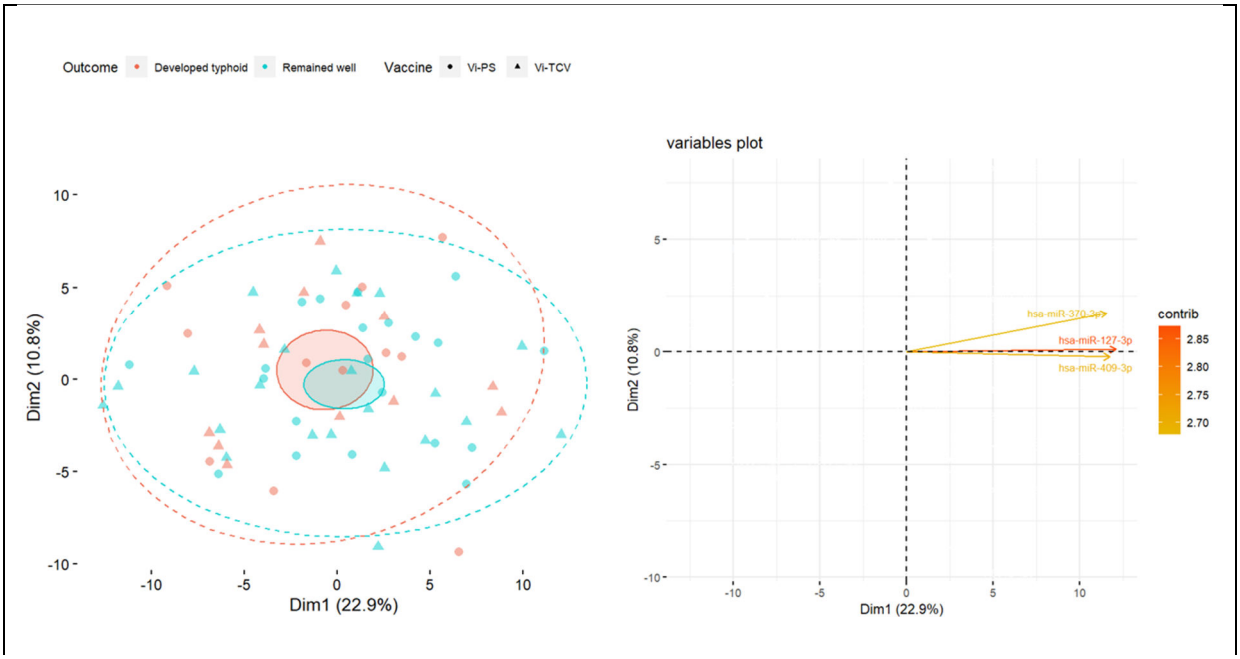
A reanalysis of the flu data did not show any miRNAs were DE longitudinally post-vaccination after multiple testing correction, despite reasonable sample sizes (~50 per time point). No miRNAs DE over time in VAST were validated as DE after flu vaccination. There were, however, weak but significant correlations between miRNA foldchanges, the best being baseline versus 10 days in VAST and baseline and 7 days in the flu study (Supplementary figure *4*). Interestingly, this correlation improved when the analysis was restricted to participants who did not develop typhoid (rho 0.18,p=0.008) and those who received Vi-PS (rho=0.2,p = 0.003). This possibly signals that vaccination induces changes in miRNA expression, and these may be more profound amongst people having a protective vaccine response, but that effect is weak and may vary between vaccine technologies.



Supplementary figure 4 Correlation in log_2 foldchange in miRNA expression after Vi and flu vaccination



Appendix Figure 33 Dot plots of sRNAs DE 7 days after vaccination in ViTCV group, broken down by vaccine type and outcome



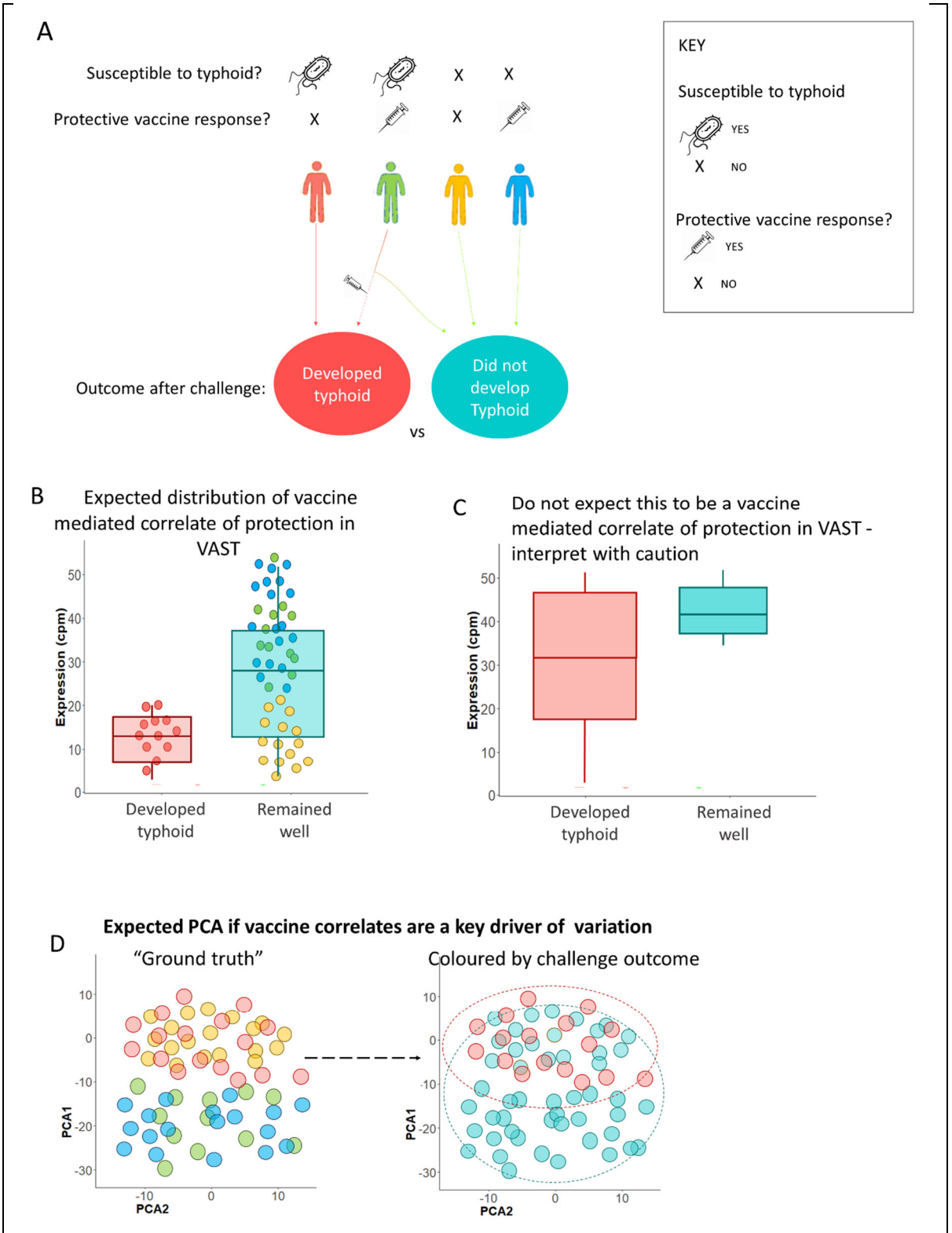
Appendix Figure 34 PCA of baseline samples showing no separation between outcome groups. Dotted lines show 95% confidence intervals of the data. Shaded ellipses represent 95% confidence intervals of each groups centroid

9.5.3 The challenge of identifying correlates of protection in VAST because of participants who are inherently resistant to typhoid.

Before exploring potential sRNA correlates of protection, it is important to explain the expression pattern expected of a correlate in VAST. As noted in the introduction, the challenge model is set up so that around 65% of non-vaccinated participants will develop typhoid. This means that 35% of participants have an inherent resistance to developing typhoid, which is an issue for identifying vaccine correlates as some of these participants had a non-protective response to vaccination (potentially around 50% based on vaccine effectiveness data) yet remained well after challenge (Appendix Figure 35 A - yellow participants). These inherently resistant participants cannot be easily identified because there are no markers of typhoid susceptibility or resistance. Certainty is only present for those who developed typhoid because they must be susceptible and have an ineffective vaccine response (Appendix Figure 35A - red participants). Inherently-resistant vaccine non-responders make it impossible to identify a perfect correlate through differential expression analyses (even if one existed); instead, a vaccine-mediated correlate in this study will have a predictable distribution of expression across outcome groups, as shown in Appendix Figure 35B. Any sRNAs that show a distribution seen in Appendix

Figure 35C are unlikely to be a vaccine-mediated correlate of protection. A similar prediction can be made for the expected output of a PCA Appendix Figure 35D&E.

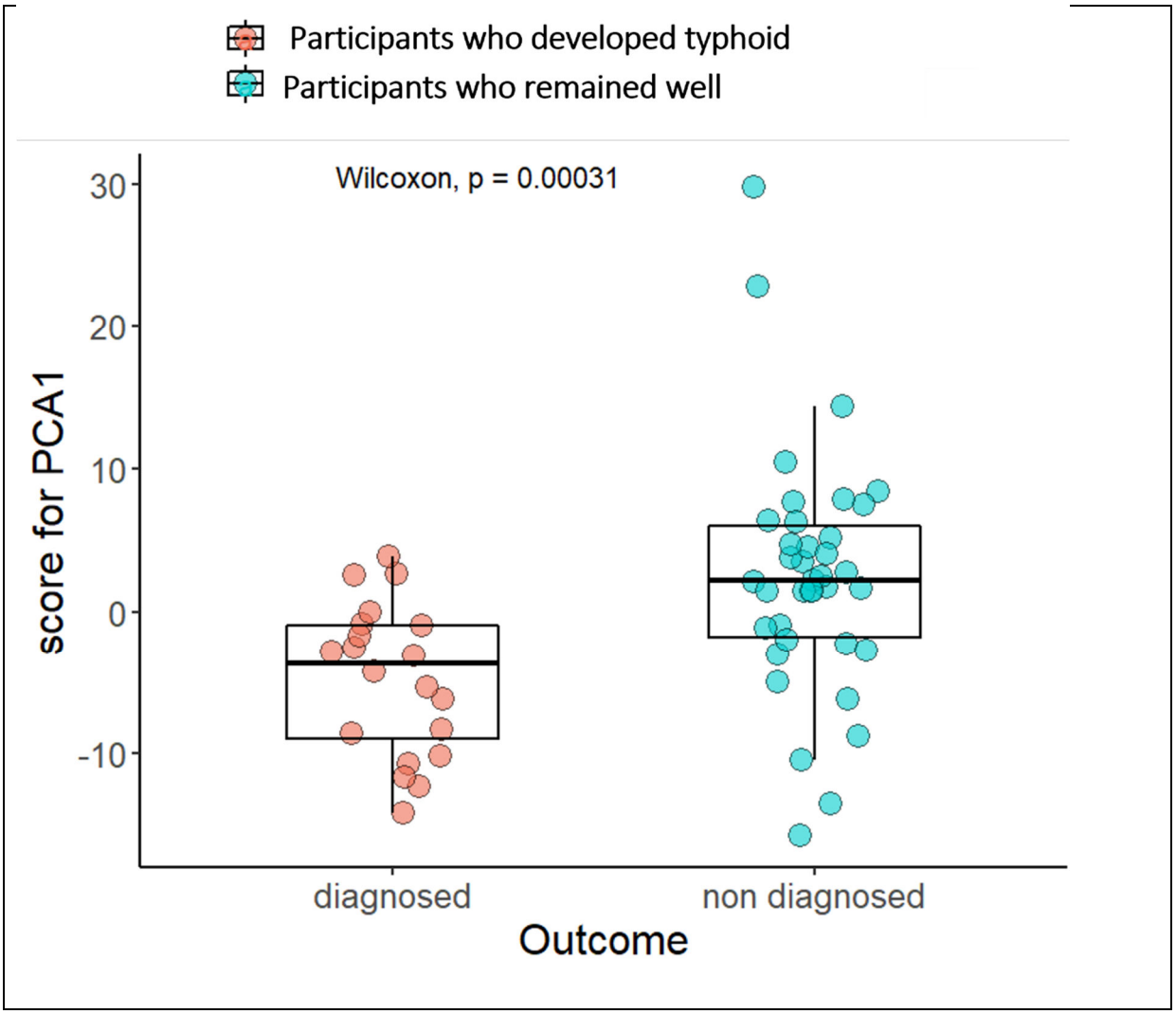
Models identifying vaccine correlates can be trained on sensitivity for diagnosis or specificity for protection. However, training a single model based on accuracy alone is likely to lead to overfitting unless it can pick out a signature that can predict both susceptibility to typhoid and effective vaccine response, e.g., based on a bagged tree algorithm.



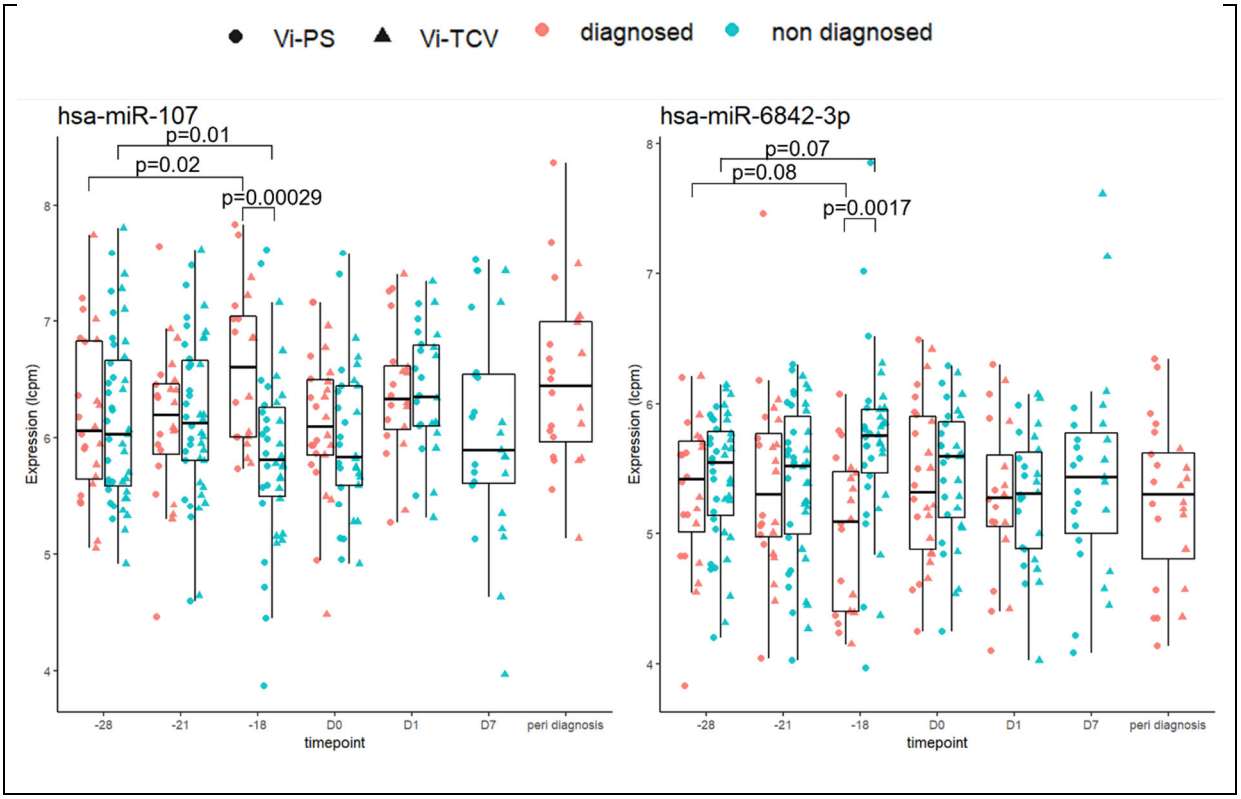
Appendix Figure 35 The correlates problem in VAST

A: The ground truth—a person's propensity for typhoid and how vaccination can interact

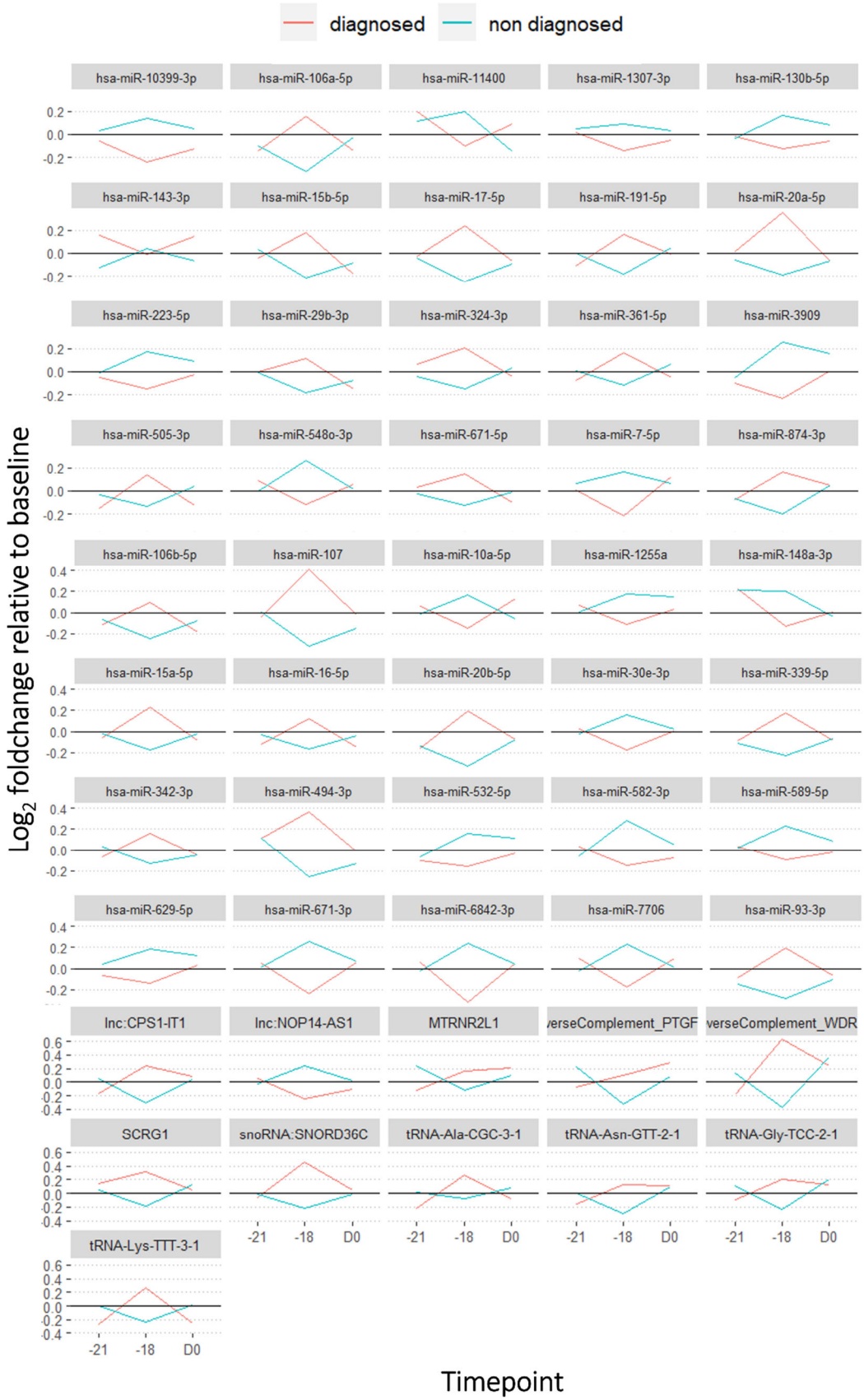
with this. Note the yellow person has a poor vaccine response but was not susceptible to typhoid—this group of people (whom cannot be identified from the data) make it harder to find vaccine correlates of protection. Based on an efficacy of around 54%, the yellow group will consist of 12 participants ~ 1/3 of people who were not diagnosed.
B. Prediction for the distribution of counts of a vaccine mediated correlate of protection in diagnosed and non-diagnosed participants. The samples at the bottom of the distribution in those that remain well would be from people who had an ineffective vaccine response but were protected from challenge due to innate resistance to typhoid.
C: A vaccine mediated correlate of protection is less likely to have this distribution of counts in diagnosed and non-diagnosed participants, because participants with a good vaccine response should not overlap with participants who developed typhoid.
D.A similar prediction can be made for the expected output of a PCA if sRNAs that correlate with vaccine mediated protection are source of a significant variation at a particular time point



Appendix Figure 36 Association between PC1 of the PCA in Figure 72 and outcome.



Appendix Figure 37 Two examples of miRNAs showing divergent patterns of expression at 10 days after vaccination between outcome groups

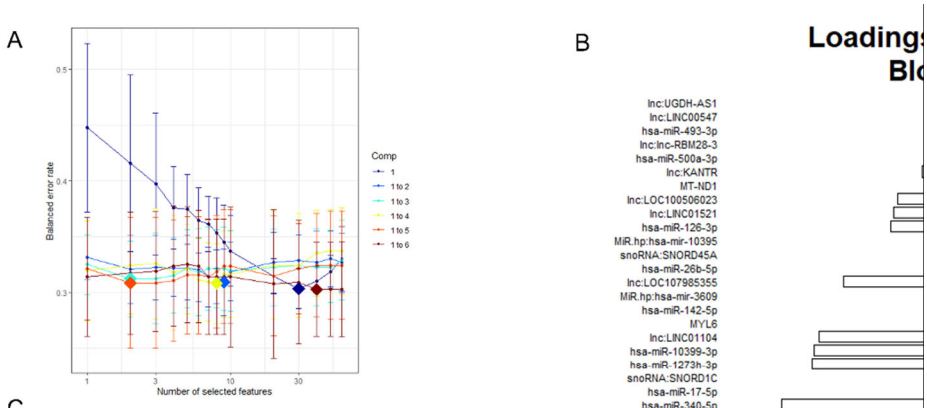

diagnosed
non diagnosed
hsa-miR-10399-3p
hsa-miR-106a-5p
hsa-miR-11400
hsa-miR-1307-3p
hsa-miR-130b-5p
hsa-miR-143-3p
hsa-miR-15b-5p
hsa-miR-17-5p
hsa-miR-191-5p
hsa-miR-20a-5p
hsa-miR-223-5p
hsa-miR-29b-3p
hsa-miR-324-3p
hsa-miR-361-5p
hsa-miR-3909
hsa-miR-505-3p
hsa-miR-548o-3p
hsa-miR-671-5p
hsa-miR-7-5p
hsa-miR-874-3p
hsa-miR-106b-5p
hsa-miR-107
hsa-miR-10a-5p
hsa-miR-1255a
hsa-miR-148a-3p
hsa-miR-15a-5p
hsa-miR-16-5p
hsa-miR-20b-5p
hsa-miR-30e-3p
hsa-miR-339-5p
hsa-miR-342-3p
hsa-miR-494-3p
hsa-miR-532-5p
hsa-miR-582-3p
hsa-miR-589-5p
hsa-miR-629-5p
hsa-miR-671-3p
hsa-miR-6842-3p
hsa-miR-7706
hsa-miR-93-3p
lnc:CPS1-IT1
lnc:NOP14-AS1
MTRNR2L1
verseComplement_PTGF
verseComplement_WDR
SCRG1
snoRNA:SNORD36C
tRNA-Ala-CGC-3-1
tRNA-Asn-GTT-2-1
tRNA-Gly-TCC-2-1
tRNA-Lys-TTT-3-1
-21
-18
D0
Log2 foldchange relative to baseline
Timepoint

Appendix Figure 38 Divergent log_2 foldchanges in responses 10 days post-vaccination between diagnosed and non-diagnosed participants. Note—foldchanges compared with baseline are modest in each group, which is why these sRNAs are not DE over time. Foldchanges at 7 days after vaccination (time point day -21) were more similar between the groups than at 10 days after vaccination (timepoint day -18).

9.5.1 sPLS-DA to identify predictors of vaccine outcome

A sparse PLS-DA model was built to determine whether sRNA expression 10 days after vaccination can predict outcome. sRNA expression was corrected for flow-cell prior to input to the model. Feature selection was performed using the 20 sRNAs with the highest loadings on the first 3 principle components of a PCA of 10-day post-vaccination samples. The samples were randomly split 80:20 into a training and test set. Model parameters were tuned to select the optimum number of components and features within those components. Tuning was done using M-fold cross-validation to minimise the balanced error rate (BER). Tuning found that the optimal model contained 1 component and 16 features. The performance of the sPLS-DA model was evaluated on the test data. The model performed well on the training data but not the test data, indicating overfitting. The number of features selected from the PCA was reduced but did not improve model performance on the test data. The inclusion of Vi-antibody titres in the model also did not improve the result.

Appendix 7.1



C

Training data

		Truth	
		Developed typhoid	Did not develop typhoid
predicted	Developed typhoid	11	6
	Did not develop typhoid	5	24

Test data

		Truth	
		Developed typhoid	Did not develop typhoid
predicted	Developed typhoid	3	3
	Did not develop typhoid	1	5

Appendix Figure 39 PLS-DA model for differentiating outcome groups based on sRNA expression at 10 days post vaccination.

A: Hyperparameter tuning, the optimum number of components was 1.

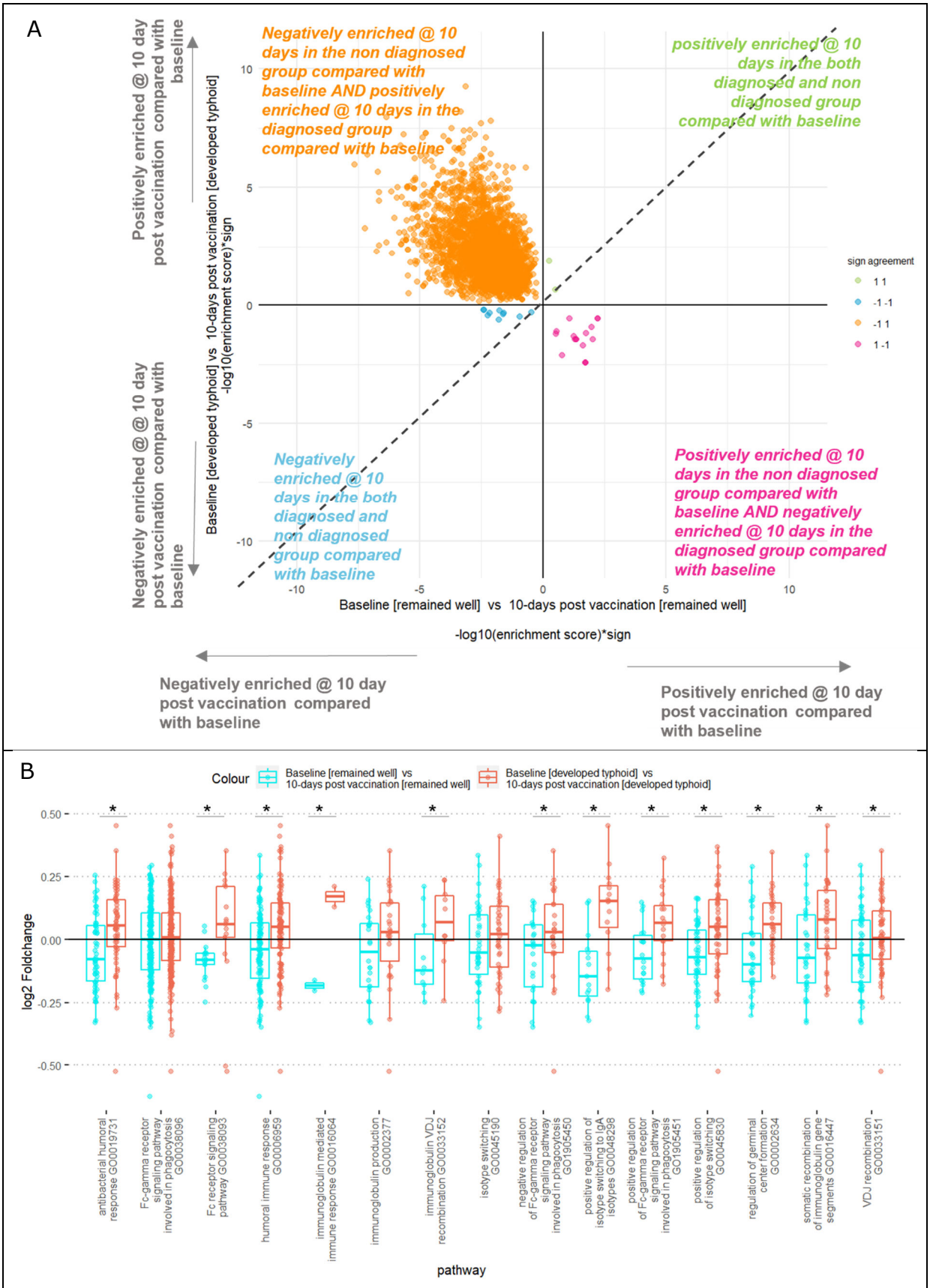
B: feature contribution to the model

C: Model performance

9.5.2 Establishing whether pathways are positively enriched for miRNA targeting by at 10 days after vaccination in diagnosed people or negatively enriched for targeting in people who did not develop typhoid.

MIEAA2 miRNA enrichment analysis showed that at 10 days after vaccination, most pathways were *positively enriched* for miRNA targeting in the diagnosed group compared with the non-diagnosed group. But enrichment analysis is

relative—positive enrichment in the diagnosed group could alternatively mean negative enrichment in the non-diagnosed group. Assessing enrichment with regards to baseline makes it possible to deduce, which scenario is more likely. Figure 75B shows an agreement plot comparing baseline vs 10 days post-vaccination enrichment results in diagnosed and non-diagnosed participants. This plot reveals an anti-correlation—most pathways were positively enriched 10 days after vaccination in the diagnosed group and negatively enriched in the non-diagnosed group. Visualising the $\log_2$ foldchanges of the miRNAs contained within differentially enriched pathways supports this conclusion. For example, Figure 75C shows miRNAs targeting antibody pathways were generally upregulated in those that developed typhoid, and downregulated in those that remained well (relative to baseline, Figure 75*C).* Median miRNA log2 foldchange for each pathway was always higher in those that developed typhoid (significantly so in most pathways * $p<0.05$). These results suggest changes in miRNA expression in PBMCs 10 days after vaccination leads to increased targeting of immune pathways in those that developed typhoid but the biological mechanism is not clear.



Supplementary figure 5 A: Agreement plot of enrichment results based on the results of the miRNA differential expression analysis between baseline and 10 days after

vaccination. Significance and direction of enrichment in the non-diagnosed group is presented on the x-axis. Significance and direction of enrichment in the diagnosed group is presented on the y-axis. An explanation of what how to interpret the results is included in coloured text in each quadrant. The diagonal dotted line is y=x.
B: log2 foldchanges from baseline vs day 10 post vaccination plotted for miRNAs that target immunoglobulin pathways in the orange quadrant of panel B., i.e., immunoglobulin pathways that were positively enriched for targeting in those that developed typhoid, and negatively enriched for targeting in those that remained well after challenge

9.5.3 Differentially expressed sRNA classes after *S.* Typhi challenge

Supplementary Table 7 Number of DE features per sRNA class for each analysis post challenge. Colour intensity relates to number of DE features. Numbers in brackets represents number of unique features analysed.

Theme	Analysis	RNA Class													
		mature miRNA (310)	lncRNA (858)	offset miRNA (66)	protein coding (723)	rRNA (7)	snoRNA (327)	snRNA (19)	tRNA (81)	Vault RNA (3)	yRNA (3)	antisense to lncRNA (24)	antisense to miRNA (1)	antisense to protein in coding (331)	antisense to rRNA (2)
Does S. Typhi alter sRNA expression? (ie timepoint vs timepoint)	D0 vs D1	130	239	14	315	2	59	4	15	1	1	13	1	127	1
	D7 vs typhoid (developed typhoid)nosis	133	41	8	181	1	115	0	9	1	0	1	0	20	1
	D7 Vi-PS vs typhoid (developed typhoid)nosis Vi-PS	92	15	7	104	0	28	0	1	0	0	1	0	8	1
	D7 Vi-TCV vs typhoid (developed typhoid)nosis Vi-TCV	73	11	1	82	0	39	0	2	1	0	0	0	5	0
	D0 (remained well) vs D1 (remained well)	100	180	17	236	1	39	4	12	1	0	12	1	106	0
	D0 (developed typhoid) vs D1 (developed typhoid)	0	1	0	1	0	0	0	2	0	0	0	0	1	0
	D0 vs typhoid diagnosis	158	76	14	302	1	72	1	11	0	0	4	0	49	1
	D0 vs D7	28	29	1	31	0	3	1	3	0	0	3	0	20	0
	D0 Vi-PS vs D1 Vi-PS	87	43	2	37	0	17	0	6	0	0	4	0	14	0
	D0 Vi-TCV vs D1 Vi-TCV	24	55	4	142	0	6	2	5	0	0	11	1	57	0
	D0 Vi-PS (developed typhoid) vs typhoid (developed typhoid)nosis Vi-PS (developed typhoid)	92	15	2	118	0	16	0	0	0	0	1	0	12	1
	D0 Vi-TCV (developed typhoid) vs D1 Vi-TCV (developed typhoid)	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	D0 Vi-TCV (developed typhoid) vs typhoid (developed typhoid)nosis Vi-TCV (developed typhoid)	85	2	2	112	0	21	0	2	0	0	1	0	9	0
	D0 Vi-PS (remained well) vs D1 Vi-PS (remained well)	32	12	1	5	0	1	0	0	0	0	1	0	2	0
	D0 Vi-PS (remained well) vs D7 Vi-PS (remained well)	21	0	0	2	0	0	0	0	0	0	1	0	2	0
	D0 Vi-TCV (remained well) vs D1 Vi-TCV (remained well)	14	32	3	56	0	1	1	2	0	0	5	0	31	0
	D0 Vi-TCV (remained well) vs D7 Vi-TCV (remained well)	0	1	0	0	0	1	1	1	0	0	0	0	1	0
	D0 vs 12hr	3	27	7	53	0	8	1	12	0	0	3	0	23	0
	D0 Vi-PS vs 12hrs Vi-PS	0	5	0	16	0	0	0	1	0	0	0	0	12	0
	D0 Vi-TCV vs 12hrs Vi-TCV	1	12	7	27	0	8	1	9	0	0	1	0	4	0
	D0 (developed typhoid) vs 12hrs (developed typhoid)	9	1	1	0	0	0	0	1	0	0	0	0	0	0

	D0 (remained well) vs 12hrs (remained well)	9	8	2	8	0	2	0	6	0	0	0	0	2	0
	D0 Vi-PS (developed typhoid) vs 12hrs Vi-PS (developed typhoid)	0	5	0	5	0	0	0	4	0	0	0	0	6	0
	D0 Vi-PS (remained well) vs 12hrs Vi-PS (remained well)	1	1	0	4	0	0	0	0	0	0	0	0	1	0
	D0 Vi-TCV (remained well) vs 12hrs Vi-TCV (remained well)	1	1	0	4	0	0	1	3	0	0	0	0	1	0
	D0 any VI vaccine (developed typhoid) vs 12hrs any VI vaccine (developed typhoid)	0	0	0	0	0	0	0	1	0	0	0	0	1	0
	D0 any VI vaccine (remained well) vs 12hrs any VI vaccine (remained well)	1	14	2	13	0	0	0	5	0	0	0	0	6	0
	D0 control vaccine vs D7 control vaccine	5	0	1	0	0	0	0	0	0	0	0	0	0	0
	D0 control vaccine vs typhoid (developed typhoid)nosis control vaccine	129	44	12	156	0	63	2	17	1	0	2	0	19	0
	D7 control vaccine (remained well)nosed vs typhoid (developed typhoid)nosis control vaccine (developed typhoid)nosed	8	1	1	24	0	6	0	2	0	0	0	0	0	0
	D0 control vaccine vs 12hrs control vaccine	13	0	0	0	0	0	0	0	0	0	0	0	0	0
	D0 control vaccine (developed typhoid) vs 12hrs control vaccine (developed typhoid)	2	0	0	0	0	0	0	0	0	0	0	0	0	0
	D0 control vaccine vs D1 control vaccine	31	0	0	0	0	0	0	0	0	0	0	0	0	0
	D0 control vaccine (developed typhoid) vs D1 control vaccine (developed typhoid)	13	0	0	0	0	0	0	0	0	0	0	0	0	0
	D0.PS.diag_vs_D1.PS.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D0.control.non.diag_vs_D_12hrs.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D0.TCV.diag_vs_D_12hrs.TCV.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D0.control.non.diag_vs_D1.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Does vaccination influence sRNA expression after S. Typhi exposure?	D7 Vi-PS vs D7 Vi-TCV	13	0	0	0	0	1	0	0	0	0	0	0	0	0
	D0 control vaccine vs D0 Vi-PS	0	2	0	12	0	0	0	0	0	0	0	0	10	0
	D0 control vaccine vs D0 Vi-TCV	0	5	0	18	0	5	0	0	0	0	1	0	2	0
	D0 control vaccine vs D0 any VI vaccinevacc	0	15	0	61	0	9	1	3	0	0	2	1	21	0
	D0 Vi-PS (developed typhoid) vs D0 control vaccine (developed typhoid)	0	5	0	7	0	0	0	4	0	0	0	0	9	0
	D0 any VI vaccine (developed typhoid) vs D0 control vaccine (developed typhoid)	0	3	0	2	0	2	0	3	0	0	0	0	7	0
	D0 Vi-PS (remained well) vs D0 control vaccine (remained well)	0	1	0	19	0	0	0	0	0	0	0	0	1	0
	D0 any VI vaccine (remained well) vs D0 control vaccine (remained well)	0	0	0	11	0	1	0	0	0	0	1	0	0	0
	D1 control vaccine vs D1 Vi-PS	0	22	0	43	0	0	0	5	0	0	2	0	26	0
	D1 control vaccine vs D1 Vi-TCV	0	3	0	8	0	0	0	0	0	0	0	0	2	0
	D1 control vaccine vs D1 any VI vaccinevacc	0	28	0	52	0	1	0	6	0	0	2	0	36	0

	D1 Vi-PS (developed typhoid) vs D1 control vaccine (developed typhoid)	0	14	0	25	0	0	0	1	0	0	2	0	14	0
	D1 Vi-TCV (developed typhoid) vs D1 control vaccine (developed typhoid)	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	D1 any VI vaccine (developed typhoid) vs D1 control vaccine (developed typhoid)	0	10	0	19	0	0	0	1	0	0	1	0	9	0
	D0 Vi-TCV (remained well) vs D0 control vaccine (remained well)	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.Vi.PS_vs_D1.Vi.TCV	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.PS.diag_vs_D1.TCV.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	TD.Vi.PS_vs_TD.Vi.TCV	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.PS.non.diag_vs_D1.TCV.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D0.TCV.diag_vs_D0.control.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.control_vs_D_12hrs.Vi.PS	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.control_vs_D_12hrs.Vi.TCV	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.control_vs_D_12hrs.Enteric.vacc	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.PS.diag_vs_D_12hrs.control.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.TCV.diag_vs_D_12hrs.control.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.enteric.diag_vs_D_12hrs.control.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.PS.non.diag_vs_D_12hrs.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.TCV.non.diag_vs_D_12hrs.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.enteric.non.diag_vs_D_12hrs.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.Vi.PS_vs_D_12hrs.Vi.TCV	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.PS.diag_vs_D_12hrs.TCV.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.PS.non.diag_vs_D_12hrs.TCV.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.PS.non.diag_vs_D1.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.TCV.non.diag_vs_D1.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.enteric.non.diag_vs_D1.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Are there correlates of outcome?	12hrs (developed typhoid) vs 12hrs (remained well)	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	D1.diag_vs_D1.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.PS.diag_vs_D1.PS.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.TCV.diag_vs_D1.TCV.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	D0.control.diag_vs_D0.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.control.diag_vs_D_12hrs.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.enteric.diag_vs_D_12hrs.enteric.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.PS.diag_vs_D_12hrs.PS.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D_12hrs.TCV.diag_vs_D_12hrs.TCV.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D1.control.diag_vs_D1.control.non.diag	0	0	0	0	0	0	0	0	0	0	0	0	0	0

9.5.4 Pathway enriched amongst sRNA-dfs differentially expressed at 12 and 24-hour post challenge

9.5.4.1 sRNA-dfs that were unilaterally upregulated 12 hours post-challenge

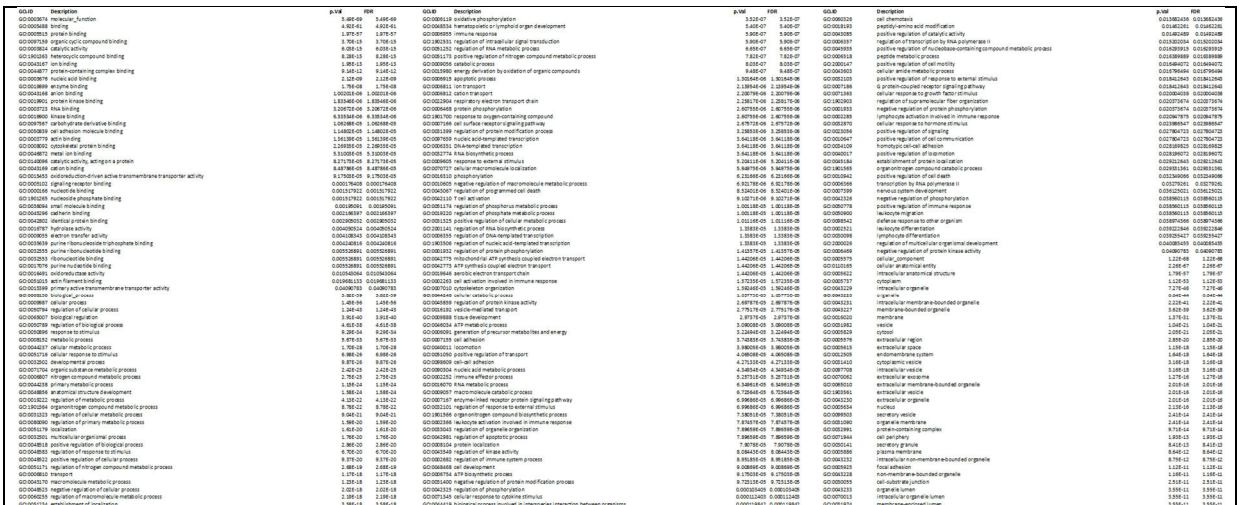
Appendix Table 8 GO molecular functions enriched (ORA) in. Major themes were ion transport, phosphorylation, nucleobase metabolism, and ATP metabolism. Taken from GProfilr web output

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$
molecular_function	GO:0003674	4.491×10^{-3}	
extracellular matrix structural constituent	GO:0005201	4.491×10^{-3}	
phosphotransferase activity, phosphate group as acce...	GO:0016776	1.130×10^{-2}	
cation binding	GO:0043169	1.395×10^{-2}	
metal ion binding	GO:0046872	1.395×10^{-2}	
ion binding	GO:0043167	1.543×10^{-2}	
binding	GO:0005488	2.180×10^{-2}	
kinase activity	GO:0016301	3.410×10^{-2}	
transferase activity, transferring phosphorus-containin...	GO:0016772	4.601×10^{-2}	
inositol hexakisphosphate kinase activity	GO:0000828	4.736×10^{-2}	
inositol heptakisphosphate kinase activity	GO:0000829	4.736×10^{-2}	
inositol hexakisphosphate 5-kinase activity	GO:0000832	4.736×10^{-2}	
catalytic activity	GO:0003824	4.736×10^{-2}	
nucleoside diphosphate kinase activity	GO:0004550	4.736×10^{-2}	
cyclin-dependent protein serine/threonine kinase activ...	GO:0004693	4.736×10^{-2}	
protein binding	GO:0005515	4.736×10^{-2}	
adenylate cyclase regulator activity	GO:0010854	4.736×10^{-2}	
cyclase inhibitor activity	GO:0010852	4.736×10^{-2}	
G protein-coupled receptor activity	GO:0004930	4.736×10^{-2}	
ionotropic glutamate receptor activity	GO:0004970	4.736×10^{-2}	
NMDA glutamate receptor activity	GO:0004972	4.736×10^{-2}	
structural molecule activity	GO:0005198	4.736×10^{-2}	
5'-nucleotidase activity	GO:0008253	4.736×10^{-2}	
nucleotidase activity	GO:0008252	4.736×10^{-2}	
glutamate receptor activity	GO:0008066	4.736×10^{-2}	
cannabinoid receptor activity	GO:0004949	4.736×10^{-2}	
protein serine kinase activity	GO:0106310	4.736×10^{-2}	
telethonin binding	GO:0031433	4.736×10^{-2}	
glutamate-gated calcium ion channel activity	GO:0022849	4.736×10^{-2}	
adenylate cyclase inhibitor activity	GO:0010855	4.736×10^{-2}	
diphosphoinositol-pentakisphosphate kinase activity	GO:0033857	4.736×10^{-2}	
IMP 5'-nucleotidase activity	GO:0050483	4.736×10^{-2}	
muscle alpha-actinin binding	GO:0051371	4.736×10^{-2}	
alpha-actinin binding	GO:0051393	4.736×10^{-2}	
transmitter-gated channel activity	GO:0022835	4.736×10^{-2}	
inositol hexakisphosphate 1-kinase activity	GO:0052723	4.736×10^{-2}	
aromatase activity	GO:0070330	4.736×10^{-2}	
cyclin-dependent protein kinase activity	GO:0097472	4.736×10^{-2}	
estrogen 16-alpha-hydroxylase activity	GO:0101020	4.736×10^{-2}	
5-diphosphoinositol pentakisphosphate 3-kinase activ...	GO:0102092	4.736×10^{-2}	
XMP 5'-nucleosidase activity	GO:0106411	4.736×10^{-2}	
hydroperoxy icosatetraenoate dehydratase activity	GO:0106256	4.736×10^{-2}	
inositol hexakisphosphate 3-kinase activity	GO:0052724	4.736×10^{-2}	
transmitter-gated ion channel activity	GO:0022824	4.736×10^{-2}	
inositol-1,3,4,5,6-pentakisphosphate kinase activity	GO:0000827	4.736×10^{-2}	
glycine binding	GO:0016594	4.736×10^{-2}	
oxidoreductase activity, acting on paired donors, with ...	GO:0016712	4.736×10^{-2}	
hydro-lyase activity	GO:0016836	4.736×10^{-2}	
carbon-oxygen lyase activity	GO:0016835	4.736×10^{-2}	
nucleobase-containing compound kinase activity	GO:0019205	4.736×10^{-2}	

9.5.4.2 sRNA-dfs differentially expressed 24 hours post-challenge

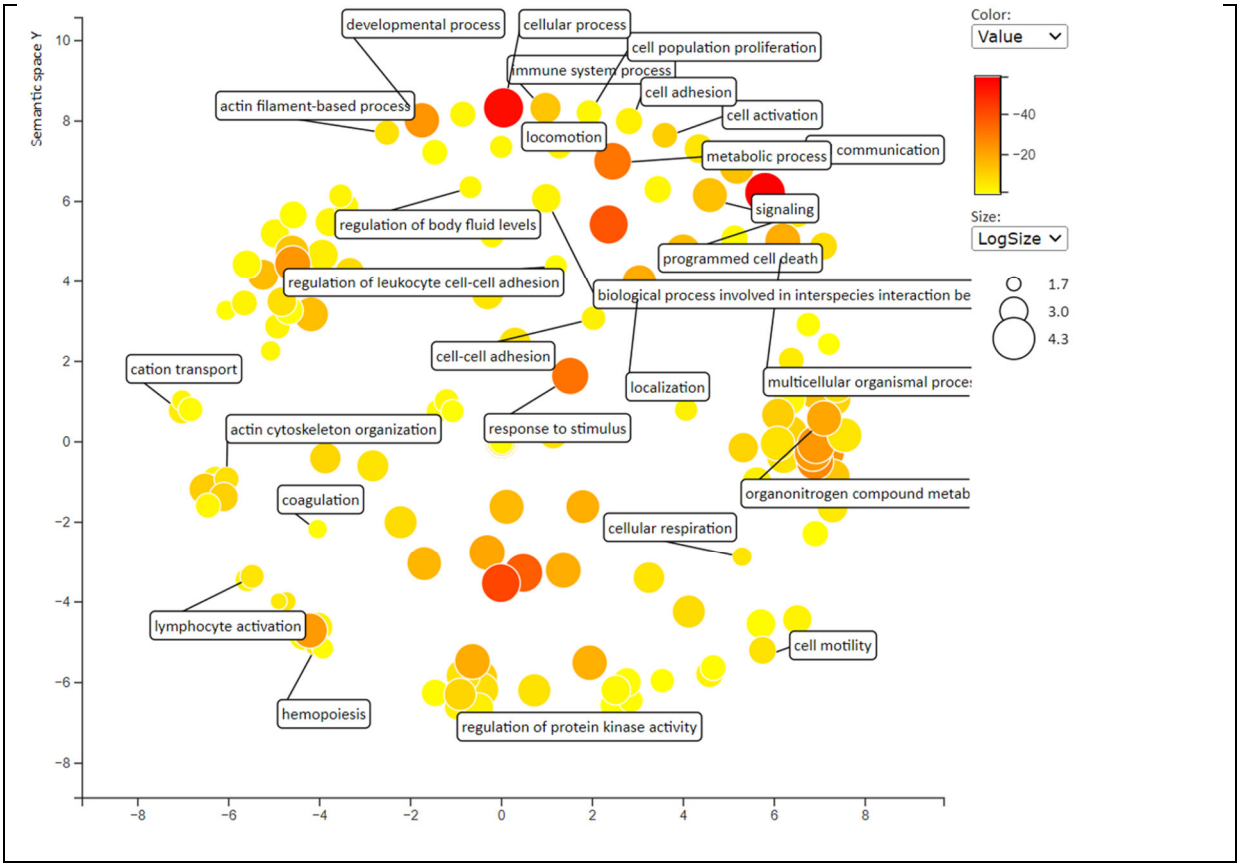
Many pathways were enriched in sRNA-dfs upregulated 24 hours after challenge. Some pathways related to the immune system and leukocytes, whilst others related to the actin cytoskeleton, ion transport and cellular respiration (Appendix Figure 43, Appendix Table 9)

Appendix Table 9 ORA results for GO terms and Reactome and Kegg pathways for upregulated sRNA-dfs



GO.ID	Description	p.Val	FDR
GO:0003674	molecular_function	5.68E-69	5.68E-69
GO:0005488	binding	4.92E-61	4.92E-61
GO:0005515	protein binding	1.87E-57	1.87E-57
GO:0097159	organic cyclic compound binding	3.70E-53	3.70E-53
GO:0036094	small molecule binding	6.03E-53	6.03E-53
GO:1901363	heterocyclic compound binding	8.28E-53	8.28E-53
GO:0043167	ion binding	1.95E-53	1.95E-53
GO:0044877	protein-containing complex binding	9.14E-52	9.14E-52
GO:0000976	nucleic acid binding	2.12E-09	2.12E-09
GO:0016829	enzyme binding	1.76E-08	1.76E-08
GO:0043168	anion binding	1.00201E-06	1.00201E-06
GO:0019901	protein kinase binding	1.83564E-06	1.83564E-06
GO:0000723	RNA binding	3.20672E-06	3.20672E-06
GO:0019900	kinase binding	6.93934E-06	6.93934E-06
GO:0097367	carbohydrate derivative binding	1.04268E-05	1.04268E-05
GO:0050839	cell adhesion molecule binding	1.14602E-05	1.14602E-05
GO:0000779	actin binding	1.36139E-05	1.36139E-05
GO:0008092	cytoskeletal protein binding	2.28935E-05	2.28935E-05
GO:0046872	metal ion binding	5.31003E-05	5.31003E-05
GO:0140096	catalytic activity, acting on a protein	8.27170E-05	8.27170E-05
GO:0043169	cation binding	8.48788E-05	8.48788E-05
GO:0015653	oxidoreduction-driven active transmembrane transporter activity	9.17502E-05	9.17502E-05
GO:0005515	signaling receptor binding	0.000176408	0.000176408
GO:0000166	nucleotide binding	0.001317922	0.001317922
GO:1901265	nucleoside phosphate binding	0.001317922	0.001317922
GO:0036094	small molecule binding	0.00195091	0.00195091
GO:0045296	cadherin binding	0.002186597	0.002186597
GO:0042802	identical protein binding	0.002905052	0.002905052
GO:0016787	hydrolase activity	0.004090514	0.004090514
GO:0009055	electron transfer activity	0.004108343	0.004108343
GO:0035639	purine ribonucleoside triphosphate binding	0.004248316	0.004248316
GO:0032555	purine ribonucleotide binding	0.005528891	0.005528891
GO:0032553	ribonucleotide binding	0.005528891	0.005528891
GO:0017076	purine nucleotide binding	0.005528891	0.005528891
GO:0016491	oxidoreductase activity	0.010540064	0.010543064
GO:0051015	actin filament binding	0.014981133	0.014981133
GO:0015399	primary active transmembrane transporter activity	0.04067783	0.04067783
GO:0008324	cation transmembrane transporter activity	0.469	0.469
GO:0008087	cellular process	1.45E-56	1.45E-56
GO:0009734	regulation of cellular process	1.24E-43	1.24E-43
GO:0008007	biological regulation	3.91E-40	3.91E-40
GO:0065739	regulation of biological process	4.61E-38	4.61E-38
GO:0016896	response to stimulus	8.29E-34	8.29E-34
GO:0008152	metabolic process	5.67E-33	5.67E-33
GO:0044237	cellular metabolic process	1.70E-28	1.70E-28
GO:0051716	cellular response to stimulus	6.98E-26	6.98E-26
GO:0032502	developmental process	9.87E-26	9.87E-26
GO:0071704	organic substance metabolic process	2.42E-25	2.42E-25
GO:0006807	nitrogen compound metabolic process	2.76E-25	2.76E-25
GO:0050789	regulation of biological process	1.18E-24	1.18E-24
GO:0048856	anatomical structure development	1.38E-24	1.38E-24
GO:0019222	regulation of metabolic process	4.12E-22	4.12E-22
GO:1901564	organonitrogen compound metabolic process	8.76E-22	8.76E-22
GO:0031323	regulation of cellular metabolic process	9.04E-21	9.04E-21
GO:0080090	regulation of primary metabolic process	1.36E-20	1.36E-20
GO:0051179	localization	1.61E-20	1.61E-20
GO:0051301	cell division	1.76E-20	1.76E-20
GO:0048518	positive regulation of biological process	2.86E-20	2.86E-20
GO:0048583	regulation of response to stimulus	6.70E-20	6.70E-20
GO:0048522	positive regulation of cellular process	9.37E-20	9.37E-20
GO:0051171	regulation of nitrogen compound metabolic process	2.68E-19	2.68E-19
GO:0006810	transport	1.17E-18	1.17E-18
GO:0043170	macromolecule metabolic process	1.25E-18	1.25E-18
GO:0048523	negative regulation of cellular process	2.02E-18	2.02E-18
GO:0060255	regulation of macromolecule metabolic process	2.18E-18	2.18E-18
GO:0051234	establishment of localization	3.58E-18	3.58E-18
GO:0048869	cellular developmental process	9.86E-18	9.86E-18
GO:0009893	negative regulation of biological process	1.34E-17	1.34E-17
GO:0032504	cell differentiation	2.27E-17	2.27E-17
GO:0010033	response to organic substance	2.39E-17	2.39E-17
GO:0051641	cellular component organization	4.66E-17	4.66E-17
GO:0009890	negative regulation of biosynthetic process	5.22E-17	5.22E-17
GO:0071840	cellular component organization or biogenesis	1.52E-16	1.52E-16
GO:0007275	multicellular organism development	2.12E-16	2.12E-16
GO:0007165	signal transduction	2.27E-16	2.27E-16
GO:0007154	cell communication	2.75E-16	2.75E-16
GO:0023052	signaling	4.30E-16	4.30E-16
GO:0042221	response to chemical	5.03E-15	5.03E-15
GO:0018958	phenol-containing compound metabolic process	5.09E-15	5.09E-15
GO:0031976	plasma membrane-bounded cell projection	5.48E-15	5.48E-15
GO:0048731	system development	7.70E-15	7.70E-15
GO:0009058	biosynthetic process	1.34E-14	1.34E-14
GO:0044249	cellular biosynthetic process	3.48E-14	3.48E-14
GO:1901576	organic substance biosynthetic process	1.21E-13	1.21E-13
GO:0022607	cellular component assembly	2.55E-13	2.55E-13
GO:0002773	cell activation	4.20E-13	4.20E-13
GO:0044271	cellular nitrogen compound biosynthetic process	4.85E-13	4.85E-13
GO:0010468	regulation of gene expression	5.45E-13	5.45E-13
GO:0071310	cellular response to organic substance	9.97E-13	9.97E-13
GO:0048513	animal organ development	1.07E-12	1.07E-12
GO:0071887	cellular response to chemical stimulus	1.17E-12	1.17E-12
GO:0048513	protein-containing complex organization	1.92E-12	1.92E-12
GO:0044085	cellular component biogenesis	3.55E-12	3.55E-12
GO:0000796	phosphate-containing compound metabolic process	5.28E-12	5.28E-12
GO:0044281	small molecule metabolic process	6.40E-12	6.40E-12
GO:0006996	organelle organization	1.07E-11	1.07E-11
GO:0006793	phosphorus metabolic process	1.15E-11	1.15E-11
GO:0051049	regulation of protein metabolic process	1.89E-11	1.89E-11
GO:0051641	cellular localization	2.30E-11	2.30E-11
GO:0010467	gene expression	3.50E-11	3.50E-11
GO:0051128	regulation of cellular component organization	5.09E-11	5.09E-11
GO:0006753	nucleoside-containing complex assembly	6.77E-11	6.77E-11
GO:0009966	regulation of signal transduction	1.39E-10	1.39E-10
GO:0009719	response to endogenous stimulus	6.27E-10	6.27E-10
GO:0023051	regulation of signaling	6.64E-10	6.64E-10
GO:0051716	cellular response to stimulus	6.64E-10	6.64E-10
GO:0051130	heterocycle biosynthetic process	8.79E-10	8.79E-10
GO:0044283	aromatic compound biosynthetic process	8.79E-10	8.79E-10
GO:0009889	regulation of biosynthetic process	1.35E-09	1.35E-09
GO:0009059	macromolecule biosynthetic process	1.56E-09	1.56E-09
GO:0006334	regulation of response to stress	1.64E-09	1.64E-09
GO:0006219	cell death	1.68E-09	1.68E-09
GO:0019438	nucleobase-containing compound biosynthetic process	1.76E-09	1.76E-09
GO:0006950	positive regulation of metabolic process	2.23E-09	2.23E-09
GO:0010604	regulation of macromolecule biosynthetic process	2.48E-09	2.48E-09
GO:0010556	regulation of macromolecule metabolic process	2.92E-09	2.92E-09
GO:0006677	cell migration	3.35E-09	3.35E-09
GO:0030036	actin cytoskeleton organization	3.72E-09	3.72E-09
GO:0022879	regulation of localization	5.35E-09	5.35E-09
GO:0031326	regulation of cellular biosynthetic process	5.57E-09	5.57E-09
GO:1901293	organic cyclic compound biosynthetic process	7.51E-09	7.51E-09
GO:0048870	cell motility	1.10E-08	1.10E-08
GO:0051639	actin filament-based process	1.17E-08	1.17E-08
GO:0043933	protein modification process	1.18E-08	1.18E-08
GO:0071495	cellular response to endogenous stimulus	1.45E-08	1.45E-08
GO:0044283	heterocycle metabolic process	1.46E-08	1.46E-08
GO:0043412	macromolecule modification	1.48E-08	1.48E-08
GO:0006609	regulation of molecular function	1.70E-08	1.70E-08
GO:0010604	positive regulation of macromolecule metabolic process	1.86E-08	1.86E-08
GO:0006725	cellular aromatic compound metabolic process	1.86E-08	1.86E-08
GO:0051049	regulation of transport	2.41E-08	2.41E-08
GO:0051049	establishment of localization in cell	2.34E-08	2.34E-08
GO:0010519	regulation of nucleobase-containing compound metabolic process	2.70E-08	2.70E-08
GO:0051239	regulation of multicellular organismal process	3.31E-08	3.31E-08
GO:0006139	nucleobase-containing compound metabolic process	4.37E-08	4.37E-08
GO:0050878	regulation of body fluid levels	5.12E-08	5.12E-08
GO:0044533	cellular respiration	5.48E-08	5.48E-08
GO:0048468	cell development	5.66E-08	5.66E-08
GO:0031324	negative regulation of cellular metabolic process	7.19E-08	7.19E-08
GO:0007275	regulation of developmental process	7.19E-08	7.19E-08
GO:0006520	immune system development	7.55E-08	7.55E-08
GO:0048584	positive regulation of response to stimulus	9.01E-08	9.01E-08
GO:0051172	negative regulation of nitrogen compound metabolic process	1.01E-07	1.01E-07
GO:1901360	organic cyclic compound metabolic process	1.23E-07	1.23E-07
GO:0009790	regulation of catalytic activity	1.82E-07	1.82E-07
GO:0033036	macromolecule localization	1.86E-07	1.86E-07
GO:0048585	negative regulation of response to stimulus	1.94E-07	1.94E-07
GO:0001841	regulation of cell death	2.07E-07	2.07E-07
GO:0012501	programmed cell death	2.15E-07	2.15E-07
GO:0044531	leukocyte activation	2.31E-07	2.31E-07
GO:0009892	negative regulation of metabolic process	2.32E-07	2.32E-07
GO:0051248	negative regulation of protein metabolic process	2.66E-07	2.66E-07
GO:0030097	hemopoiesis	3.10E-07	3.10E-07

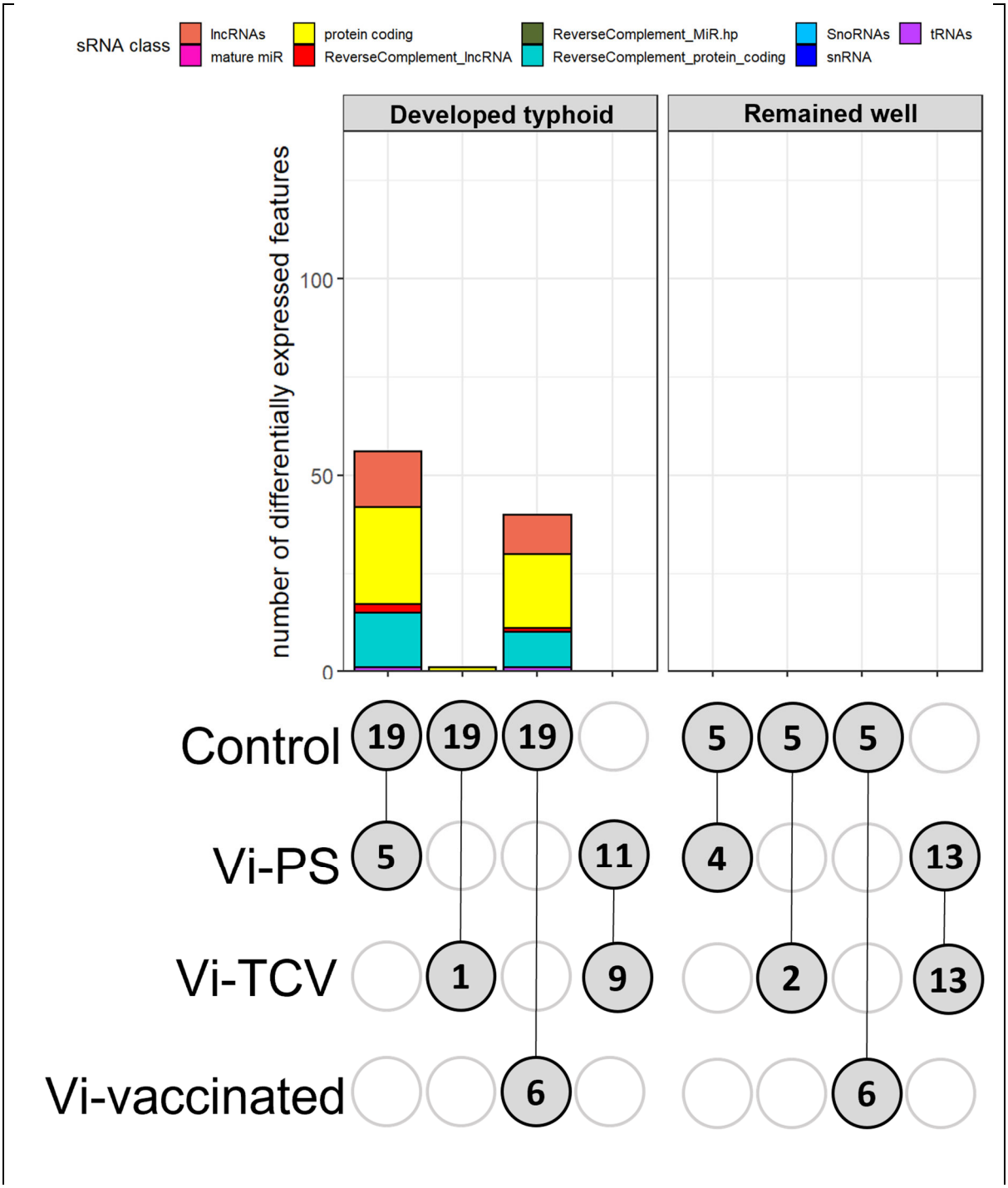
GO.ID	Description	p.Val	FDR
GO:0006119	oxidative phosphorylation	3.32E-07	3.32E-07
GO:0048534	hematopoietic or lymphoid organ development	5.40E-07	5.40E-07
GO:0006955	immune response	5.90E-07	5.90E-07
GO:1902531	regulation of intracellular signal transduction	5.90E-07	5.90E-07
GO:0051252	regulation of RNA metabolic process	6.65E-07	6.65E-07
GO:0051173	positive regulation of nitrogen compound metabolic process	7.82E-07	7.82E-07
GO:0009056	catabolic process	8.03E-07	8.03E-07
GO:0015980	energy derivation by oxidation of organic compounds	9.48E-07	9.48E-07
GO:0006915	apoptotic process	1.30264E-06	1.30264E-06
GO:0006811	ion transport	2.13864E-06	2.13864E-06
GO:0006812	cation transport	2.20879E-06	2.20879E-06
GO:0022904	respiratory electron transport chain	2.23617E-06	2.23617E-06
GO:0006468	protein phosphorylation	2.60755E-06	2.60755E-06
GO:1901700	response to oxygen-containing compound	2.60755E-06	2.60755E-06
GO:0007166	cell surface receptor signaling pathway	2.67572E-06	2.67572E-06
GO:0018193	peptidyl-amino acid modification	3.23831E-06	3.23831E-06
GO:0006351	DNA-templated transcription	3.64118E-06	3.64118E-06
GO:0032774	RNA biosynthetic process	3.64118E-06	3.64118E-06
GO:0006915	apoptotic process	3.52641E-06	3.52641E-06
GO:0006811	ion transport	2.13864E-06	2.13864E-06
GO:0006812	cation transport	2.20879E-06	2.20879E-06
GO:0022904	respiratory electron transport chain	2.23617E-06	2.23617E-06
GO:0006468	protein phosphorylation	2.60755E-06	2.60755E-06
GO:0071363	cellular response to growth factor stimulus	2.67572E-06	2.67572



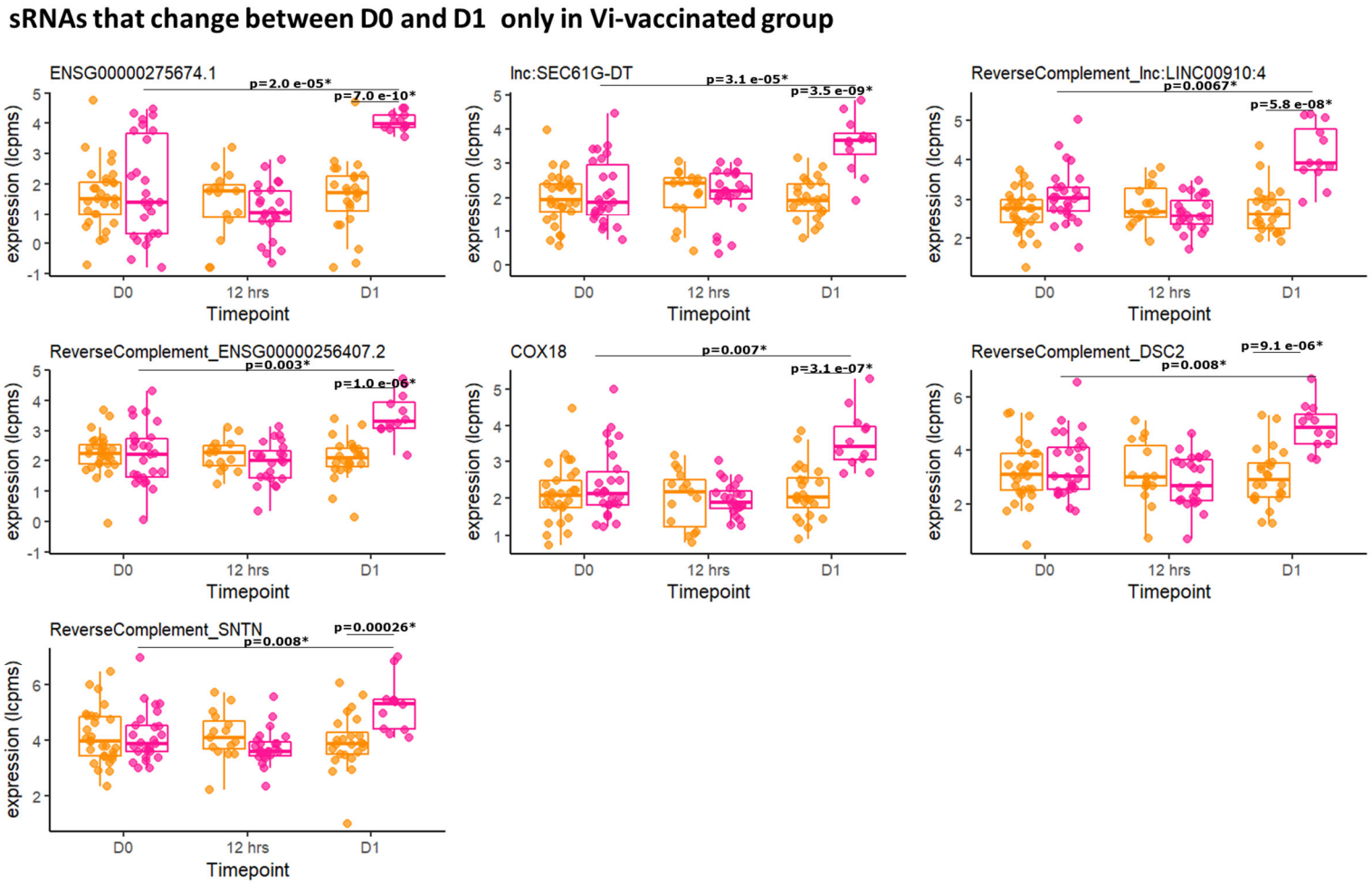
Appendix Figure 40 GO terms enriched in the genes from, which sRNA-dfs DE 24 hours post challenge are derived. The long list of pathways in appendix are summarised by semantically grouping them into clusters and labelling those clusters. MDS plot created using Revigo[565].

Supplementary Table 8 MIEAA 2 D0 vs D1 enriched pathways. Relates to Figure 79 of main text

process	Subcategory	Observed	Enrichment	P-value	FDR
RNA silencing	RISC complex GO0016442	147	enriched	4.89E-07	0.004
	pre-miRNA processing GO0031054	134	enriched	4.40E-05	0.027
	miRNA loading onto RISC involved in gene silencing by miRNA GO0035280	113	enriched	5.47E-05	0.027
	positive regulation of miRNA mediated inhibition of translation GO1905618	43	enriched	7.57E-05	0.027
	endoribonuclease activity GO0004521	124	enriched	7.58E-05	0.027
	miRNA metabolic process GO0010586	122	enriched	9.10E-05	0.027
	double-stranded RNA binding GO0003725	189	enriched	9.73E-05	0.027
	endoribonuclease activity, cleaving miRNA-paired mRNA GO0090624	60	enriched	1.01E-04	0.027
	RISC-loading complex GO0070578	111	enriched	1.33E-04	0.032
	siRNA binding GO0035197	74	enriched	1.68E-04	0.037
	production of siRNA involved in RNA interference GO0030422	82	enriched	1.81E-04	0.039
	production of miRNAs involved in gene silencing by miRNA GO0035196	128	enriched	2.34E-04	0.045
	posttranscriptional gene silencing by RNA GO0035194	130	enriched	2.63E-04	0.045
	miRNA binding GO0035198	168	enriched	2.66E-04	0.045
	mRNA cleavage involved in gene silencing by miRNA GO0035279	95	enriched	2.89E-04	0.046
	mRNA cleavage involved in gene silencing by siRNA GO0090625	95	enriched	2.89E-04	0.046
	siRNA loading onto RISC involved in RNA interference GO0035087	62	enriched	3.04E-04	0.047
cell division and proliferation	cyclin-dependent protein serinethreonine kinase regulator activity GO0016538	155	enriched	5.09E-05	0.027
	positive regulation of telomerase activity GO0051973	142	enriched	6.84E-05	0.027
	SMAD protein complex assembly GO0007183	93	enriched	7.07E-05	0.027
	negative regulation of G0 to G1 transition GO0070317	161	enriched	8.26E-05	0.027
	mitotic cell cycle phase transition GO0044772	133	enriched	1.47E-04	0.034
	DNA damage checkpoint GO0000077	117	enriched	1.83E-04	0.039
	macrophage colony-stimulating factor receptor binding GO0005157	11	enriched	3.40E-04	0.048
	regulation of stem cell proliferation GO0072091	71	enriched	3.64E-04	0.048
	regulation of stem cell proliferation GO0072091	71	enriched	3.64E-04	0.048
nervous system	axon guidance GO0007411	229	enriched	5.42E-06	0.020
	lens development in camera-type eye GO0002088	43	enriched	1.61E-05	0.022
	disordered domain specific binding GO0097718	174	enriched	3.81E-05	0.027
	positive regulation of microglial cell migration GO1904141	13	enriched	8.03E-05	0.027
	neurotransmitter secretion GO0007269	92	enriched	1.02E-04	0.027
	axon GO0030424	228	enriched	1.50E-04	0.034
	regulation of synapse assembly GO0051963	21	enriched	2.12E-04	0.042
	syntaxin-1 binding GO0017075	62	enriched	2.40E-04	0.045
	tau protein binding GO0048156	170	enriched	3.51E-04	0.048
	Spemann organizer formation GO0060061	20	enriched	3.54E-04	0.048
leukocyte	positive regulation of mononuclear cell proliferation GO0032946	15	enriched	1.78E-05	0.022
	immune response GO0006955	226	enriched	4.66E-05	0.027
	dendritic cell differentiation GO0097028	26	enriched	8.29E-05	0.027
	negative regulation of leukocyte tethering or rolling GO1903237	10	enriched	1.19E-04	0.030
	positive regulation of chemotaxis GO0050921	66	enriched	1.88E-04	0.039
	spleen development GO0048536	34	enriched	2.29E-04	0.045
	antibacterial humoral response GO0019731	59	enriched	3.51E-04	0.048
	regulation of immune response GO0050776	192	enriched	3.65E-04	0.048
signalling	GTPase activity GO0003924	256	enriched	8.11E-06	0.022
	calcium channel complex GO0034704	94	enriched	1.47E-05	0.022
	intracellular signal transduction GO0035556	268	enriched	1.74E-05	0.022
	GTP binding GO0005525	245	enriched	7.45E-05	0.027
	regulation of nitric-oxide synthase activity GO0050999	140	enriched	7.89E-05	0.027
	negative regulation of phosphoprotein phosphatase activity GO0032515	96	enriched	1.85E-04	0.039
	G-protein beta-subunit binding GO0031681	53	enriched	2.84E-04	0.046
	protein kinase activity GO0004672	263	enriched	2.91E-04	0.046
stress	response to hypoxia GO0001666	191	enriched	4.46E-07	0.004
	protein refolding GO0042026	150	enriched	3.85E-05	0.027
	unfolded protein binding GO0051082	205	enriched	7.43E-05	0.027
	protein destabilization GO0031648	163	enriched	7.44E-05	0.027
	positive regulation of stress fiber assembly GO0051496	137	enriched	1.42E-04	0.034
	cellular response to heat GO0034605	162	enriched	2.67E-04	0.045
transcripton	transcription coactivator activity GO0003713	259	enriched	3.68E-05	0.027
	DNA-binding transcription repressor activity, RNA polymerase II-specific GO0001227	260	enriched	5.35E-05	0.027
	regulation of gene expression GO0010468	221	enriched	6.61E-05	0.027
	RNA polymerase II regulatory region sequence-specific DNA binding GO0000977	268	enriched	1.01E-04	0.027
	positive regulation of gene expression GO0010628	275	enriched	2.43E-04	0.045
	histone H4-K20 trimethylation GO0034773	4	depleted	3.60E-04	0.048
apoptosis	positive regulation of endothelial cell apoptotic process GO2000353	65	enriched	1.24E-05	0.022
	hippo signaling GO0035329	125	enriched	4.37E-05	0.027
	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process GO0043154	186	enriched	6.17E-05	0.027
	positive regulation of macroautophagy GO0016239	133	enriched	1.60E-04	0.035
cytoskeleton	cytoskeleton organization GO0007010	191	enriched	2.70E-04	0.045
	actin filament binding GO0051015	207	enriched	3.06E-04	0.047
	actin filament reorganization involved in cell cycle GO0030037	3	depleted	3.13E-04	0.047
	regulation of cellular component movement GO0051270	3	depleted	3.13E-04	0.047
ion	ion transport GO0006811	147	enriched	9.43E-05	0.027
	response to calcium ion GO0051592	114	enriched	1.05E-04	0.027
	zinc ion binding GO0008270	269	enriched	1.24E-04	0.030
	chloride transmembrane transport GO1902476	89	enriched	1.94E-04	0.039
protein	regulation of cellular protein localization GO1903827	88	enriched	9.04E-05	0.027
	protein folding GO0006457	219	enriched	1.24E-04	0.030
	positive regulation of protein complex disassembly GO0043243	60	enriched	2.51E-04	0.045
	regulation of ubiquitin-dependent protein catabolic process GO2000058	9	enriched	2.62E-04	0.045
metabolism	cellular response to glucose stimulus GO0071333	153	enriched	2.52E-04	0.045
	negative regulation of oxidative phosphorylation GO0090324	42	enriched	3.43E-04	0.048
	positive regulation of ATP metabolic process GO1903580	4	depleted	3.60E-04	0.048
vesicle	early endosome GO0005769	231	enriched	9.30E-06	0.022
	secretory granule GO0030141	107	enriched	7.76E-05	0.027
	late endosome GO0005770	201	enriched	8.63E-05	0.027
clotting	negative regulation of plasminogen activation GO0010757	36	enriched	2.08E-05	0.023
	negative regulation of blood coagulation GO0030195	61	enriched	7.87E-05	0.027
Cell migration	positive regulation of cell migration GO0030335	258	enriched	9.86E-05	0.027
cytoskeleton	vesicle transport along actin filament GO0030050	73	enriched	8.99E-05	0.027
lipid metabolism	very-low-density lipoprotein particle clearance GO0034447	15	enriched	5.83E-05	0.027
other	positive regulation of smooth muscle cell proliferation GO0048661	123	enriched	2.73E-05	0.027
	erythrocyte homeostasis GO0034101	33	enriched	1.49E-04	0.034
	positive regulation of aspartic-type endopeptidase activity involved in amyloid precursor protein catabolic process GO1902961	60	enriched	3.41E-04	0.048
	procollagen-proline 4-dioxygenase complex GO0016222	6	depleted	3.27E-06	0.016
	procollagen-proline 4-dioxygenase complex GO0016222	6	depleted	3.27E-06	0.016
	intracellular GO0005622	195	enriched	1.06E-05	0.022
	intracellular GO0005622	195	enriched	1.06E-05	0.022
	leftright axis specification GO0070986	14	enriched	2.17E-05	0.023
	leftright axis specification GO0070986	14	enriched	2.17E-05	0.023
	blood vessel development GO0001568	79	enriched	5.75E-05	0.027
	blood vessel development GO0001568	79	enriched	5.75E-05	0.027
	1-phosphatidylinositol-3-kinase regulator activity GO0046935	87	enriched	5.89E-05	0.027
	1-phosphatidylinositol-3-kinase regulator activity GO0046935	87	enriched	5.89E-05	0.027
	establishment of protein localization GO0045184	150	enriched	7.70E-05	0.027
	establishment of protein localization GO0045184	150	enriched	7.70E-05	0.027
	regulation of mitochondrial mRNA stability GO0044528	34	enriched	7.81E-05	0.027
	positive regulation of epithelial to mesenchymal transition GO0010718	183	enriched	1.02E-04	0.027
	positive regulation of epithelial to mesenchymal transition GO0010718	183	enriched	1.02E-04	0.027
	spectrin binding GO0030507	89	enriched	1.06E-04	0.027
	spectrin binding GO0030507	89	enriched	1.06E-04	0.027
	phosphatidylinositol 3-kinase complex GO0005942	112	enriched	2.11E-04	0.042
	phosphatidylinositol 3-kinase complex GO0005942	112	enriched	2.11E-04	0.042
	core promoter binding GO0001047	134	enriched	2.40E-04	0.045
	core promoter binding GO0001047	134	enriched	2.40E-04	0.045
	GDP binding GO0019003	196	enriched	2.46E-04	0.045
	GDP binding GO0019003	196	enriched	2.46E-04	0.045
	positive regulation of TORC2 signaling GO1904515	9	enriched	2.62E-04	0.045
	positive regulation of TORC2 signaling GO1904515	9	enriched	2.62E-04	0.045
	dendrite GO0030425	229	enriched	2.85E-04	0.046
	dendrite GO0030425	229	enriched	2.85E-04	0.046
	cellular response to amyloid-beta GO1904646	158	enriched	3.03E-04	0.047
	cellular response to amyloid-beta GO1904646	158	enriched	3.03E-04	0.047
	circadian rhythm GO0007623	163	enriched	3.53E-04	0.048
	circadian rhythm GO0007623	163	enriched	3.53E-04	0.048
	ADP-ribosylation factor binding GO0030306	38	enriched	3.59E-04	0.048
	ADP-ribosylation factor binding GO0030306	38	enriched	3.59E-04	0.048

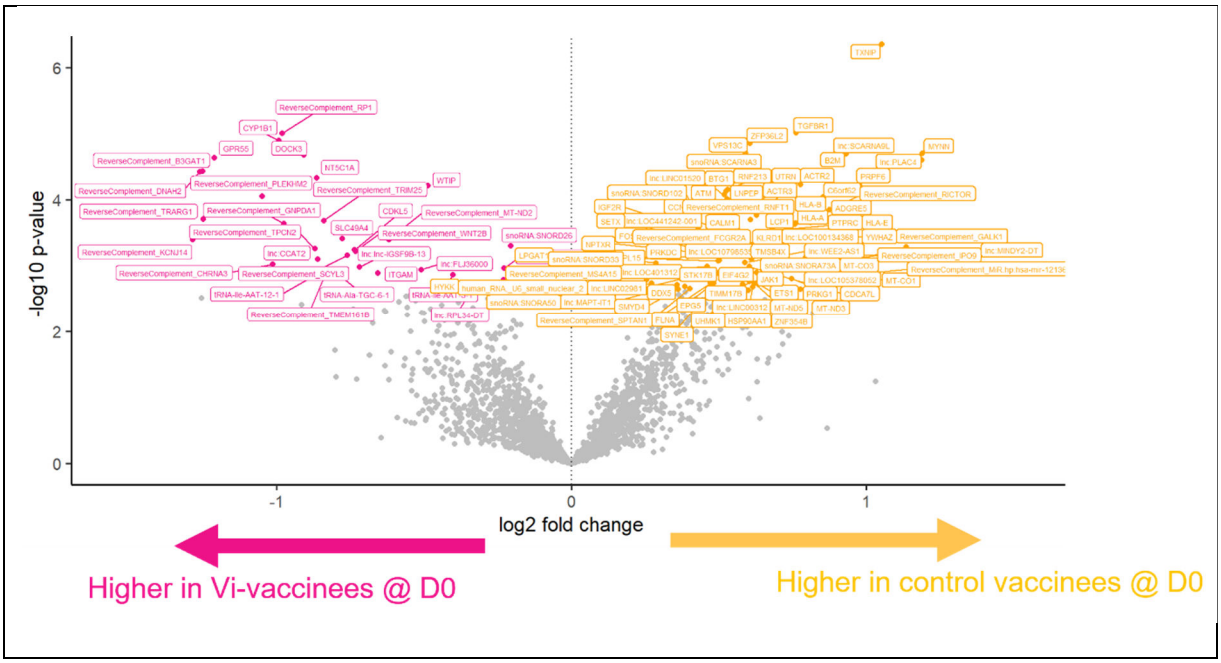


Appendix Figure 41 Summary plot influence of vaccine on D1 post challenge broken down by outcome

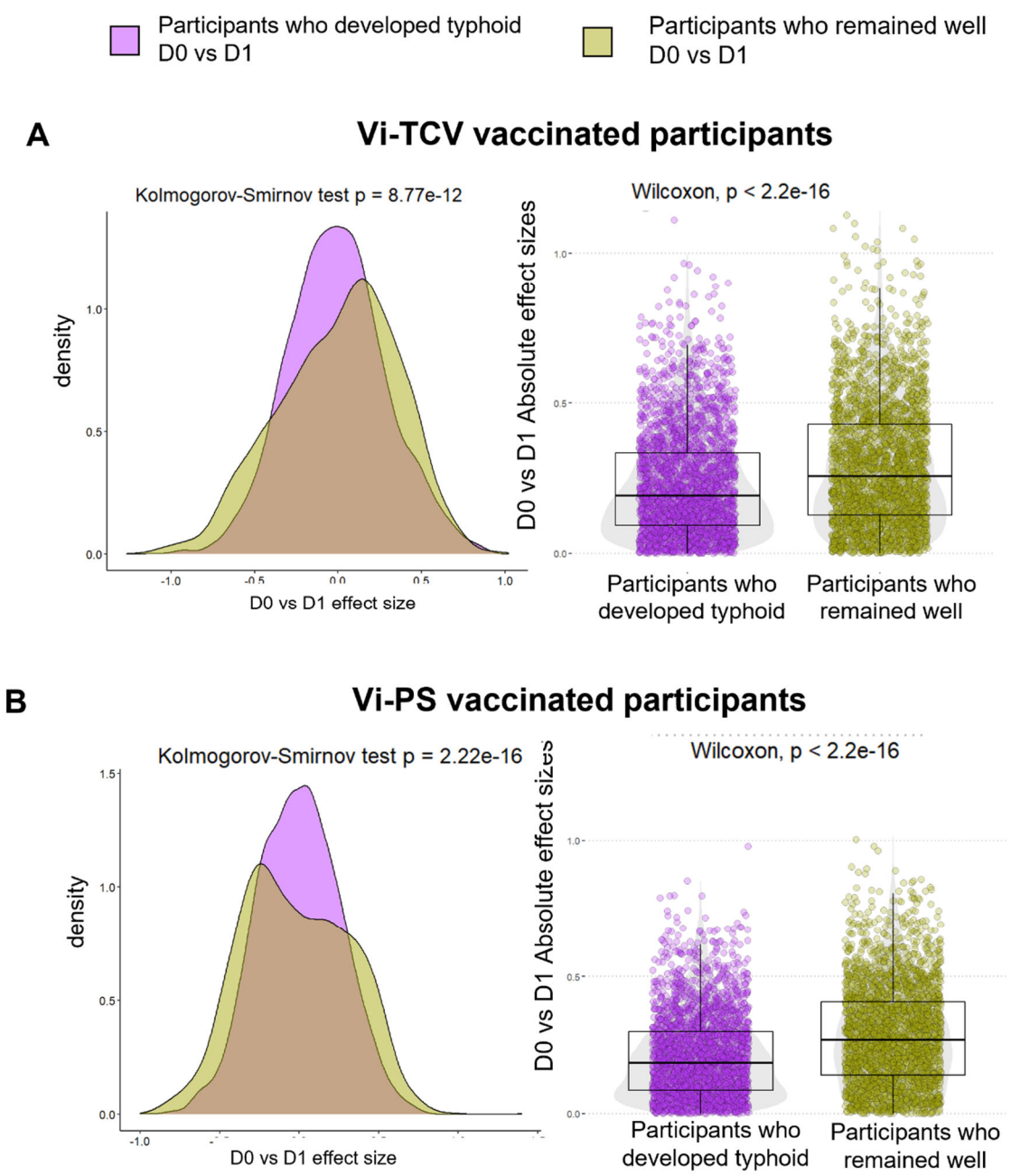


Supplementary figure 6 sRNAs that are unilaterally DE between D0 and D1 in Vi-vaccinees only. sRNAs identified as DE between D0 and D1 in the Vi-vaccine group AND DE at D1 between Vi-vaccinees and control vaccinees AND DE in terms of D0 to D1 foldchange between Vi and control vaccines.

9.5.5 Differences between control vaccinated individuals and Vi-vaccinated individuals at D0 (28 days after vaccination, pre-challenge)



Appendix Figure 42 Volcano plot of DE sRNAs between Vi-vaccinees and control vaccinees at D0. Coloured labelled points had FDR<0.05. Pathway enrichment results (not shown) showed positive enrichment for pathways pertaining to T-cell immunity, negative regulation of NK cell responses and immunoglobulin somatic hypermutation in the control vaccinees.



Appendix Figure 43 compare D0 vs D1 effect sizes between diagnosed and non-diagnosed participants broken down by Vi-vaccine.
A: Vi-TCV vaccinees left panel—distribution of D0 vs D1 effect sizes in diagnosed (purple) and non-diagnosed (green) participants. Kolmogorov Smirnov statistic is used to compare the distributions. The tails of effect sizes were longer in those that did not develop typhoid indicating more robust changes compared with the those that developed typhoid. The plot on the right shows a similar picture, with data presented and analysed by absolute effect size.
B: Vi-PS vaccinees left panel—distribution of D0 vs D1 effect sizes in diagnosed (purple)

and non-diagnosed (green) participants. Kolmogorov Smirnov statistic is used to compare the distributions. The tails of effect sizes were longer in those that did not develop typhoid indicating more robust changes compared with the those that developed typhoid. The plot on the right shows a similar picture, with data presented and analysed by absolute effect size.

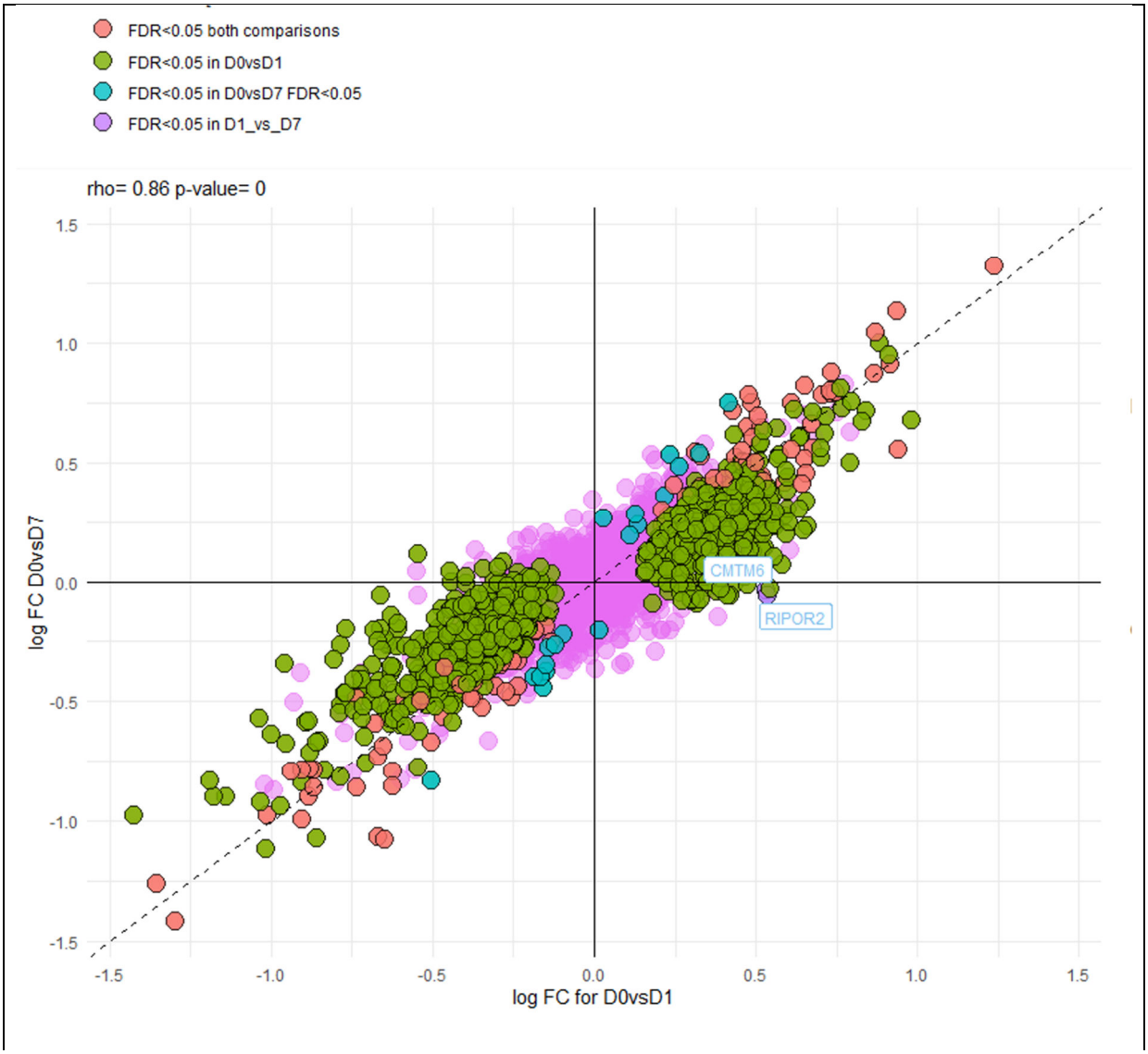
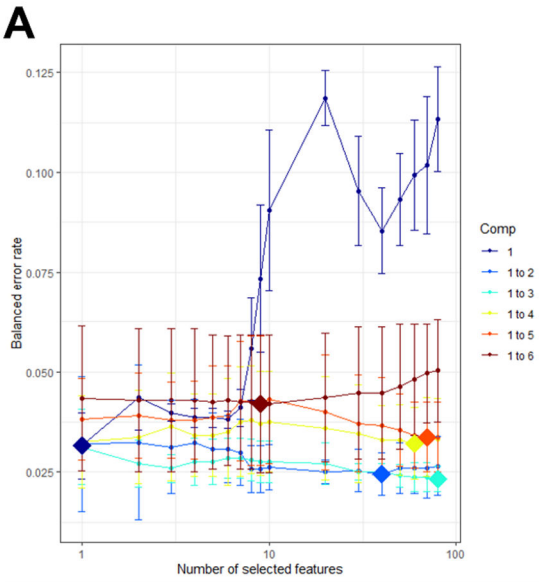
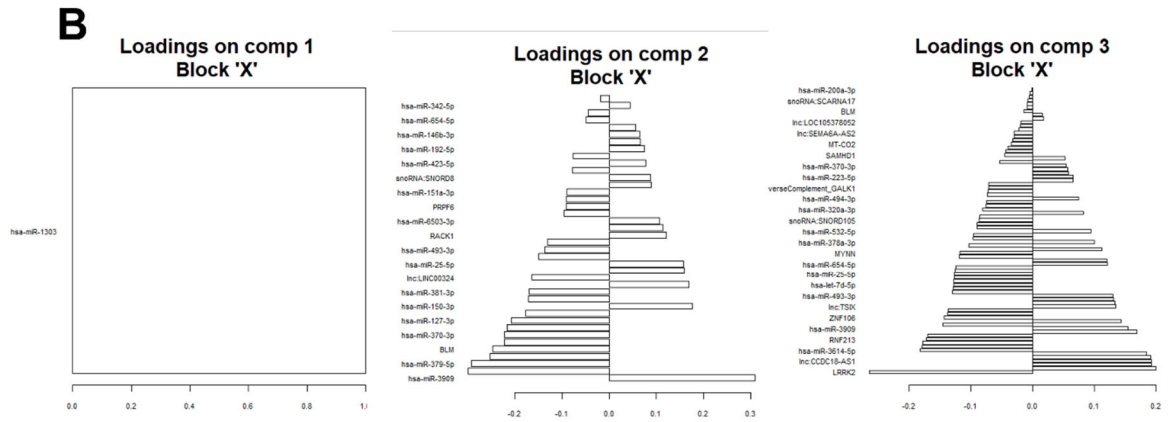


Figure 107 Agreement plot comparing log_2 foldchanges between D0 and D1 (x-axis) and D0 and D7 (y-axis) in Vi-vaccinees.

9.5.6 sPLS-DA to differentiate healthy samples from acute typhoid samples.

A sparse PLS-DA model was built to determine whether sRNA expression could differentiate samples taken at diagnosis from samples taken at D0 and D7. sRNA expression was corrected for flow cell prior to input to the model. Feature selection was performed using the 10 sRNAs with the highest loadings on the first principle components of a PCA of baseline, D0 and D7 samples. The samples were randomly split 80:20 into a training and test set stratified for time point. The model parameters were tuned as in the previous PLS-DA analysis above. The optimal model contained 3 components. The sPLS-D1 model differentiated diagnosis from baseline, D0 and D7 samples with an accuracy of 98% and 95% and a BER of 0.02 and 0.07 in test and training data, respectively.





C

Training

		Truth	
		D0 or D7	Typhoid diagnosis
predicted	D0 or D7	111	1
	Typhoid diagnosis	1	30

Test

		Truth	
		D0 or D7	Typhoid diagnosis
predicted	D0 or D7	34	1
	Typhoid diagnosis	1	7

Appendix Figure 44 PLS-DA model for differentiating D7 and D0 samples from diagnosis samples.

A: Parameter tuning, the optimum number of components was 1.

B: feature contribution to the model

C: Model performance

Appendix Table 10 Features DE features D0 vs Typhoid diagnosis in Vi-vaccinated participants. Features in bold were validated in the control sub group.

Ensembl name	logFC	CI.L	CI.R	AveExpr	t	P.Value	adj.P.Val
GBP1	**3.14**	**2.79**	**3.50**	**5.81**	**16.98**	**3.60E-39**	**6.48E-36**
hsa-miR-1303	**1.92**	**1.67**	**2.19**	**6.22**	**14.56**	**1.34E-37**	**3.52E-35**
LAP3	**1.73**	**1.50**	**1.97**	**7.16**	**14.41**	**4.25E-37**	**5.66E-35**
RNF213	**1.60**	**1.38**	**1.83**	**5.10**	**13.91**	**2.21E-35**	**1.91E-33**
LIN52	**1.64**	**1.41**	**1.88**	**5.28**	**13.51**	**1.24E-33**	**7.28E-32**
WARS1	**2.61**	**2.20**	**3.02**	**7.87**	**12.63**	**9.82E-30**	**1.88E-27**
GBP2	**2.07**	**1.73**	**2.41**	**7.13**	**12.18**	**2.23E-28**	**3.68E-26**
FGL2	**1.83**	**1.53**	**2.13**	**7.80**	**12.00**	**6.46E-28**	**2.17E-25**
hsa-miR-95-3p	**1.83**	**1.52**	**2.13**	**7.87**	**11.91**	**9.54E-28**	**6.08E-25**
CH7A	**1.08**	**0.88**	**1.28**	**3.33**	**10.74**	**2.17E-23**	**5.93E-21**
SCHIP1	**1.06**	**1.03**	**1.09**	**10.74**	**10.74**	**2.17E-23**	**5.93E-21**
hsa-miR-21-3p	**1.19**	**0.96**	**1.42**	**6.16**	**10.19**	**1.83E-21**	**1.09E-19**
MIR3p.hsa-mir-21	**1.19**	**0.96**	**1.42**	**6.16**	**10.19**	**1.83E-21**	**1.09E-19**
GNB4	**1.19**	**1.11**	**1.66**	**1.71**	**9.82**	**8.35E-20**	**6.79E-18**
TIFA	**0.85**	**1.28**	**1.28**	**5.05**	**9.82**	**8.35E-20**	**6.79E-18**
PARP14	**1.23**	**0.98**	**1.49**	**7.20**	**9.52**	**3.35E-19**	**6.09E-17**
VNP1	**1.24**	**0.98**	**1.49**	**7.20**	**9.52**	**3.35E-19**	**6.09E-17**
LPCAT1	**1.22**	**0.96**	**1.48**	**2.56**	**9.25**	**2.45E-18**	**3.84E-16**
GBP7	**2.40**	**1.95**	**2.85**	**6.62**	**9.25**	**2.45E-18**	**3.84E-16**
FGL2	**1.83**	**1.53**	**2.13**	**2.80**	**11.00**	**6.46E-28**	**2.17E-25**
hsa-miR-95-5p	**1.83**	**1.52**	**2.13**	**7.87**	**11.91**	**9.54E-28**	**6.08E-25**
hsa-miR-127-5p	**-0.74**	**-0.98**	**-0.57**	**6.29**	**-9.34**	**2.11E-17**	**1.79E-15**
CMTM6	**1.09**	**0.85**	**1.34**	**1.90**	**9.30**	**5.07E-17**	**7.20E-15**
MOSC2	**1.08**	**1.11**	**1.81**	**2.75**	**9.30**	**5.07E-17**	**7.20E-15**
hsa-miR-302-5p	**0.80**	**0.62**	**0.98**	**5.15**	**8.73**	**3.78E-17**	**5.35E-15**
IRF1	**0.98**	**0.74**	**1.22**	**8.17**	**8.68**	**1.14E-16**	**9.88E-15**
hsa-miR-148b-3p	**0.86**	**0.65**	**1.06**	**4.49**	**8.68**	**1.14E-16**	**1.56E-14**
hsa-miR-210-3p	**0.62**	**-0.77**	**-0.48**	**6.33**	**-8.48**	**8.99E-16**	**2.71E-14**
hsa-miR-3614-5p	**1.09**	**1.04**	**1.85**	**5.15**	**8.25**	**1.70E-15**	**1.23E-13**
STC11	**0.93**	**0.71**	**1.15**	**2.08**	**8.27**	**1.68E-15**	**1.68E-13**
PTPN1	**0.96**	**0.73**	**1.19**	**2.11**	**8.27**	**1.68E-15**	**1.68E-13**
ZNF106	**1.01**	**0.77**	**1.25**	**3.49**	**8.24**	**1.61E-15**	**1.72E-13**
TRIM25	**1.08**	**0.78**	**1.27**	**3.49**	**8.24**	**1.61E-15**	**1.72E-13**
hsa-miR-4772-3p	**-1.06**	**-1.32**	**-0.81**	**4.54**	**-8.21**	**4.56E-15**	**1.45E-13**
DDX60L	**1.05**	**0.80**	**1.30**	**2.63**	**8.21**	**4.56E-15**	**5.42E-13**
CYB5R	**1.47**	**1.12**	**1.82**	**3.26**	**8.19**	**5.34E-15**	**6.08E-13**
ATP6V1B2	**1.17**	**1.00**	**1.44**	**3.26**	**8.19**	**5.34E-15**	**6.08E-13**
IRF1	**0.98**	**0.74**	**1.22**	**8.17**	**8.68**	**1.14E-16**	**9.88E-15**
GF2R	**0.71**	**0.61**	**1.01**	**4.49**	**8.68**	**1.14E-16**	**1.56E-14**
ADAR	**1.11**	**0.88**	**1.39**	**2.80**	**7.96**	**2.47E-14**	**2.17E-12**
SPTLC2	**0.84**	**0.65**	**1.04**	**3.68**	**7.93**	**2.70E-14**	**2.33E-12**
WASF2	**1.14**	**0.85**	**1.42**	**3.91**	**7.90**	**3.04E-14**	**6.29E-12**
CD74	**1.09**	**1.03**	**1.39**	**7.33**	**7.84**	**3.94E-14**	**5.52E-12**
KIAA0830	**0.74**	**0.53**	**0.95**	**2.66**	**7.80**	**5.37E-14**	**5.91E-12**
DNM1Z	**0.79**	**0.57**	**1.02**	**2.66**	**7.80**	**5.37E-14**	**5.91E-12**
hsa-miR-361-3p	**1.08**	**0.78**	**1.39**	**3.91**	**7.80**	**5.37E-14**	**5.91E-12**
MCL1	**1.08**	**0.78**	**1.39**	**6.94**	**7.80**	**5.37E-14**	**5.91E-12**
LIN7A	**0.99**	**1.71**	**1.27**	**2.61**	**7.80**	**5.37E-14**	**5.91E-12**
hsa-miR-150-5p	**-0.58**	**-0.75**	**-0.41**	**4.49**	**-7.80**	**5.37E-14**	**5.91E-12**
IQSEC1	**0.88**	**0.63**	**1.13**	**2.88**	**7.80**	**5.37E-14**	**5.91E-12**
LAMA01	**0.98**	**0.70**	**1.26**	**3.68**	**7.80**	**5.37E-14**	**5.91E-12**
AKAP13	**0.65**	**0.46**	**0.84**	**3.86**	**6.82**	**4.02E-11**	**1.95E-09**
hsa-miR-192-5p	**-0.67**	**-0.86**	**-0.48**	**4.49**	**-6.82**	**4.02E-11**	**1.95E-09**
hsa-miR-769-5p	**-0.53**	**-0.68**	**-0.37**	**7.49**	**-6.74**	**6.62E-11**	**1.26E-09**
CLTC	**0.82**	**0.58**	**1.06**	**3.68**	**6.74**	**6.62E-11**	**1.26E-09**
PSAP	**1.12**	**0.79**	**1.45**	**4.64**	**6.67**	**1.05E-10**	**4.75E-09**
hsa-miR-4772-5p	**-0.59**	**-1.14**	**-0.63**	**6.01**	**-6.62**	**1.34E-10**	**2.33E-09**
hsa-miR-6126-5p	**0.59**	**0.41**	**0.76**	**8.22**	**6.61**	**1.50E-10**	**2.66E-09**
hsa-miR-342-5p	**0.59**	**11.14**	**0.63**	**3.68**	**6.61**	**1.50E-10**	**2.66E-09**
TGOLN2	**0.89**	**0.62**	**1.16**	**2.87**	**6.59**	**1.66E-10**	**7.37E-09**
hsa-miR-196-5p	**-0.62**	**-0.81**	**-0.43**	**6.50**	**-6.50**	**2.86E-10**	**4.22E-09**
NACC2	**0.69**	**0.48**	**0.91**	**3.07**	**6.48**	**3.29E-10**	**1.45E-08**
C14orf73	**0.63**	**0.44**	**0.82**	**3.07**	**6.48**	**3.29E-10**	**1.45E-08**
ReverseComplement_FCGR2A	**1.03**	**0.71**	**1.36**	**8.77**	**6.36**	**6.58E-10**	**2.77E-08**
SLK8	**0.80**	**0.51**	**1.09**	**3.68**	**6.36**	**6.58E-10**	**3.85E-08**
ReverseComplement_SNX11	**-1.09**	**-1.44**	**-0.75**	**4.60**	**-6.27**	**1.10E-09**	**6.45E-08**
LOC401072	**0.81**	**0.56**	**1.07**	**3.68**	**6.27**	**1.10E-09**	**6.45E-08**
OSFRL1	**1.01**	**0.69**	**1.33**	**2.81**	**6.26**	**1.16E-09**	**6.54E-08**
IL7R	**-0.97**	**-1.27**	**-0.66**	**2.71**	**-6.21**	**1.20E-09**	**6.62E-08**
CEBPB	**0.92**	**0.63**	**1.21**	**3.13**	**6.21**	**1.23E-09**	**6.62E-08**
hsa-miR-873-5p	**-1.36**	**-1.71**	**-1.01**	**4.49**	**-6.21**	**1.23E-09**	**6.62E-08**
BACH1	**0.64**	**0.44**	**0.85**	**2.12**	**6.20**	**1.66E-09**	**6.17E-08**
ABHD2	**1.13**	**0.77**	**1.49**	**5.18**	**6.18**	**1.83E-09**	**6.76E-08**
CB9G	**0.79**	**0.54**	**1.05**	**2.24**	**6.15**	**2.14E-09**	**7.69E-08**
hsa-miR-1246	**0.74**	**0.51**	**0.98**	**2.24**	**6.15**	**2.14E-09**	**7.69E-08**
LOC100002288292.1	**0.80**	**0.54**	**1.06**	**2.31**	**6.13**	**2.36E-09**	**8.31E-08**
CPD	**0.74**	**0.50**	**0.97**	**2.31**	**6.13**	**2.36E-09**	**8.31E-08**
BZW	**0.96**	**0.65**	**1.27**	**4.80**	**6.10**	**2.85E-09**	**9.75E-08**
ReverseComplement_DNHD1	**-1.04**	**-1.38**	**-0.70**	**4.60**	**-6.10**	**2.85E-09**	**9.75E-08**
PPARGC1B	**0.71**	**0.48**	**0.93**	**2.73**	**6.05**	**3.70E-09**	**1.23E-07**
MBNL1	**0.74**	**0.49**	**0.98**	**3.10**	**6.03**	**4.12E-09**	**1.34E-07**
snoRNA:SNORA54	**-0.61**	**-0.81**	**-0.41**	**6.29**	**-5.91**	**6.77E-09**	**2.19E-07**
snoRNA:SNORD_105B	**-0.66**	**-0.91**	**-0.42**	**6.50**	**-5.91**	**6.77E-09**	**2.19E-07**
NBEAL2	**0.75**	**0.50**	**1.00**	**2.79**	**5.92**	**7.82E-09**	**2.46E-07**
MEGF9	**0.80**	**0.54**	**1.07**	**2.91**	**5.92**	**7.82E-09**	**2.46E-07**
ABHD2	**0.79**	**0.53**	**1.06**	**2.91**	**5.89**	**8.91E-09**	**2.75E-07**
PIM1K	**0.89**	**0.60**	**1.19**	**3.26**	**5.89**	**8.91E-09**	**2.75E-07**
hsa-miR-191-5p	**0.55**	**0.37**	**0.73**	**10.86**	**5.89**	**8.18E-09**	**1.25E-07**
hsa-miR-181a-5p	**-0.34**	**-0.45**	**-0.23**	**10.06**	**-5.89**	**8.18E-09**	**1.25E-07**
AP1G1	**0.66**	**0.44**	**0.87**	**2.64**	**5.87**	**1.04E-08**	**3.11E-07**
GAS7	**0.87**	**0.58**	**1.16**	**2.76**	**5.87**	**1.04E-08**	**3.11E-07**
DYNC1H1	**0.48**	**0.32**	**0.65**	**3.85**	**5.86**	**1.09E-08**	**3.17E-07**
SIN3	**0.74**	**0.49**	**0.99**	**3.26**	**5.86**	**1.09E-08**	**3.17E-07**
hsa-miR-29c-3p	**-0.74**	**-0.98**	**-0.49**	**7.49**	**-5.81**	**1.28E-08**	**3.68E-07**
LCP1	**1.05**	**0.69**	**1.40**	**3.65**	**5.82**	**1.34E-08**	**3.85E-07**
IRAK3	**0.84**	**0.56**	**1.12**	**3.01**	**5.81**	**1.47E-08**	**4.11E-07**
ReverseComplement_EIF2K	**0.91**	**1.22**	**0.60**	**4.37**	**5.79**	**1.64E-08**	**4.60E-07**
C6orf62	**0.95**	**0.63**	**1.27**	**1.95**	**5.79**	**1.62E-08**	**4.64E-07**
TMEM179B	**0.78**	**0.51**	**1.05**	**4.79**	**5.72**	**2.28E-08**	**6.32E-07**
hsa-miR-186-5p	**-0.43**	**-0.58**	**-0.28**	**11.68**	**-5.71**	**2.42E-08**	**2.89E-07**
LBP4	**0.72**	**0.47**	**0.97**	**2.91**	**5.71**	**2.47E-08**	**6.68E-07**
RASEF2	**0.79**	**0.51**	**1.06**	**2.91**	**5.68**	**2.91E-08**	**2.75E-07**
DOCK8	**0.70**	**0.46**	**0.94**	**2.77**	**5.68**	**2.91E-08**	**7.54E-07**
ASPA73	**1.11**	**0.72**	**1.49**	**1.68**	**5.66**	**3.18E-08**	**8.24E-07**
hsa-miR-29a-3p	**-0.68**	**-0.64**	**-0.31**	**7.52**	**-5.66**	**3.18E-08**	**4.03E-07**
ReverseComplement_lncRNA:LINC00434	**0.75**	**0.49**	**1.01**	**5.13**	**5.64**	**3.41E-08**	**9.18E-07**
DUSP6	**0.67**	**0.44**	**0.91**	**2.38**	**5.64**	**3.41E-08**	**1.31E-07**
snoRNA:SNORD_17	**-0.50**	**-0.68**	**-0.33**	**11.25**	**-5.62**	**3.86E-08**	**9.97E-07**
TUBA1B	**0.96**	**0.64**	**0.75**	**3.11**	**5.60**	**4.08E-08**	**1.06E-06**
APOL6	**0.78**	**0.51**	**1.05**	**2.90**	**5.60**	**4.02E-08**	**1.06E-06**
SMOX01	**0.74**	**0.47**	**0.99**	**2.56**	**5.60**	**4.02E-08**	**1.06E-06**
hsa-miR-142-5p	**-0.45**	**-0.61**	**-0.29**	**14.34**	**-5.58**	**4.09E-08**	**1.52E-07**
PDK3	**0.70**	**0.45**	**0.94**	**2.71**	**5.57**	**4.31E-08**	**1.25E-06**
GNS	**1.01**	**0.65**	**1.37**	**2.00**	**5.51**	**5.85E-08**	**1.39E-06**
hsa-miR-181-5p	**-0.74**	**-1.01**	**-0.48**	**6.40**	**-5.50**	**5.77E-08**	**1.37E-06**
hsa-miR-331-3p	**-0.48**	**-0.61**	**-0.31**	**12.87**	**-5.49**	**7.84E-08**	**8.15E-07**
SLU1	**0.65**	**0.42**	**0.89**	**2.96**	**5.48**	**8.17E-08**	**8.51E-06**
lncRNA:LINC02-AS1	**-1.30**	**-1.77**	**-0.84**	**4.74**	**-5.48**	**8.22E-08**	**1.92E-06**
hsa-miR-148-3p	**-0.58**	**-0.78**	**-0.37**	**13.60**	**-5.48**	**8.22E-08**	**8.23E-07**
hsa-miR-1275	**-0.90**	**-1.22**	**-0.57**	**5.01**	**-5.45**	**9.88E-08**	**9.57E-07**
hsa-miR-1255-5p	**-0.67**	**-0.91**	**-0.43**	**7.11**	**-5.45**	**1.57E-07**	**1.25E-06**
EIF1A1	**0.80**	**0.51**	**1.09**	**5.39**	**5.39**	**1.31E-07**	**2.07E-06**
SUDS3	**0.97**	**0.61**	**1.32**	**2.88**	**5.38**	**1.36E-07**	**3.11E-06**
TRIP12	**0.60**	**0.38**	**0.82**	**2.70**	**5.37**	**1.49E-07**	**1.37E-06**
hsa-miR-29b-3p	**-0.57**	**-0.76**	**-0.36**	**9.95**	**-5.35**	**1.63E-07**	**1.49E-06**
hsa-miR-200b-3p	**-0.37**	**-0.51**	**-0.23**	**6.40**	**-5.33**	**1.79E-07**	**1.58E-06**
SZRN1	**0.73**	**0.46**	**1.00**	**2.70**	**5.32**	**1.87E-07**	**4.18E-06**
KLF13	**0.51**	**0.32**	**0.69**	**3.45**	**5.30**	**2.05E-07**	**4.52E-06**
PHC1	**0.61**	**0.38**	**0.84**	**3.45**	**5.30**	**2.13E-07**	**4.66E-06**
UBP1	**0.65**	**0.41**	**0.89**	**4.20**	**5.29**	**2.14E-07**	**4.67E-06**
hsa-miR-146b-5p	**-0.61**	**-0.77**	**-0.34**	**11.68**	**-5.28**	**2.28E-07**	**1.97E-06**
MTHFD2	**0.69**	**0.43**	**0.94**	**2.23**	**5.25**	**2.68E-07**	**5.75E-06**
SETX	**0.76**	**0.48**	**1.07**	**2.71**	**5.25**	**2.68E-07**	**5.75E-06**
hsa-miR-92b-3p	**0.45**	**0.28**	**0.62**	**3.27**	**5.25**	**2.70E-07**	**2.26E-06**
hsa-miR-200c-3p	**-0.43**	**-0.59**	**-0.27**	**9.76**	**-5.24**	**2.81E-07**	**2.29E-06**
CTSB	**0.69**	**0.43**	**0.95**	**2.86**	**5.23**	**2.90E-07**	**6.15E-06**
hsa-miR-664-5p	**-0.45**	**-0.62**	**-0.28**	**9.21**	**-5.23**	**2.93E-07**	**2.33E-06**
PLOD2	**0.54**	**0.33**	**0.74**	**2.53**	**5.23**	**2.95E-07**	**6.25E-06**
NCTC12	**0.75**	**0.47**	**1.04**	**2.91**	**5.23**	**2.95E-07**	**6.08E-06**
snoRNA:SNORD_105	**-0.58**	**-0.80**	**-0.36**	**7.96**	**-5.17**	**4.01E-07**	**2.88E-06**
snoRNA:SNORD_4A	**-0.36**	**-0.50**	**-0.22**	**8.62**	**-5.11**	**5.49E-07**	**3.91E-06**
DIAPH1	**0.77**	**0.47**	**1.06**	**2.60**	**5.11**	**5.45E-07**	**1.11E-05**
EIF3A	**0.58**	**0.35**	**0.80**	**2.97**	**5.10**	**5.52E-07**	**1.11E-05**
lncRNA:GAS3.2-AS1	**1.10**	**0.68**	**1.53**	**2.05**	**5.08**	**6.35E-07**	**1.27E-05**
hsa-miR-99a-5p	**-0.58**	**-0.80**	**-0.35**	**8.62**	**-5.08**	**6.35E-07**	**5.28E-06**
hsa-miR-4772-5p	**-0.53**	**-0.32**	**-0.74**	**2.34**	**-5.05**	**7.25E-07**	**1.45E-05**
hsa-miR-1290	**-0.42**	**-0.59**	**-0.25**	**9.76**	**-5.05**	**7.25E-07**	**6.03E-06**
hsa-miR-4286	**0.64**	**0.39**	**0.89**	**3.92**	**5.03**	**7.97E-07**	**1.95E-06**
SNHG5	**0.53**	**0.32**	**0.74**	**2.34**	**5.03**	**7.25E-07**	**1.45E-05**
ACTB	**0.74**	**0.45**	**1.03**	**3.26**	**5.03**	**8.94E-07**	**7.43E-06**
NAIP	**0.69**	**0.42**	**0.96**	**2.34**	**4.97**	**1.09E-06**	**2.06E-05**
hsa-miR-378i-3p	**0.60**	**0.37**	**0.83**	**2.34**	**4.97**	**1.09E-06**	**2.06E-05**
WSB1	**0.56**	**0.34**	**0.78**	**2.36**	**4.95**	**1.18E-06**	**2.23E-05**
hsa-miR-9-5p	**0.39**	**0.23**	**0.54**	**4.91**	**4.93**	**1.27E-06**	**9.16E-06**
NAMPT	**0.76**	**0.46**	**1.07**	**2.75**	**4.92**	**1.31E-06**	**2.45E-05**
hsa-miR-1468-5p	**-0.57**	**-0.78**	**-0.36**	**9.95**	**-4.92**	**1.34E-06**	**9.66E-06**
hsa-miR-223-3p	**0.77**	**0.46**	**1.08**	**11.64**	**4.91**	**1.42E-06**	**2.68E-05**
hsa-miR-198-5p	**-0.38**	**-0.53**	**-0.23**	**5.60**	**-4.89**	**1.54E-06**	**1.11E-05**
hsa-miR-221-5p	**-0.25**	**-0.35**	**-0.15**	**7.35**	**-4.89**	**1.54E-06**	**1.04E-05**
hsa-miR-29a-3p	**0.54**	**0.32**	**0.76**	**3.86**	**4.85**	**1.86E-06**	**2.13E-05**
lncRNA:lnc-CNTNAP5B-69:2	**-0.52**	**-0.73**	**-0.31**	**3.00**	**-4.84**	**1.96E-06**	**3.57E-05**
snoRNA:SNORD_35A	**-0.52**	**-0.73**	**-0.31**	**7.60**	**-4.84**	**1.96E-06**	**1.41E-05**
hsa-miR-1249-3p	**0.63**	**0.37**	**0.88**	**3.92**	**4.82**	**2.17E-06**	**1.85E-05**
ReverseComplement_rRNA:Homo_sapiens_mitochondrially_encoded_16S_RNA_(MT-RNR2)	**0.74**	**0.44**	**1.04**	**2.91**	**4.82**	**2.17E-06**	**3.88E-05**
SLC35A1	**0.67**	**0.38**	**0.97**	**2.64**	**4.82**	**2.17E-06**	**3.88E-05**
TPK1	**0.73**	**0.43**	**1.03**	**3.25**	**4.82**	**2.17E-06**	**3.88E-05**
hsa-miR-24-3p	**0.28**	**0.17**	**0.40**	**13.25**	**4.79**	**2.58E-06**	**1.62E-05**
hsa-miR-363-3p	**-0.37**	**-0.52**	**-0.22**	**11.68**	**-4.79**	**2.61E-06**	**1.62E-05**
snoRNA:SNORD_116-23	**-0.67**	**-0.94**	**-0.39**	**4.52**	**-4.77**	**2.72E-06**	**4.78E-05**
hsa-let-7c-5p	**-0.52**	**-0.73**	**-0.30**	**7.60**	**-4.77**	**2.72E-06**	**1.41E-05**
snoRNA:SNORD_116-19	**-0.50**	**-0.70**	**-0.29**	**3.96**	**-4.77**	**2.77E-06**	**4.88E-05**
CNOT1	**0.63**	**0.37**	**0.89**	**2.86**	**4.77**	**2.85E-06**	**5.14E-05**
ARRB1	**0.54**	**0.33**	**0.80**	**3.38**	**4.76**	**2.96E-06**	**5.26E-05**
CTLD	**0.46**	**0.27**	**0.65**	**2.79**	**4.74**	**3.15E-06**	**5.45E-05**
HSPSSB1	**0.76**	**0.44**	**1.08**	**2.93**	**4.72**	**3.58E-06**	**5.98E-05**
CTSC	**0.61**	**0.38**	**0.93**	**3.38**	**4.69**	**4.03E-06**	**6.81E-05**
ETL	**0.72**	**0.42**	**1.02**	**3.54**	**4.68**	**4.09E-06**	**6.92E-05**
snoRNA:SNORD_85	**-0.62**	**-0.62**	**-0.21**	**11.99**	**-4.68**	**4.12E-06**	**7.05E-05**
hsa-miR-6168-5p	**0.63**	**0.35**	**0.92**	**3.09**	**4.68**	**4.24E-06**	**2.53E-05**
CZD11	**0.56**	**0.32**	**0.80**	**3.09**	**4.68**	**4.24E-06**	**7.25E-05**
hsa-miR-1207-5p	**-0.60**	**-0.56**	**-0.23**	**4.35**	**-4.66**	**4.56E-06**	**2.64E-05**
hsa-miR-218-1-3p	**0.42**	**0.25**	**0.60**	**3.85**	**4.65**	**4.68E-06**	**7.98E-05**
CCDC12A	**0.69**	**0.40**	**0.99**	**2.58**	**4.64**	**4.95E-06**	**8.31E-05**
hsa-miR-6511-5p	**-0.36**	**-0.52**	**-0.21**	**5.60**	**-4.64**	**4.95E-06**	**3.21E-05**
snoRNA:SNORD_81L3	**-0.66**	**-0.95**	**-0.38**	**4.97**	**-4.63**	**5.21E-06**	**8.52E-05**
hsa-miR-421-3p	**0.51**	**0.31**	**0.71**	**10.34**	**4.63**	**5.21E-06**	**3.02E-05**
ReverseComplement_GINAP4	**0.79**	**0.45**	**1.12**	**5.15**	**4.62**	**5.47E-06**	**8.89E-05**
hsa-miR-1	**0.37**	**0.21**	**0.53**	**3.45**	**4.62**	**5.47E-06**	**3.17E-05**
lncRNA_LOC100130231	**1.03**	**0.59**	**1.48**	**2.41**	**4.56**	**7.22E-06**	**1.14E-04**
snoRNA:SNORD_38	**-0.45**	**-0.63**	**-0.26**	**11.99**	**-4.56**	**7.22E-06**	**1.14E-04**
snoRNA:SNORD_116-6	**-0.80**	**-1.14**	**-0.45**	**3.92**	**-4.53**	**7.80E-06**	**1.34E-04**
MOXD1L	**0.67**	**0.38**	**0.96**	**2.95**	**4.53**	**7.80E-06**	**1.23E-04**
ITGB1	**0.67**	**0.27**	**0.80**	**3.45**	**4.49**	**9.76E-06**	**1.53E-04**
FLII	**0.79**	**0.44**	**1.13**	**3.45**	**4.49**	**9.76E-06**	**1.53E-04**
BLM	**0.67**	**0.38**	**0.96**	**3.77**	**4.49**	**9.97E-06**	**1.56E-04**
TRIM41	**0.56**	**0.31**	**0.81**	**3.77**	**4.49**	**9.97E-06**	**1.56E-04**
FCN1	**0.57**	**0.32**	**0.82**	**3.51**	**4.47**	**1.07E-05**	**1.64E-04**
hsa-miR-181c-5p	**-0.43**	**-0.63**	**-0.25**	**10.34**	**-4.47**	**1.07E-05**	**1.64E-04**
ReverseComplement_mRnGAP31A	**-0.72**	**-1.03**	**-0.40**	**2.79**	**-4.43**	**1.26E-05**	**1.91E-04**
TUBB1	**0.79**	**0.44**	**1.14**	**3.45**	**4.43**	**1.26E-05**	**1.94E-04**
ReverseComplement_KCNJ14	**1.09**	**1.57**	**0.60**	**2.05**	**4.42**	**1.34E-05**	**2.03E-04**
GPATCH2L	**0.57**	**0.32**	**0.83**	**3.45**	**4.42**	**1.34E-05**	**2.03E-04**
S100A8	**0.94**	**0.52**	**1.36**	**2.09**	**4.40**	**1.47E-05**	**2.17E-04**
HIP2	**0.43**	**0.24**	**0.63**	**3.45**	**4.40**	**1.47E-05**	**2.17E-04**
snoRNA:SNORD_44	**-0.59**	**-0.84**	**-0.34**	**4.35**	**-4.39**	**1.52E-05**	**2.24E-04**
hsa-miR-23b-3p	**0.52**	**0.29**	**0.75**	**3.45**	**4.39**	**1.52E-05**	**2.24E-04**
CBS	**0.64**	**0.35**	**0.92**	**3.81**	**4.38**	**1.61E-05**	**2.35E-04**
GLG1	**0.58**	**0.32**	**0.85**	**3.06**	**4.37**	**1.66E-05**	**2.42E-04**
HLA-B	**0.76**	**0.42**	**1.11**	**4.68**	**4.37**	**1.68E-05**	**2.41E-04**
SYNGR3	**0.65**	**0.36**	**0.94**	**4.68**	**4.36**	**1.73E-05**	**2.47E-04**
NCF2	**0.71**	**0.39**	**1.03**	**2.27**	**4.36**	**1.86E-05**	**2.64E-04**
NDST1	**0.55**	**0.31**	**0.79**	**2.95**	**4.33**	**1.96E-05**	**2.77E-04**
hsa-miR-99-5p	**0.24**	**0.13**	**0.35**	**10.03**	**4.33**	**2.00E-05**	**1.02E-04**
hsa-miR-126-3p	**0.48**	**0.26**	**0.70**	**8.15**	**4.32**	**2.01E-05**	**1.02E-04**
ReverseComplement_lncRNA:LINC00434	**-1.15**	**-1.67**	**-0.63**	**3.73**	**-4.31**	**2.12E-05**	**2.97E-04**
lncRNA:SEMA6A_antisense_RNA_1_SEMA6A-AS2	**0.61**	**0.34**	**1.17**	**10.03**	**4.31**	**2.15E-05**	**3.01E-04**
hsa-miR-223-5p	**0.43**	**0.22**	**0.62**	**10.34**	**4.30**	**2.21E-05**	**1.11E-04**
snoRNA:SNORD_116-8	**-0.47**	**-0.68**	**-0.25**	**5.87**	**-4.30**	**2.24E-05**	**3.18E-04**

Ensembl name	logFC	CI.L	CI.R	AveExpr	t	P.Value	adj.P.Val
HLA-B	**0.76**	**0.42**	**1.11**	**4.68**	**4.37**	**1.68E-05**	**2.41E-04**
SYNGR3	**0.65**	**0.36**	**0.94**	**4.68**	**4.36**	**1.73E-05**	**2.47E-04**
NCF2	**0.71**	**0.39**	**1.03**	**2.27**	**4.36**	**1.86E-05**	**2.64E-04**
NDST1	**0.55**	**0.31**	**0.79**	**2.95**	**4.33**	**1.96E-05**	**2.77E-04**
hsa-miR-99-5p	**0.24**	**0.13**	**0.35**	**10.03**	**4.33**	**2.00E-05**	**1.02E-04**
hsa-miR-126-3p	**0.48**	**0.26**	**0.70**	**8.15**	**4.32**	**2.01E-05**	**1.02E-04**
ReverseComplement_lncRNA:LINC02009	**-0.55**	**-0.77**	**-0.33**	**3.64**	**-4.31**	**2.12E-05**	**2.97E-04**
snoRNA:SNORD63	**0.73**	**0.41**	**1.05**	**3.64**	**4.31**	**2.15E-05**	**3.01E-04**
hsa-let-7b-5p	**0.42**	**0.22**	**0.62**	**10.34**	**4.30**	**2.21E-05**	**1.11E-04**
snoRNA:SNORD_116-8	**-0.47**	**-0.68**	**-0.25**	**5.87**	**-4.30**	**2.24E-05**	**3.18E-04**

9.6 Appendix for Chapter 7

9.6.1 The COV studies: the first clinical trials of ChAdOx1 nCoV-19

The Oxford vaccine group undertook the first clinical trials of ChAdOx1 nCoV-19, assessing it's safety, immunogenicity and effectiveness. Samples from the COV studies were explored in a multi-omics study investigating the immune response to COVID-19 and the effect of prior ChAdOx1 nCoV-19 vaccination on that response. An outline of COV studies is presented here.

9.6.1.1 COV001: Phase I/II clinical trial in healthy adults

The first in-human trial of ChAdOx1 nCoV-19 was a phase 1 / 2, participant-blinded, multicentre, randomised controlled trial (ClinicalTrials.gov, NCT04324606, and with ISRCTN, 15281137)[78]. The co-primary end points were the safety and efficacy of symptomatic virologically confirmed COVID-19. Secondary outcomes included safety, reactogenicity, and immunogenicity of ChAdOx1 nCoV-19 and efficacy against hospital-attended COVID-19, death, and seroconversion against non-spike proteins.

One thousand seventy-seven healthy adult participants aged 18–55 years with no history of SARS-CoV-2 infection, were randomised to receive ChAdOx-1/nCoV-19 at a dose of 5×10^{10} viral particles or a control vaccine, meningococcal conjugate vaccine (MenACWY). All participants provided baseline blood samples. Based on preliminary immunogenicity results, a protocol amendment allowed for a booster dose of ChAdOx-1/nCoV-19 to be given no sooner than 28 days after the priming dose.

Participants were followed up to establish whether they developed COVID-19. Participants with a cough, fever or loss of sense of taste or smell were instructed to attend for a "symptomatic visit" within 24 hours. At this visit, a nose and throat swab was sent for SARS-CoV-2 PCR, and blood samples were taken. Participants reattended 7 days later for repeat SARS-CoV-2 PCR testing, and more blood samples were taken. Not all symptomatic visits were due to COVID-19; however, this was only known in retrospect as the SARS-CoV-2 PCR result took days to be reported. Thus, samples were also taken from participants with illnesses with similar symptoms to COVID-19 that SARS-CoV-2 did not cause.

The result of COV001showed that ChAdOx1 nCoV-19 had an acceptable safety and reactogenicity profile and induced antibodies against spike that neutralised SARS-CoV-2 *in-vitro*[78].

9.6.1.2 COV002: phase II/III trial in older adults and people with comorbidities

COV002 was a phase II/III trial similar in design to the COV001 above but enrolled adults aged up to 70 years old and included people with co-morbidities, providing they were medically controlled[79]. Participants with a high frailty score were excluded. Participants were randomly allocated to receive either one dose or two doses of either ChAdOx1 nCoV-19 or MenACWY. The coprimary outcomes of the trial were efficacy and safety. Five hundred and sixty participants were enrolled. The study showed that the vaccine was safe, tolerable and immunogenic in a broader, more representative group. The study contributed

towards efficacy estimates determined by pooling outcomes of CO002 with CO001, COV003 and COV005 studies (see below).

9.6.1.3 COV003 and COV005: phase III trials in non-UK settings

COV003 and COV005 were similar to the COV studies above but were conducted in Brazil and South Africa, respectively[566]. In South Africa, the control arm was a saline placebo.

The first interim efficacy results of ChAdOx1 nCoV-19 based on 11,636 participants from efficacy cohorts across all of the above studies showed 70·4% (95·8% CI 54·8–80·6) vaccine efficacy against symptomatic infection [566]. Efficacy was even higher against severe disease. Subsequent studies using real-world data showed vaccine effectiveness of 73% five weeks after vaccination in people aged 70 and over [567]. In addition, there was a further 37% reduction in hospital admissions with COVID-19[567].

The safety profile of ChAdOx1 nCoV-19 is excellent; however, a rare, life-threatening adverse reaction—vaccine-induced immune thrombotic thrombocytopenia—was noted after the commencement of ChAdOx1 nCoV-19 mass immunisation programmes. This incidence is around 1 in 67,302 to 518,181 vaccinations, after doses 1 and 2, respectively. Thrombotic or fatal events due to COVID-19 are much higher than vaccination[568].

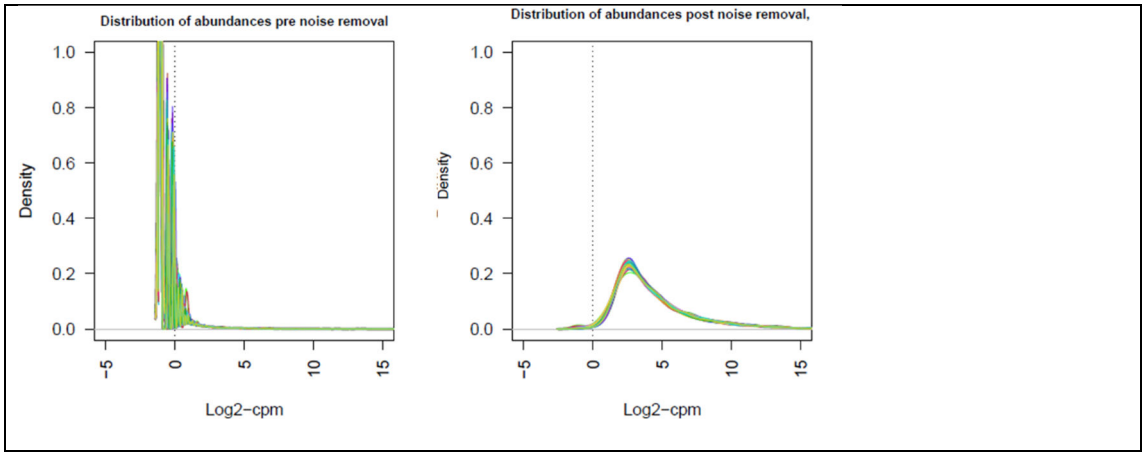
9.6.2 sRNA-seq QC

For a gene to be considered above the noise threshold, it had to have an abundance of at least 5 counts per million (equivalent to 6 reads in the smallest library) in at least 5 samples in stage I. For stage II these thresholds were 5 counts per million (7 reads in the smallest library in stage) in at least 20% of samples in stage II.

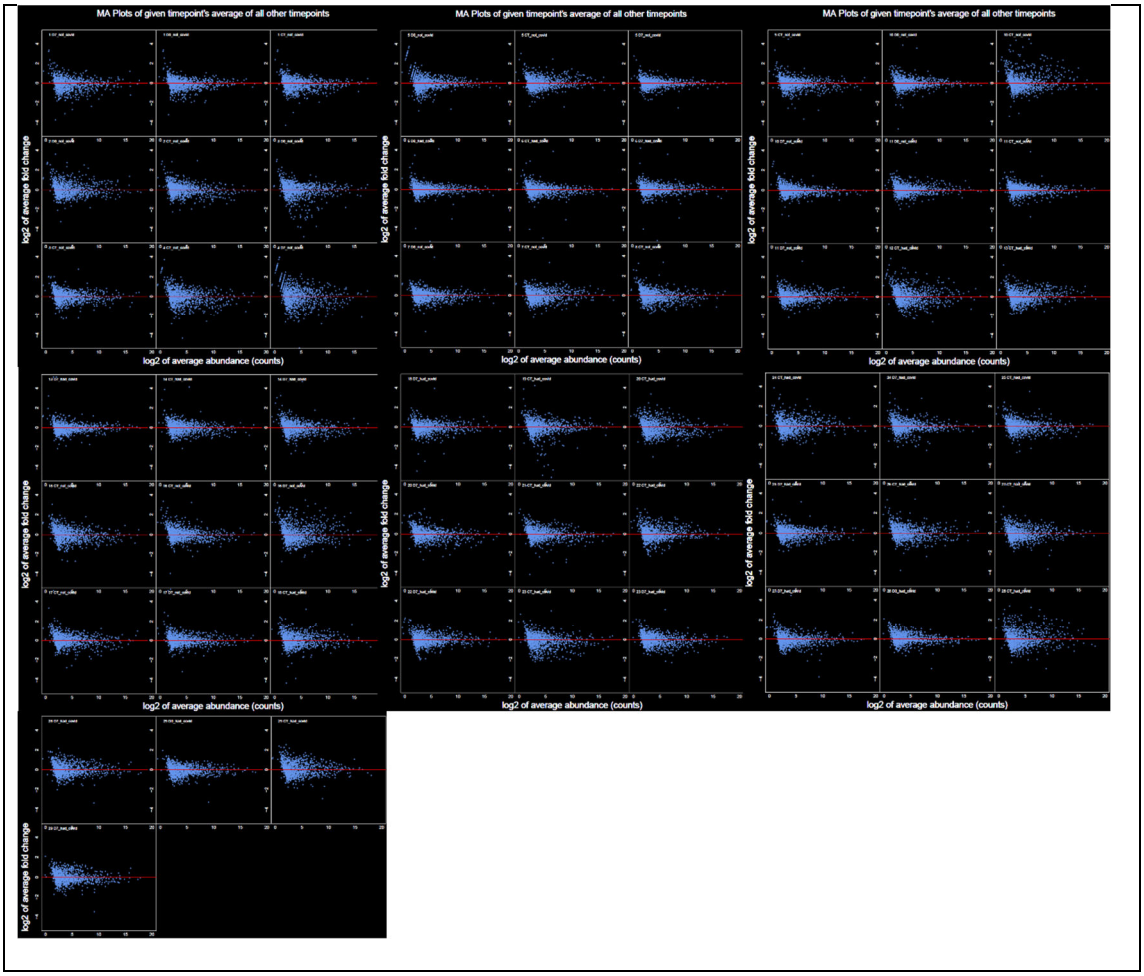
Stage I

Supplementary Table 9 ***Raw library sizes***

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1,502,578	3,765,994	4,333,886	4,25,3911	4,856,238	5,533,817



Supplementary figure 7 Stage I sRNA-seq Density Distribution of log CPM reads before (A) and after filtering out noise, and application of TMM normalisation (B). The large peak at low count level in (A) is absent in B suggesting that the noise removal and normalisaiton has worked well and there are no outliers



Supplementary figure 8 Stage I sRNA-seq M-A plots of each sample vs all others. Each dot represents the log_2 fold change of an sRNA feature vs the average expression of the sRNAs all other samples. If the removal of noise and the normalisation process is adequate, then median sRNA feature fold changes should be 0 for each sample - the red horizontal line (log_2 fold change = 0) should bisect the cloud of dots in half (i.e., there should be approximately an equal number of dots above the red line as below it). This is the case, suggesting that the noise removal and normalisation worked well and that there are no outliers.

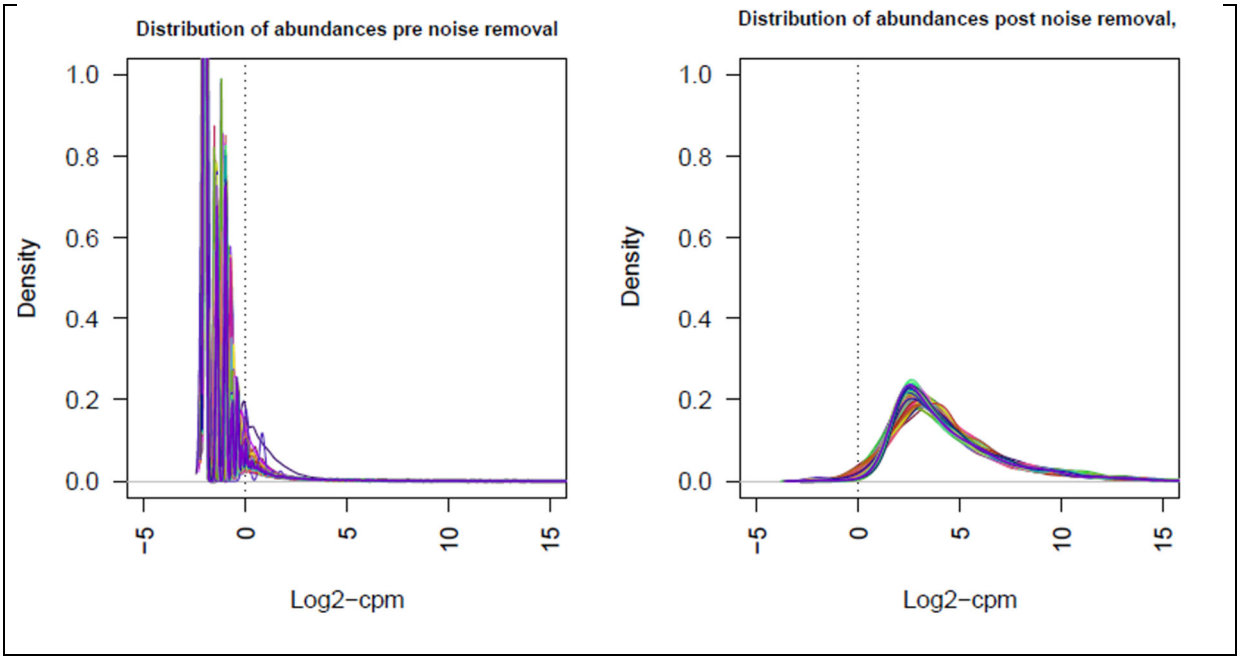
Stage II

Min. 1st Qu. Median Mean 3rd Qu. Max.

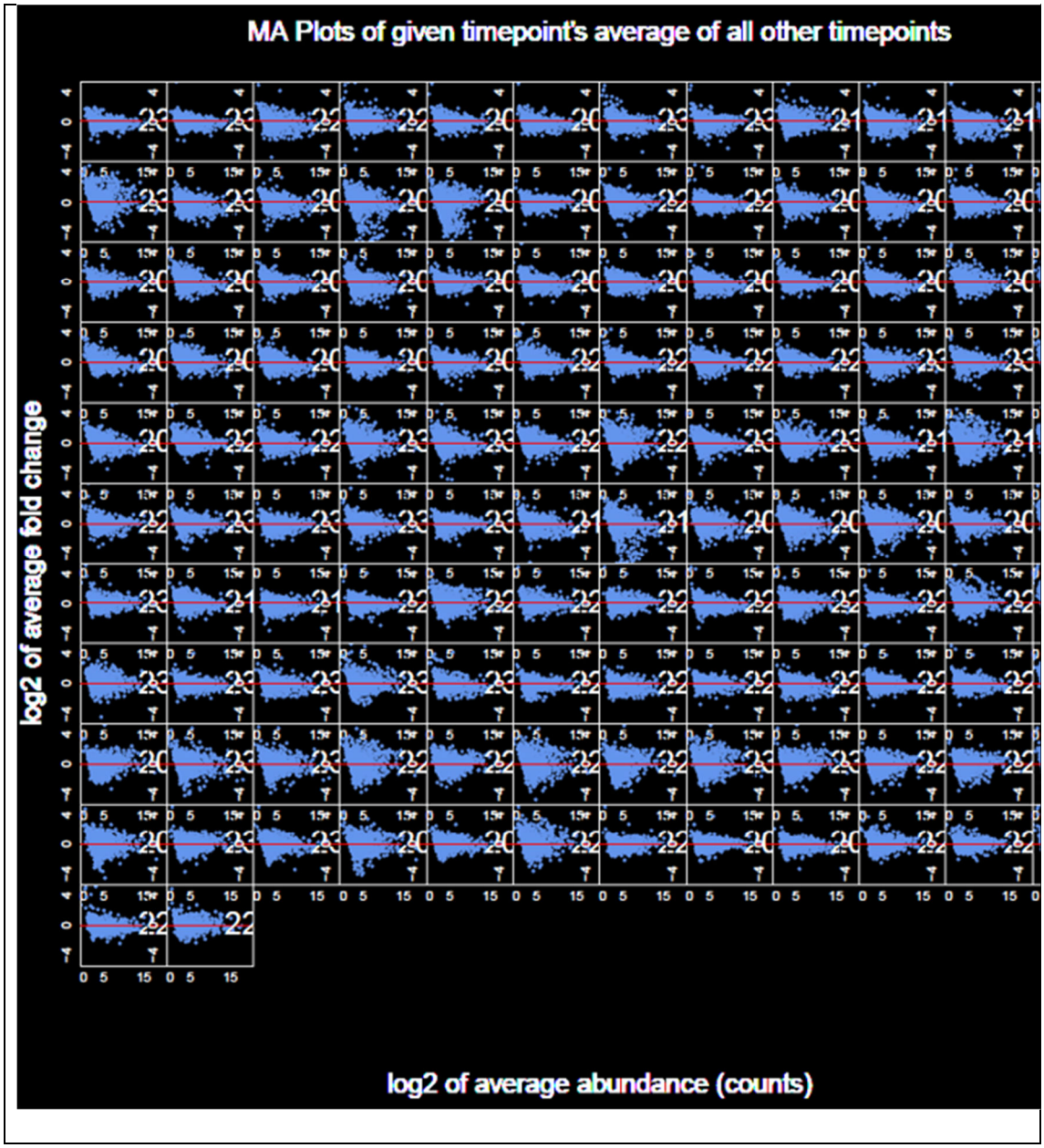
1,326,078	*7,522,174*	*8,004,520*	*8,050,032*	*9,503,816*	*1,1099,296*

Supplement

ary Table 10 Stage II Raw library sizes



Supplementary figure 9 Stage II sRNA-seq Density Distribution of log CPM reads before (A) and after filtering out noise, and application of TMM normalisation (B). The large peak at low count level in (A) is absent in B suggesting that the noise removal and normalisaiton has worked well and there are no outliers



Supplementary figure 10 Stage II sRNA-seq M-A plots of each sample vs all others. Each dot represents the log_2 fold change of an sRNA feature vs the average expression of the sRNAs all other samples. If the removal of noise and the normalisation process is adequate, then median sRNA feature fold changes should be 0 for each sample - the red horizontal line (log_2 fold change = 0) should bisect the cloud of dots in half (i.e., there should be approximately an equal number of dots above the red line as below it). This is the case,

suggesting that the noise removal and normalisation worked well and that there are no outliers.

Stage I demographics

Supplementary Table 11 Summary demographics for participants in stage I. All groups were demographically similar.

	COVID-like illness	menACWY	ChAdOx1 nCoV-19
mean BMI (kg)	28	29	26
Sex (%M)	28%	57%	57%
smoking (%)	0%	0%	14%
mean weight (Kg)	82	84	82
mean Age (years)	44	35	33
Alcohol (% drink alcohol)	100%	77%	100%

Supplementary Table 12 Stage I Individual level demographic data

covid_diagnosis	ht	wt	smoking	alcohol	age_yr	sex.mf	BMI	vaccine
not_covid	158	83	0	1	56.3	F	33.24788	ChAdOx1
not_covid	158	68.8	0	1	49.7	F	27.55969	MenACW\
not_covid	168	75	0	1	48	M	26.57313	MenACW\
not_covid	174	81	0	1	27	M	26.75386	ChAdOx1
not_covid	168	67	0	1	36	F	23.73866	MenACW\
had_covid	183	82.6	0	1	21	M	24.66482	ChAdOx1
not_covid	167	61.4	0	1	41.9	F	22.01585	MenACW\
not_covid	172	139.5	0	1	32	F	47.15387	MenACW\
not_covid	197	119	0	1	23	M	30.66299	ChAdOx1
not_covid	184	81	0	1	70	M	23.92486	ChAdOx1
not_covid	161	62.85	0	1	71	F	24.24675	ChAdOx1
had_covid	175	89	0	1	32	M	29.06122	ChAdOx1
had_covid	177	89.5	0	1	49	M	28.56778	ChAdOx1
had_covid	176	87.8	1	1	23	F	28.34452	ChAdOx1
not_covid	168	66	0	1	45	F	23.38435	MenACW\
not_covid	183	82.8	0	1	40	M	24.72454	MenACW\
not_covid	165	74	0	1	31	F	27.1809	MenACW\
had_covid	160	65	0	0	40	F	25.39063	MenACW\
had_covid	177	116	0	0	34	F	37.0264	MenACW\
had_covid	178	61	0	1	32	M	19.25262	MenACW\
had_covid	191	80	0	1	44	M	21.92922	ChAdOx1
had_covid	170	66	0	1	34	M	22.83737	MenACW\
had_covid	170	85.4	0	1	45	F	29.55017	ChAdOx1
had_covid	183	76	0	1	20	F	22.69402	MenACW\
had_covid	157	101.9	0	1	51	F	41.34042	MenACW\
had_covid	180	120	0	1	52	M	37.03704	MenACW\
had_covid	173	81	0	1	29	M	27.06405	MenACW\
had_covid	165	61	0	1	19.8	F	22.40588	ChAdOx1
had_covid	165	69.4	0	1	19.8	F	25.49128	MenACW\

Stage II demographics

Supplementary Table 13 Summary deomographics for participants in stage II. There were no baseline samples for COVID-19 positive participants as they were enrolled later in the trial, therefore matched baseline samples were selected from participants matched by demographics. All groups were demographically similar.

	Baseline	menACWY	ChAdOx1 nCoV-19
mean BMI (kg)	28	29	29
Sex (%M)	33%	29%	29%
smoking (%)	5%	6%	2%
mean weight (Kg)	78	84	86
mean Age (years)	41	38	39

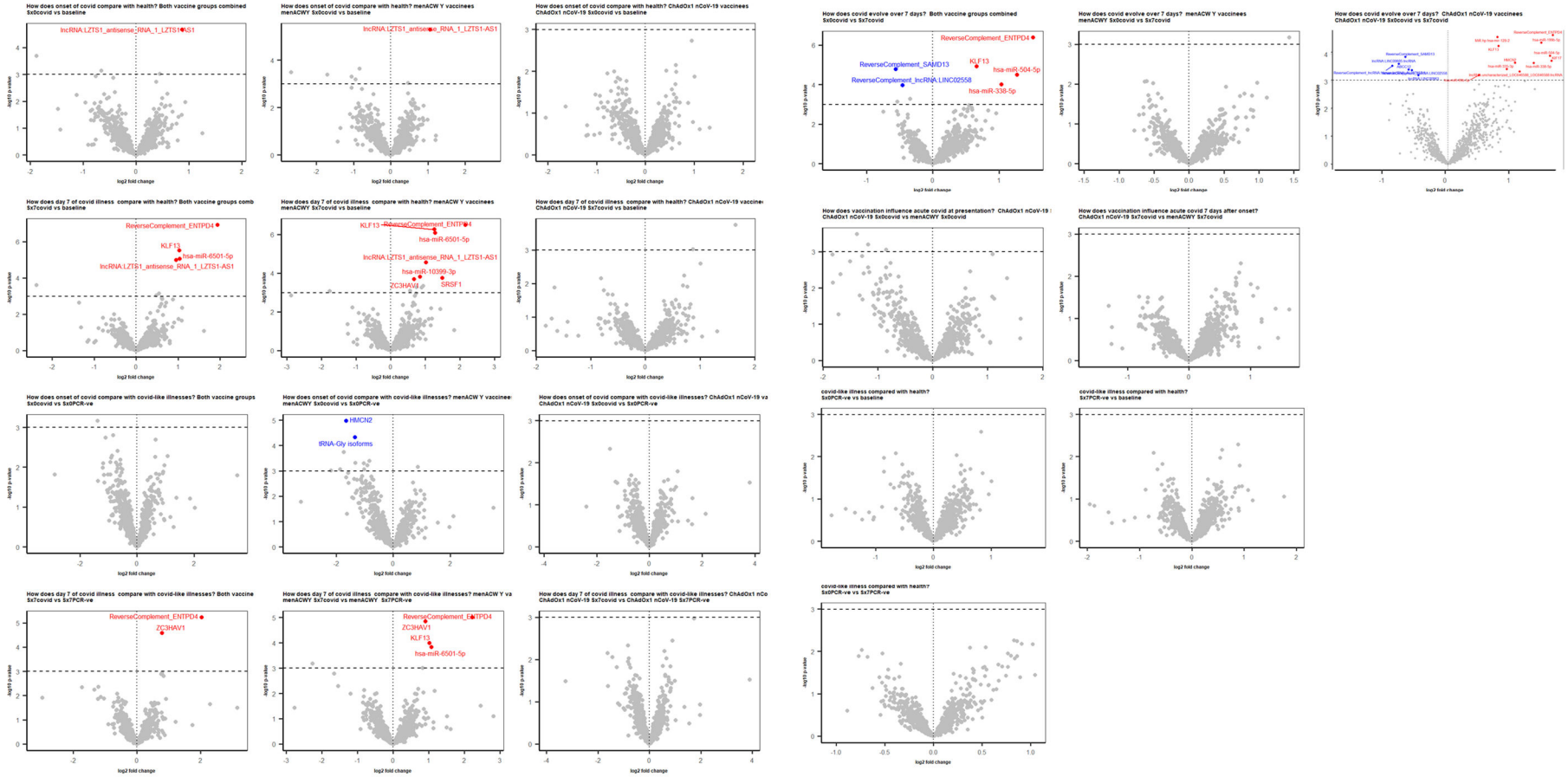
Supplementary Table 14 Stage II Individual level demographic data

covid_diagnosis	ht	wt	smoking	alcohol	age_yr	sex.mf	BMI	vaccine
had_covid	165	81	0	0	35	M	29.75206612	MenACWY
had_covid	163	97	0	0	24	F	36.50871316	MenACWY
had_covid	165	69	0	1	25	F	25.34435262	MenACWY
had_covid	173	77.15	0	1	51	F	25.77767383	ChAdOx1
had_covid	171	106	0	1	23	F	36.25047023	MenACWY
had_covid	166	67	0	1	36	F	24.31412397	MenACWY
had_covid	181	115.5	0	1	65	M	35.25533409	ChAdOx1
had_covid	179	99	0	1	37	F	30.89791205	ChAdOx1
had_covid	165	96.4	0	1	44	F	35.40863177	ChAdOx1
had_covid	181	73	0	1	33	M	22.28259211	ChAdOx1
not_covid	167	75	0	1	39.3	F	26.89232314	MenACWY
not_covid	163	65.2	0	1	57.1	F	24.5398773	ChAdOx1
not_covid	152	86.9	0	1	65	F	37.61253463	ChAdOx1
not_covid	166	74	0	1	64.3	F	26.85440557	ChAdOx1
not_covid	171	73.8	0	1	58.9	M	25.23853493	ChAdOx1
not_covid	152	51.3	0	0	63.4	F	22.20394737	MenACWY
not_covid	174	94	0	1	51.9	M	31.04769454	ChAdOx1
not_covid	165	69.8	0	1	24.9	F	25.63820018	ChAdOx1
not_covid	154	88	0	1	51.5	F	37.10575139	MenACWY
not_covid	174	91.8	0	1	26.5	M	30.32104637	ChAdOx1
not_covid	164	56	0	1	22	F	20.82093992	MenACWY
not_covid	162	64.4	0	1	35	F	24.53894223	ChAdOx1
not_covid	183	66	0	1	37	F	19.70796381	MenACWY
not_covid	193	126	1	1	41	M	33.82641145	MenACWY
not_covid	174	63	0	0	34	M	20.80856124	ChAdOx1
not_covid	160	95.8	0	1	32	F	37.421875	MenACWY
not_covid	176	76.3	0	1	21	M	24.63197314	ChAdOx1
not_covid	169	75.8	0	1	29	F	26.53968699	ChAdOx1
not_covid	175	84	0	1	34	M	27.42857143	ChAdOx1
had_covid	159	95	0	1	49	F	37.57762747	MenACWY
had_covid	160	61.1	0	1	55.1	F	23.8671875	MenACWY
had_covid	173	80	0	1	30	M	26.7299275	MenACWY
had_covid	164	76	0	0	44	F	28.25698989	MenACWY
had_covid	158	78	0	1	42	F	31.24499279	MenACWY
had_covid	170	65.3	0	1	27	F	22.59515571	MenACWY
had_covid	172	113.5	0	1	46	F	38.36533261	MenACWY
had_covid	159	55	0	1	36	F	21.75546853	ChAdOx1
had_covid	159	68.2	0	1	44	F	26.97678098	MenACWY
had_covid	162	70	0	1	42	F	26.6727633	ChAdOx1
had_covid	187	114	0	1	36	M	32.60030313	ChAdOx1
had_covid	168	96.1	0	1	52	F	34.04903628	ChAdOx1
had_covid	167	85.1	0	1	37	F	30.51382265	MenACWY
had_covid	169	79.3	1	1	49	F	27.76513427	MenACWY
had_covid	168	86	0	1	53	F	30.47052154	MenACWY
had_covid	166	116	0	0	33	F	42.09609522	MenACWY
had_covid	164	99.7	0	1	53	F	37.0687091	MenACWY
had_covid	172	105	0	1	48	F	35.49215792	MenACWY
had_covid	167	60	0	1	31	F	21.51385851	ChAdOx1
had_covid	163	59.15	0	1	24	F	22.26278746	MenACWY
had_covid	163	84.7	0	0	23	F	31.87925778	ChAdOx1
had_covid	174	110	0	1	20	M	36.33240851	MenACWY
had_covid	166	56.6	0	1	32	F	20.53999129	MenACWY
had_covid	176	129	0	1	40	F	41.64514463	ChAdOx1
had_covid	166	71	0	1	26	F	25.76571346	MenACWY
had_covid	181	99.3	0	1	42	M	30.31043008	MenACWY
had_covid	168	68.7	0	1	32.1	F	24.34098639	ChAdOx1
had_covid	165	57	0	1	41	F	20.93663912	MenACWY
had_covid	161	65.6	0	1	34	F	25.3076656	MenACWY
had_covid	183	71.3	0	1	28	M	21.29057302	MenACWY
had_covid	186	122.6	0	1	41	F	35.43762285	MenACWY
had_covid	177	83	0	0	55	F	26.49302563	MenACWY
had_covid	174	99.7	0	1	49	F	32.93037389	ChAdOx1
had_covid	181	69.9	0	1	19	M	21.33634504	ChAdOx1
had_covid	165	81.2	0	1	51	F	29.82552801	ChAdOx1
had_covid	169	77.94	1	1	42	M	27.28896047	ChAdOx1
had_covid	168	60.4	0	1	24	F	21.40022676	ChAdOx1
had_covid	174	72.65	1	1	42	F	23.99590435	MenACWY
had_covid	187	98.8	1	1	53	M	28.25359604	ChAdOx1
had_covid	164	89	0	1	29	F	33.09042237	ChAdOx1
had_covid	174	95	0	1	39	F	31.37798917	ChAdOx1
had_covid	189	96	0	0	34	M	26.87494751	MenACWY

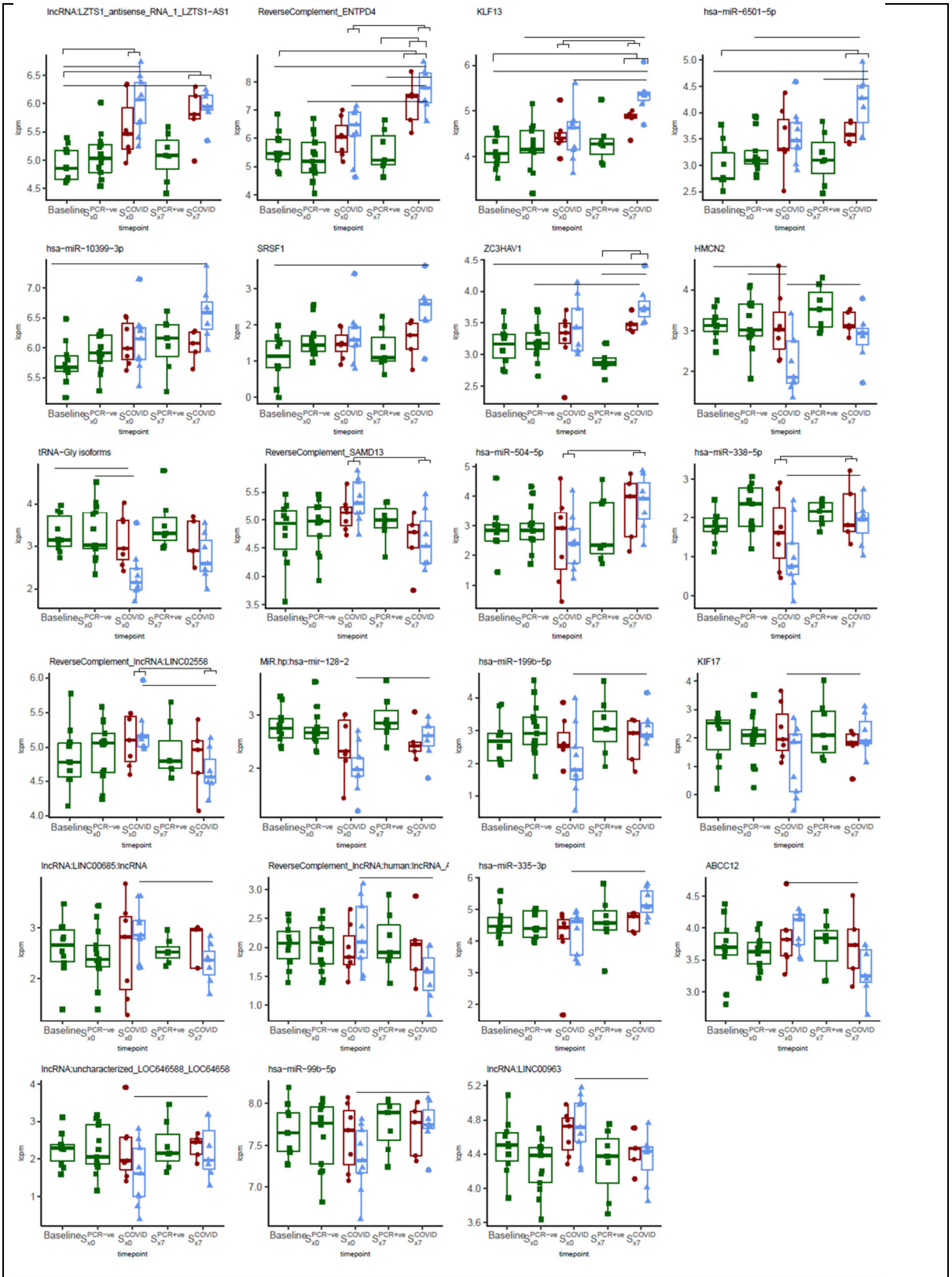
Appendix Table 11 Stage I Number of DE sRNAs in each comparison. This table summarises how many sRNA were DE within each sRNA biotype. This table complements the bar chart in

Question	Analysis	samples in analysis	mature miRNAs	offset miRNAs	tRNAs	lncRNAs	protein coding	Reverse complement	total
how does onset of COVID-19 compare with heath	$S_0^{COVID-19}$ vs baseline	27 (16 vs 11)	0	0	0	1	0	0	1
	menACWY $S_0^{COVID-19}$ vs baseline	20 (9vs11)	0	0	0	1	0	0	1
	ChAdOx1 nCoV-19 $S_0^{COVID-19}$ vs baseline	18 (7vs 11)	0	0	0	0	0	0	0
how does day 7 of COVID-19 illness compare with health?	$S_7^{COVID-19}$ vs baseline	24 (13 vs 11)	1	0	0	1	1	1	4
	menACWY $S_7^{COVID-19}$ vs baseline	17 (6 vs 11)	2	0	0	1	3	1	7
	ChAdOx1 nCoV-19 $S_7^{COVID-19}$ vs baseline	18 (7 vs 11)	0	0	0	0	0	0	0
how does onset of COVID-19 compare with COVID-like illnesses?	$S_0^{COVID-19}$ vs S_0^{PCR-ve}	29 (16 vs 13)	0	0	0	0	0	0	0
	menACWY $S_0^{COVID-19}$ vs S_0^{PCR-ve}	22 (9 vs 13)	0	0	1	0	1	0	2
	ChAdOx1 nCoV-19 $S_0^{COVID-19}$ vs S_0^{PCR-ve}	20 (7 vs 13)	0	0	0	0	0	0	0
how does day 7 of COVID-19 illness compare with COVID-like	$S_7^{COVID-19}$ vs S_7^{PCR-ve}	20 (13vs7)	0	0	0	0	1	1	2
	menACWY $S_7^{COVID-19}$ vs menACWY S_7^{PCR-ve}	13 (6vs7)	1	0	0	0	2	1	4

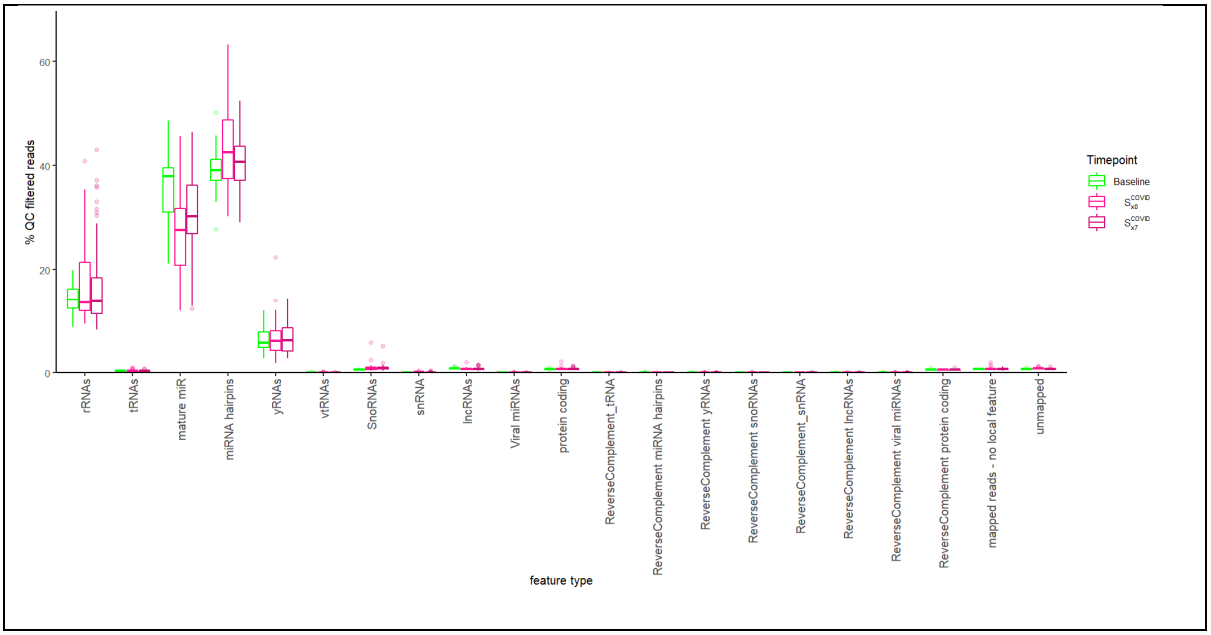
illnesses?	ChAdOx1 nCoV-19 $S_7^{COVID-19}$ vs ChAdOx1 nCoV-19 S_7^{PCR-ve}	14 (7vs7)	0	0	0	0	0	0	0
how does COVID-19 evolve over 7 days?	$S_0^{COVID-19}$ vs S_7^{COVID}	29 (16 vs 13)	2	0	0	0	1	3	6
	menACWY $S_0^{COVID-19}$ vs menACWY S_7^{COVID}	16 (9 vs 7)	5	1	0	3	4	4	17
	ChAdOx1 nCoV-19 $S_0^{COVID-19}$ vs ChAdOx1 nCoV-19 S_7^{COVID}	13 (7 vs 6)	0	0	0	0	0	0	0
How does vaccination influence acute COVID?	ChAdOx1 nCoV-19 $S_0^{COVID-19}$ vs menACWY S_0^{COVID}	16 (7 vs 9)	0	0	0	0	0	0	0
	ChAdOx1 nCoV-19 $S_7^{COVID-19}$ vs menACWY S_7^{COVID}	13 (6 vs 7)	0	0	0	0	0	0	0
COVID-like illness compared with health	S_0^{PCR-ve} vs baseline	19 (8 vs 11)	0	0	0	0	0	0	0
	S_7^{PCR-ve} vs baseline	18 (7 vs 11)	0	0	0	0	0	0	0
	S_0^{PCR-ve} vs S_7^{PCR-ve}	15 (8 vs 7)	0	0	0	0	0	0	0



Appendix Figure 45 Stage I differential expression results – summarised in



Appendix Figure 46—Dotplots of DE sRNAs in COVID study

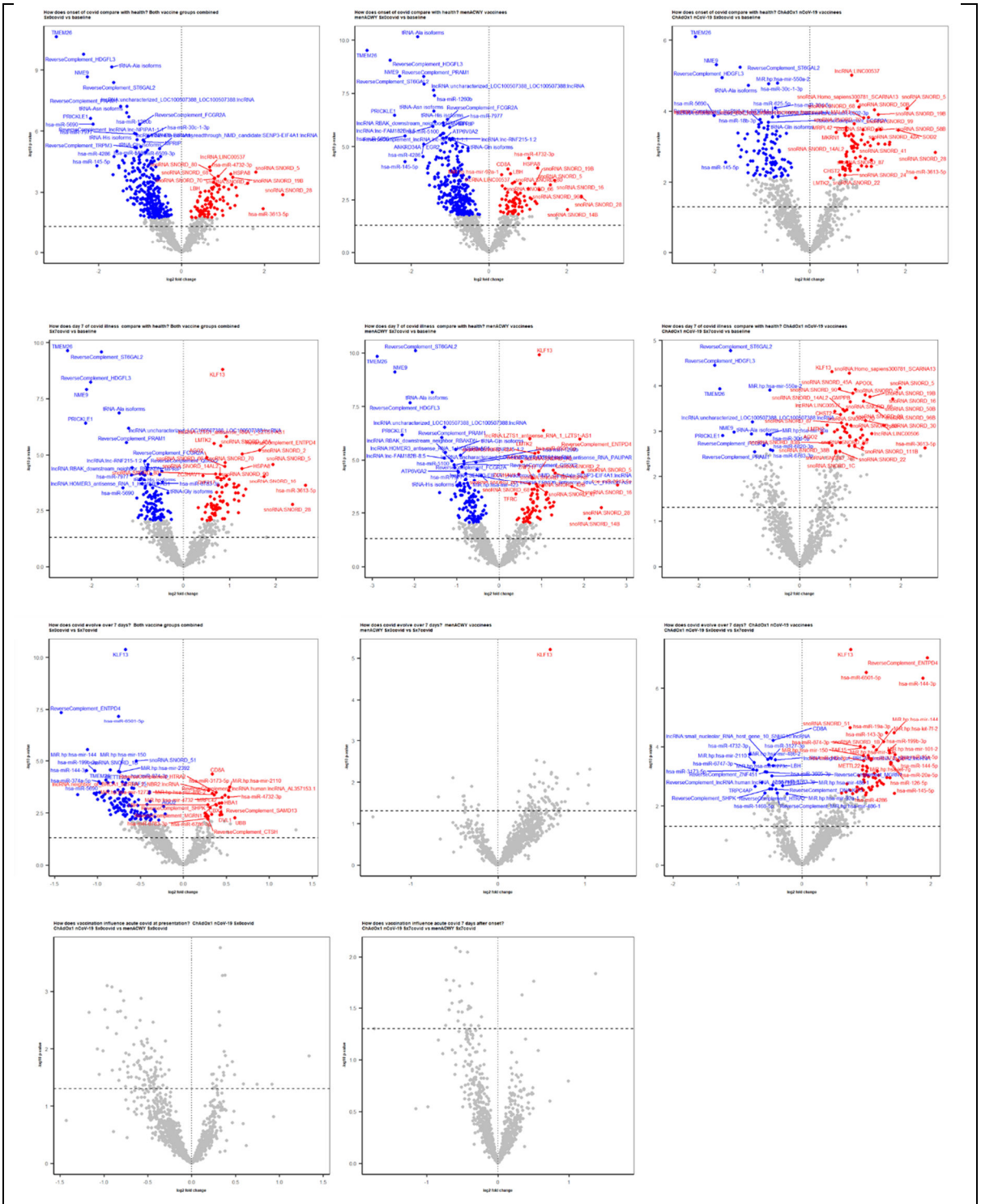


Appendix Figure 47 Stage II. Distribution of sRNA-seq reads mapping to each class of RNAs. There were no significant differences between the timepoint at the percentage mapping level. This meant all classess could be analysed together.

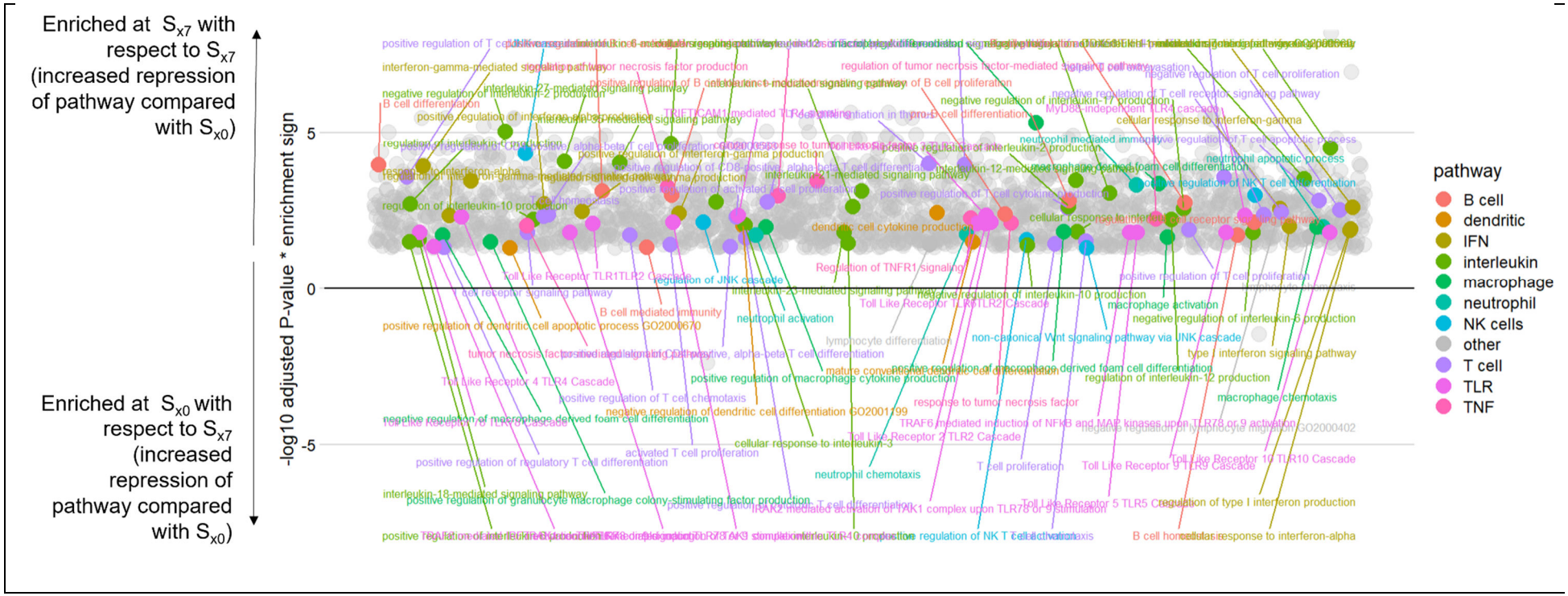
Appendix Table 12 Stage II: number of DE sRNAs in each comparison. This table summarises how many sRNA were DE within each sRNA biotype. This table complements the bar chart in Figure 95B.

Question	Timepoint	samples in analysis	mature miRNAs	off set miRNAs	tRNAs	lncRNAs	protein coding	Reverse complement	rRNA	snoRNA	vaultRNAs	total
How does onset of COVID-19 compare with heath?	$S_0^{COVID-19}$ vs baseline	70 (51 vs 19)	130	38	26	40	55	40	2	53	2	386
	menACWY $S_0^{COVID-19}$ vs baseline	49 (30 vs 19)	120	37	26	37	50	37	1	36	2	346
	ChAdOx1 nCoV-19 $S_0^{COVID-19}$ vs baseline	40 (21 vs 19)	49	15	10	10	23	10	3	50	0	170
How does day 7 of COVID-19 illness compare with health?	$S_7^{COVID-19}$ vs baseline	71 (52 vs 19)	27	10	20	26	27	15	1	66	0	192
	menACWY $S_7^{COVID-19}$ vs baseline	49 (30 vs 19)	26	11	22	28	27	17	0	62	0	193
	ChAdOx1 nCoV-19 $S_7^{COVID-19}$ vs baseline	40 (21 vs 19)	4	2	1	4	10	4	1	40	0	66
How does COVID-19 evolve over 7 days?	$S_0^{COVID-19}$ vs S_7^{COVID}	103 (51 vs 52)	94	22	4	3	15	12	0	9	0	159
	menACWY $S_0^{COVID-19}$ vs S_7^{COVID}	51 (21 vs 30)	0	0	0	0	1	0	0	0	0	1
	ChAdOx1 nCoV-19 $S_0^{COVID-19}$ vs S_7^{COVID}	61 (30 vs 31)	52	19	1	3	7	8	0	4	0	94

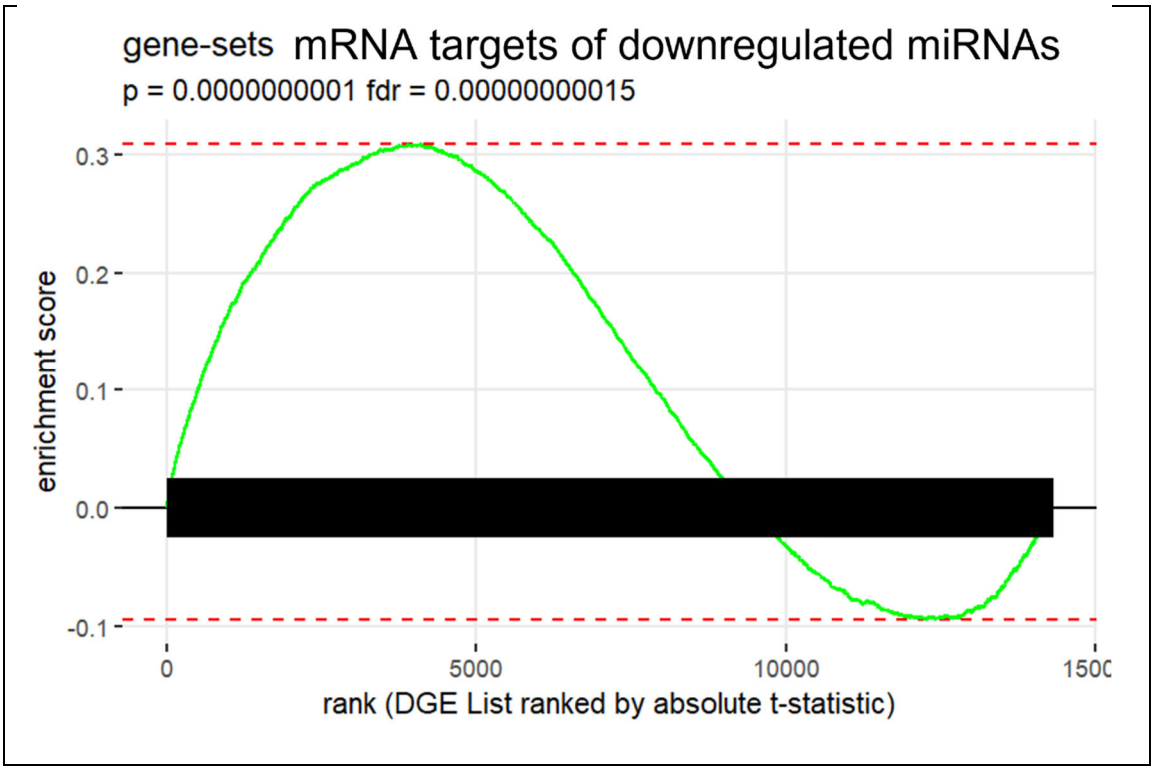
How does vaccination influence acute COVID?	ChAdOx1 nCoV-19 $S_0^{COVID-19}$ vs menACWY S_0^{COVID}	51 (30 vs 21)	0	0	0	0	0	0	0	0	0	0
	ChAdOx1 nCoV-19 $S_7^{COVID-19}$ vs menACWY S_7^{COVID}	52 (21 vs 31)	0	0	0	0	0	0	0	0	0	0



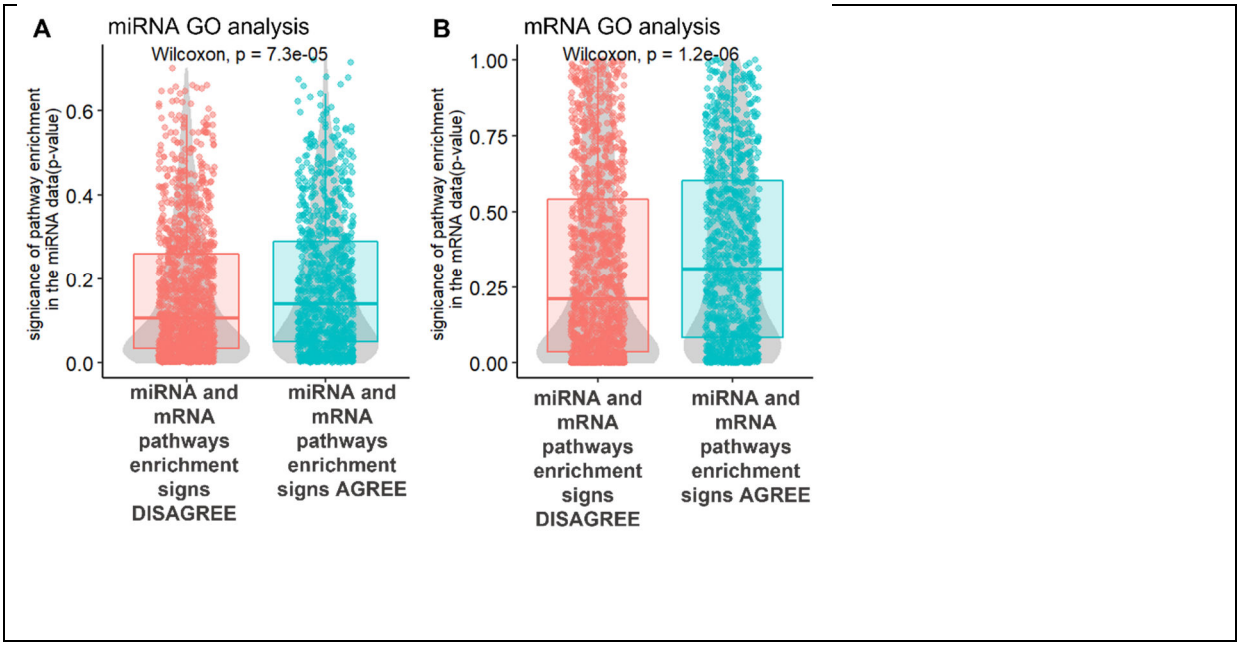
Appendix Figure 48 Stage II Differential expression results, summarised in Figure 95B.



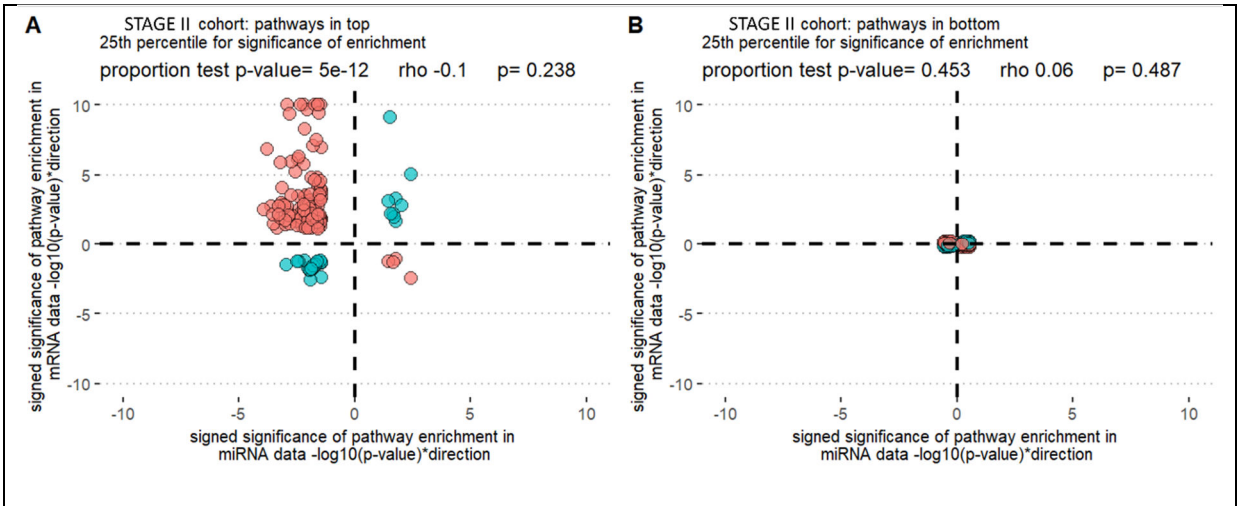
Appendix Figure 49 MIEA2 pathway enrichment analysis results run on the list of miRNAs that had been ranked by t-statistic comparing S_0 with S_7. Only FDR significant pathways are plotted. The pathways that were also differentially enriched at stage 1 are labelled.



Supplementary figure 11 Enrichment graph of mRNAs that are collectively targeted by all significantly downregulated miRNAs at S_0 vs baseline, based on the mRNA differential expression results ranked by t statistic. The mRNAs that are targeted by downregulated miRNAs are generally upregulated. This is in-line with miRNAs influencing gene expression during COVID-19.

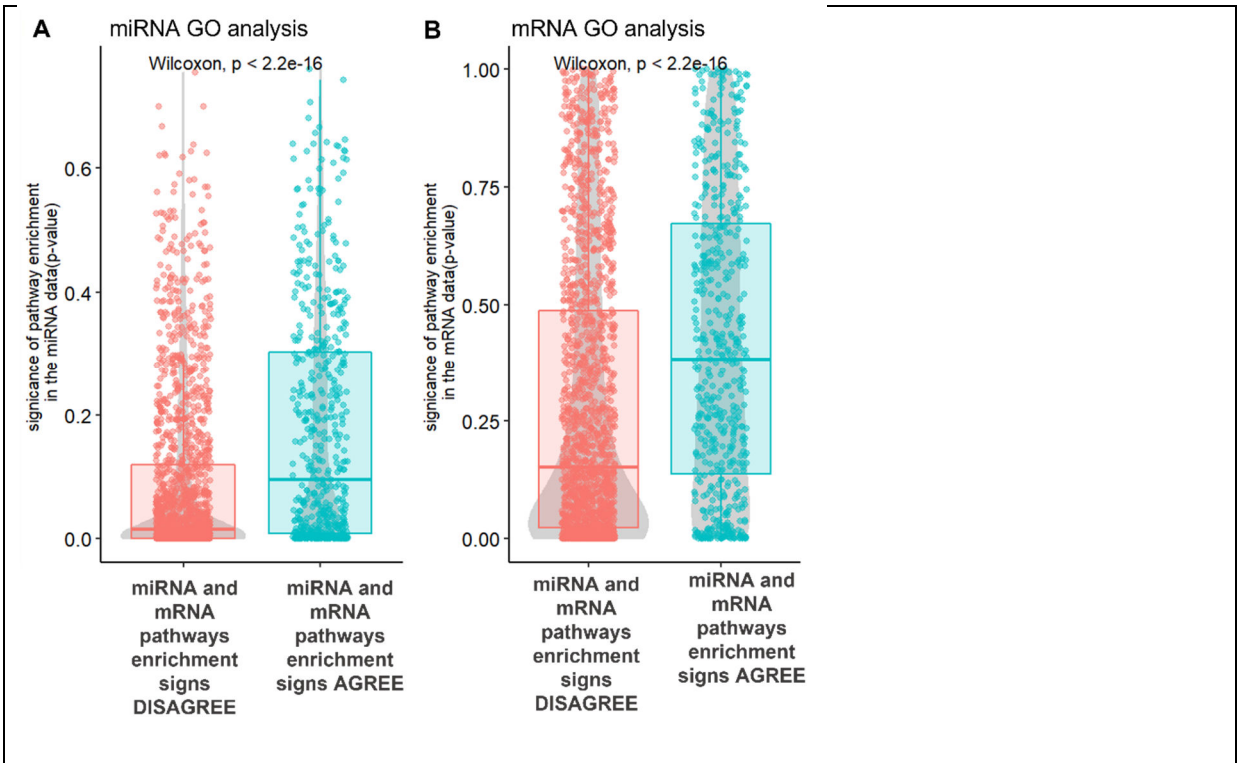


Appendix Figure 50 baseline versus COVID-19 symptom onset. GO enrichment results plotted by significance and enrichment sign agreement for the miRNA and mRNA data. Pathways, which who enrichment signs disagreed in the miRNA and mRNAs had more significant enrichment results in the miRNA (A) and mRNA (B) data.



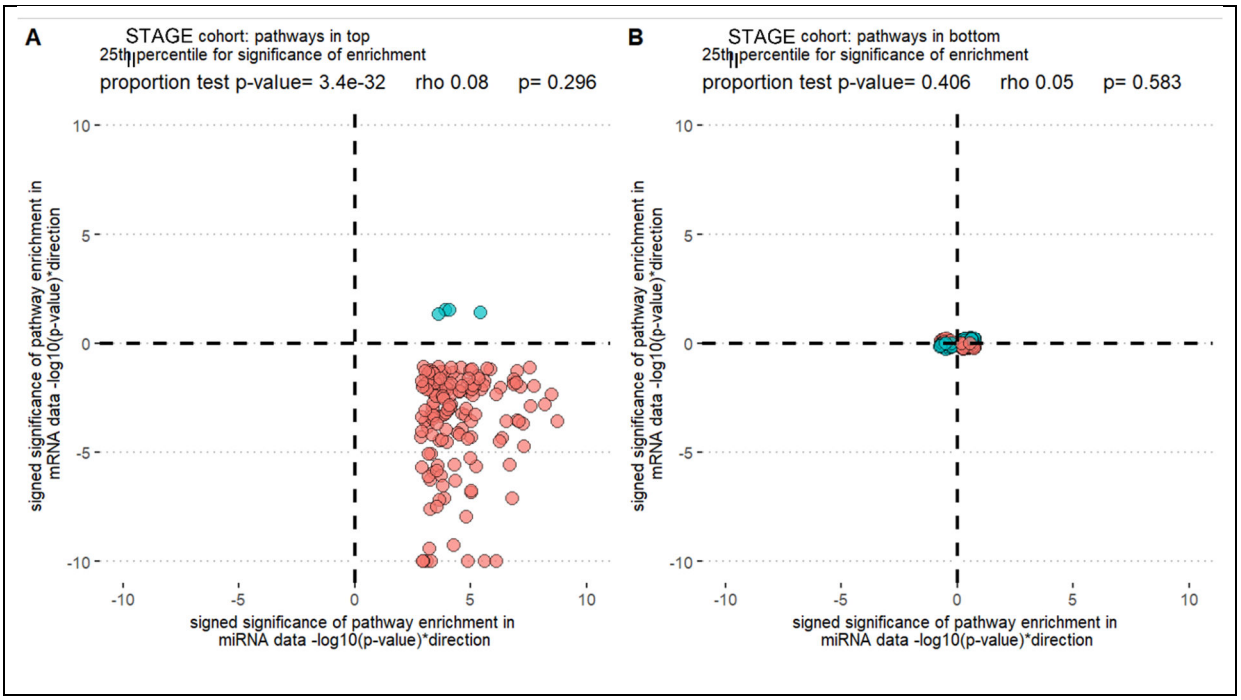
Appendix Figure 51 GO pathway analysis results at baseline vs COVID-19 symptom presented as an agreement plot between the miRNA and mRNA data for the most (A) and least (B) enriched pathays.

To shows that there is a relationship between significance of enrichment, an agreement plot was created showing the results for pathways with the most significantly enriched results (A) and least significant enriched results (B). Each dot represents the results of a pathway. The x-axis shows the log_{10} of the p-value of the pathway enrichment in the miRNA data. The y-axis shows the log_{10} of the p-value of the pathway enrichment in the mRNA data multiplied by the sign of enrichment (+1 if positive, -1 if negative). P-values are used because MIEAA does not give enrichment scores only significance values. The colour indicates whether the sign of miRNA enrichment and mRNA enrichment results for a pathway agree (blue) or disagree (red). This can also be gauged from the quadrants of the graph the pathway lies in. In panel A, most pathway enrichment signs disagree in the miRNA and mRNA and this disagreement was significant based on a proportion test. This was not the case for panel B.



Appendix Figure 52 COVID-19 symptom onset vs 7 days into illness. GO enrichment results plotted by significance and enrichment sign agreement for the miRNA and mRNA

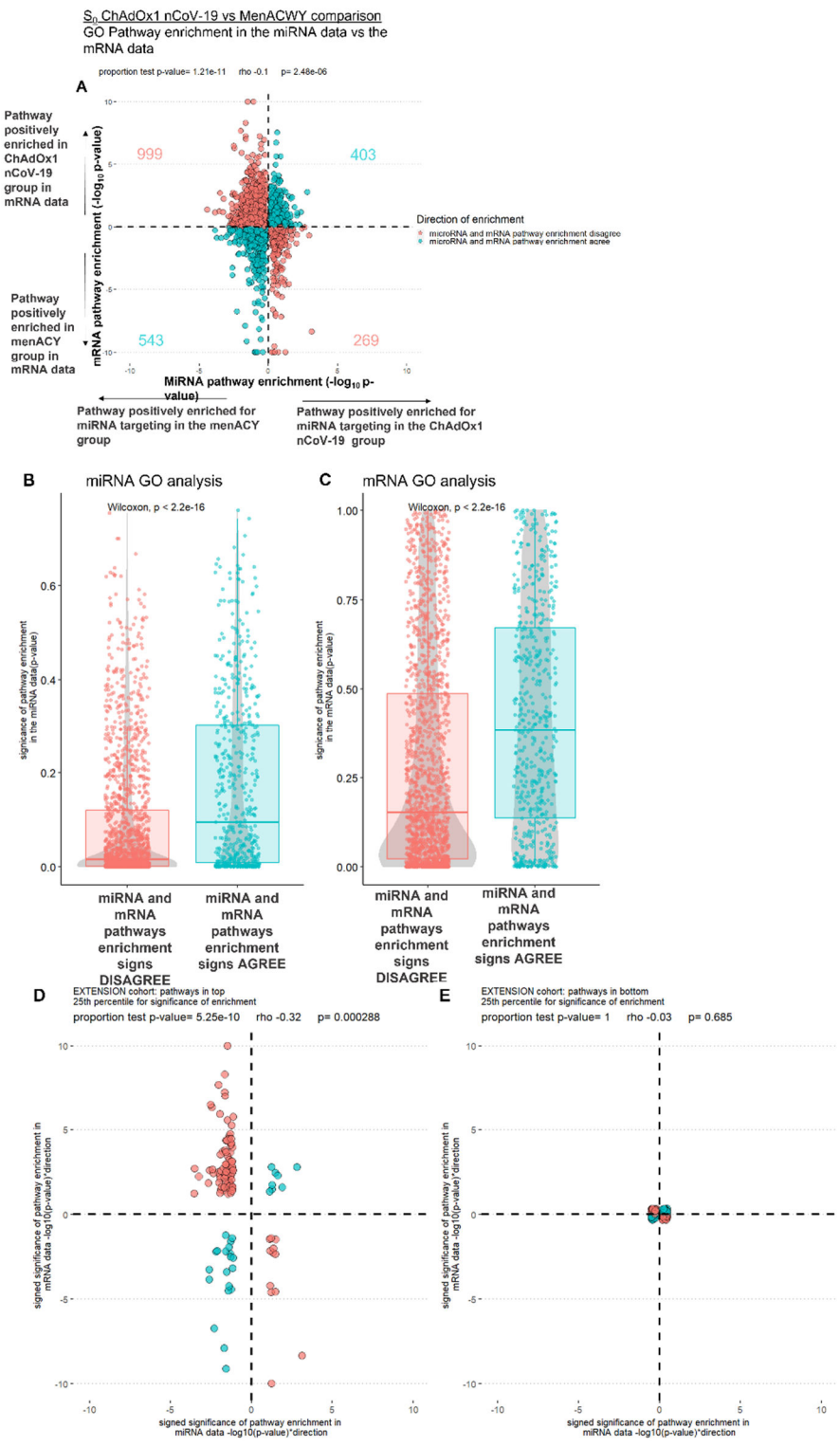
data. Pathways, which who enrichment signs disagreed in the miRNA and mRNAs had more significant enrichment results in the miRNA (A) and mRNA (B) data.



Appendix Figure 53 GO pathway analysis results at COVID-19 symptom onset vs 7 days later presented as an agreement plot between the miRNA and mRNA data for the most (A) and least (B) enriched pathays.

To shows that there is a relationship between significance of enrichment, an agreement plot was created showing the results for pathways with the most significantly enriched results (A) and least significant enriched results (B). Each dot represents the results of a pathway. The x-axis shows the log_{10} of the p-value of the pathway enrichment in the miRNA data. The y-axis shows the log_{10} of the p-value of the pathway enrichment in the mRNA data multiplied by the sign of enrichment (+1 if positive, -1 if negative). P-values are used because MIEAA does not give enrichment scores only significance values. The colour indicates whether the sign of miRNA enrichment and mRNA enrichment results for a pathway agree (blue) or disagree (red). This can also be gauged from the

quadrants of the graph the pathway lies in. In panel A, most pathway enrichment signs disagree in the miRNA and mRNA and this disagreement was significant based on a proportion test. This was not the case for panel B.



Appendix Figure 54 GO pathway analysis results at COVID-19 symptom onset in ChAdOx1 nCoV-19 vs control vaccine group showing, again, that there is antogonistic

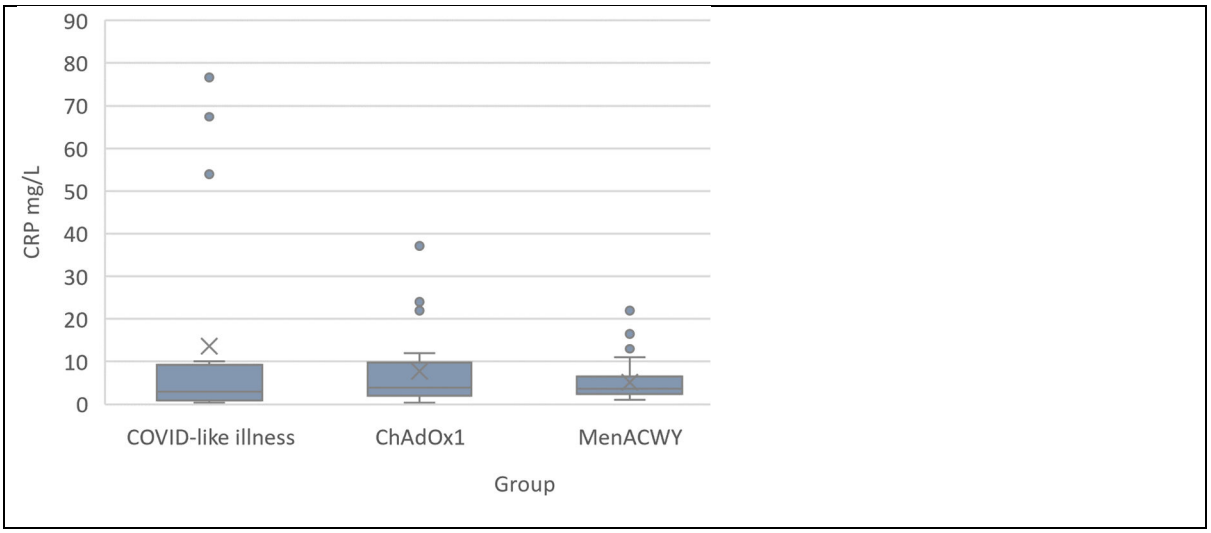
relationship between enrichment at the miRNA and mRNA level.

A: Each dot represents the results of a pathway. The x-axis shows the log_{10} of the p-value of the pathway enrichment in the miRNA data. The y-axis shows the log_{10} of the p-value of the pathway enrichment in the mRNA data multiplied by the sign of enrichment (+1 if positive, -1 if negative). P-values are used because MIEAA does not give enrichment scores only significance values. The colour indicates whether the sign of miRNA enrichment and mRNA enrichment results for a pathway agree (blue) or disagree (red). This can also be gauged from the quadrants of the graph the pathway lies in. The numbers in each quadrant show how many pathways fall in that quadrant. The results for all pathways tested in the miRNA and mRNA data is shown. For most pathways the enrichment signs in the miRNA and mRNA disagreed and this disagreement was significant based on a proportion test, and there was a weak but significant anti-correlation between strength of enrichment in the miRNA and mRNA data.

Panels B and C: GO enrichment results plotted by significance and enrichment sign agreement for the miRNA and mRNA data. Pathways, which who enrichment signs disagreed in the miRNA and mRNAs had more significant enrichment results in the miRNA (B) and mRNA (C) data.

Panels D and E To shows that there is a relationship between significance of enrichment, an agreement plot was created showing the results for pathways with the most significantly enriched results (D) and least significant enriched results (E). Each dot represents the results of a pathway. The x-axis shows the log_{10} of the p-value of the pathway enrichment in the miRNA data. The y-axis shows the log_{10} of the p-value of the pathway enrichment in the mRNA data multiplied by the sign of enrichment (+1 if positive, -1 if negative). P-values are used because MIEAA does not give enrichment scores only significance values. The colour indicates whether the sign of miRNA enrichment and mRNA enrichment results for a pathway agree (blue) or disagree (red). This can also be gauged from the quadrants of the graph the pathway lies in. In panel

D, most pathway enrichment signs disagree in the miRNA and mRNA and this disagreement was significant based on a proportion test. This was not the case for panel E.



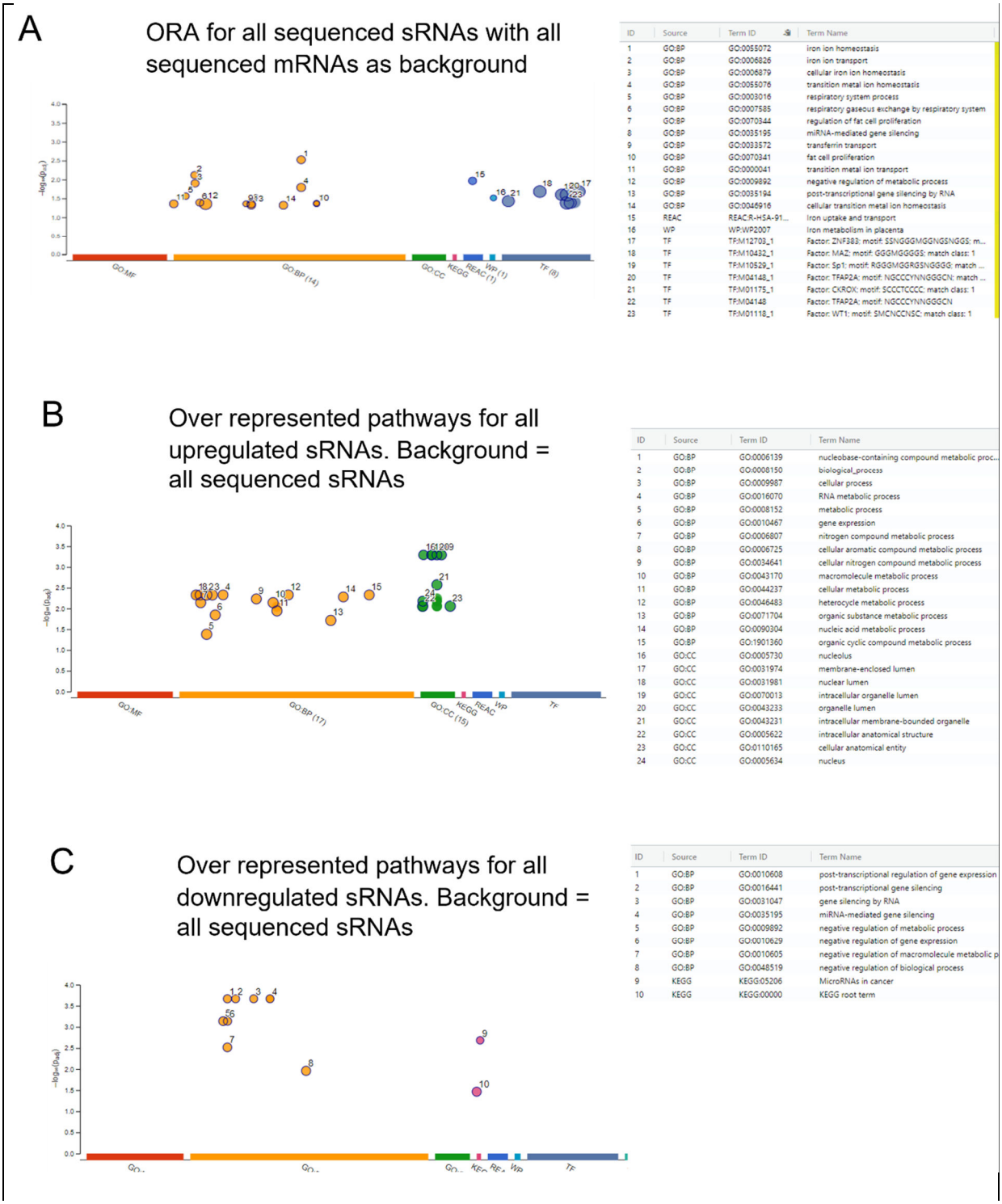
Supplementary figure 12 CRP results at S_0 for Stage I and Stage II combined. CRP measured through hospital assay.

9.6.3 differentially expressed sRNAs arise from genes involved in metabolism and RNA silencing, but there is little relationship between mRNA and sRNA expression

The function of many sRNAs is still unknown. Pathway over representation analysis (ORA) was done by querying all sRNAs against all annotated genes to gain further insights. This aimed to see whether sRNAs captured in sRNA-seq arise from genes with particular functions. To some extent, this is expected, e.g., splicing, given that snoRNA fragments are relatively common in sRNA seq data, but it was interesting to see whether other functions were also common.

ORA revealed enrichment of terms relating to post-transcriptional gene silencing, metabolic processes pertaining to RNA, nucleotides and organic compounds, and iron homeostasis and transport. Though these terms did not appear blood cell-specific, gene set enrichment analysis was repeated, querying all sRNAs using only mRNAs detected in the sequencing data as background. This avoids enrichment of terms that reflects the cellular sources. Terms mainly related to iron homeostasis and, curiously, the respiratory system were enriched (Figure 108 A).

After characterising the functions of the genes from, which sRNAs were derived, pathway enrichment analysis was done on DE sRNAs using the population of sequenced sRNAs as background (Figure 108 B&C). Upregulated sRNAs derive from full-length RNAs involved in metabolism. Downregulated sRNAs derived from full-length RNAs involved in RNA silencing but this was mainly due to offset miRNAs (e.g., sequences arriving from miRNA hairpins) being downregulated. Transcription factors were also enriched. Excluding miR and snoRNA-derived sRNAs led to no pathways being overrepresented amongst up or downregulated sRNAs.



ID	Source	Term ID	Term Name
1	GO:BP	GO:0055072	iron ion homeostasis
2	GO:BP	GO:0006826	iron ion transport
3	GO:BP	GO:0006879	cellular iron ion homeostasis
4	GO:BP	GO:0055076	transition metal ion homeostasis
5	GO:BP	GO:0003016	respiratory system process
6	GO:BP	GO:0007585	respiratory gaseous exchange by respiratory system
7	GO:BP	GO:0070344	regulation of fat cell proliferation
8	GO:BP	GO:0035195	miRNA-mediated gene silencing
9	GO:BP	GO:0033572	transferrin transport
10	GO:BP	GO:0070341	fat cell proliferation
11	GO:BP	GO:0000041	transition metal ion transport
12	GO:BP	GO:0009892	negative regulation of metabolic process
13	GO:BP	GO:0035194	post-transcriptional gene silencing by RNA
14	GO:BP	GO:0046916	cellular transition metal ion homeostasis
15	REAC	REAC:R-HSA-91...	Iron uptake and transport
16	WP	WP:WP2007	Iron metabolism in placenta
17	TF	TF:M12703_1	Factor: ZNF383: motif: SSNGGGMGGNGSNGGS: m...
18	TF	TF:M10432_1	Factor: MAZ: motif: GGGMGGGGS: match class: 1
19	TF	TF:M10529_1	Factor: Sp1: motif: RGGGMGGRGSNGGGG: match ...
20	TF	TF:M04148_1	Factor: TFAP2A: motif: NGCCCYNNGGGCN: match ...
21	TF	TF:M01175_1	Factor: CKROX: motif: SCCCTCCCC: match class: 1
22	TF	TF:M04148	Factor: TFAP2A: motif: NGCCCYNNGGGCN
23	TF	TF:M01118_1	Factor: WT1: motif: SMCNCCNSC: match class: 1

ID	Source	Term ID	Term Name
1	GO:BP	GO:0006139	nucleobase-containing compound metabolic proc...
2	GO:BP	GO:0008150	biological_process
3	GO:BP	GO:0009987	cellular process
4	GO:BP	GO:0016070	RNA metabolic process
5	GO:BP	GO:0008152	metabolic process
6	GO:BP	GO:0010467	gene expression
7	GO:BP	GO:0006807	nitrogen compound metabolic process
8	GO:BP	GO:0006725	cellular aromatic compound metabolic process
9	GO:BP	GO:0034641	cellular nitrogen compound metabolic process
10	GO:BP	GO:0043170	macromolecule metabolic process
11	GO:BP	GO:0044237	cellular metabolic process
12	GO:BP	GO:0046483	heterocycle metabolic process
13	GO:BP	GO:0071704	organic substance metabolic process
14	GO:BP	GO:0090304	nucleic acid metabolic process
15	GO:BP	GO:1901360	organic cyclic compound metabolic process
16	GO:CC	GO:0005730	nucleolus
17	GO:CC	GO:0031974	membrane-enclosed lumen
18	GO:CC	GO:0031981	nuclear lumen
19	GO:CC	GO:0070013	intracellular organelle lumen
20	GO:CC	GO:0043233	organelle lumen
21	GO:CC	GO:0043231	intracellular membrane-bounded organelle
22	GO:CC	GO:0005622	intracellular anatomical structure
23	GO:CC	GO:0110165	cellular anatomical entity
24	GO:CC	GO:0005634	nucleus

ID	Source	Term ID	Term Name
1	GO:BP	GO:0010608	post-transcriptional regulation of gene expression
2	GO:BP	GO:0016441	post-transcriptional gene silencing
3	GO:BP	GO:0031047	gene silencing by RNA
4	GO:BP	GO:0035195	miRNA-mediated gene silencing
5	GO:BP	GO:0009892	negative regulation of metabolic process
6	GO:BP	GO:0010629	negative regulation of gene expression
7	GO:BP	GO:0010605	negative regulation of macromolecule metabolic p
8	GO:BP	GO:0048519	negative regulation of biological process
9	KEGG	KEGG:05206	MicroRNAs in cancer
10	KEGG	KEGG:00000	KEGG root term

Figure 108 Over-representation analysis showing, which pathways were overrepresented in the sRNA data. A: all sRNAs, B: sRNAs upregulated at S0 vs baseline, C: downregulated at S0 vs baseline. Figures created on gProfiler. WP = wiki pathways, TF = transcription factors.

The next step was to look at whether sRNA expression is related to the expression or foldchange of their full-length mRNA counterparts. sRNAs arose from more abundant mRNAs (using log counts per million p=0.00000056, using reads per kilobase per million p=0.01). Though this relationship was significant, the difference in abundance between mRNAs with and without sRNAs in the sRNA seq data was slight, and there were plenty of examples of sRNAs deriving from low-abundance mRNAs and vice versa (Figure 109 A). Furthermore, there was no correlation between the average abundance of an mRNA and their sRNA counterpart(Figure 109 B&C). Potential correlations can be destroyed due to sequencing bias arising from different library preparation methods; however, no relationship between sRNA and mRNA foldchanges existed either, and this result is less effected by sequencing bias. Restricting correlations to mRNAs and sRNAs that were significantly DE did not uncover a correlation either.

Overall, there is a disconnect between sRNA expression and the full-length RNAs from, which they derive. This might be expected for sRNAs derived from cleavage of full-length RNAs by endonucleases, e.g., DICER, in the case of snoRNAs. sRNAs derived from lncRNA and protein-coding genes could be degradation products, but why are some differentially expressed? Moreover, why is their expression mainly unrelated to their full-length RNA counterparts? The possibility that these fragments represent biologically function entities rather than degradation products remains unresolved. More work is needed to characterise these protein-coding and lncRNA fragments.

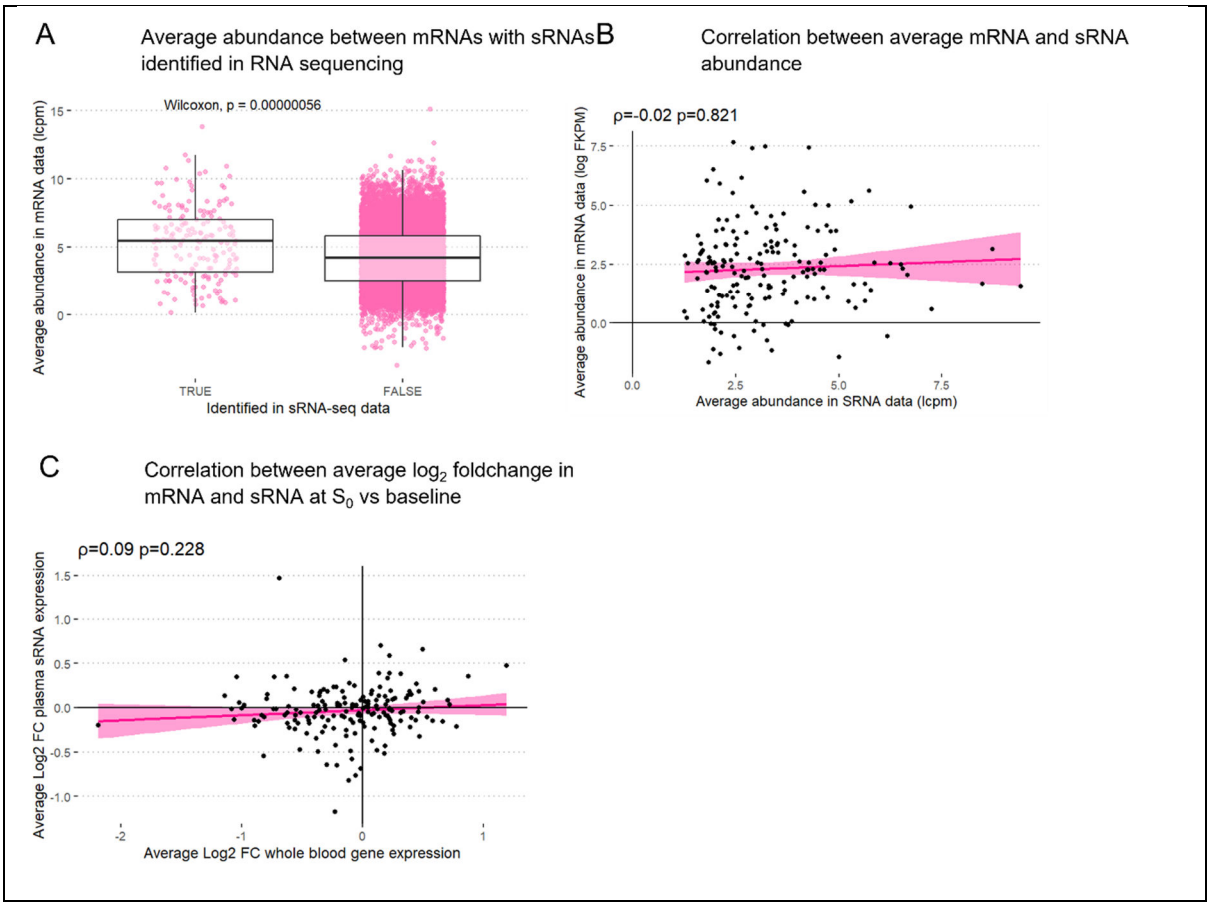


Figure 109 There is a limited relationship between sRNA and expression and the full-length mRNAs they derive from.

Supplementary Table 15 Indication of relative abundance of each RNA class in each substrate subject to sRNA sequencing in this thesis. +++ very abundant (~ 50% or more) ; ++ moderately abundant (~20-50%) : + lowly abundant (~10%) : . Barely detectable above noise (<2%)

RNA class	plasma	PBMCs	whole blood
miR	+	+++	++
tRNA	+++	++	.
snoRNA	.	++	+
offset miRNA	.	.	+++
yRNA	+	+	+
lncRNA	./+	./+	./+

protein coding	./+	./+	.
ReverseComplement	./+	.	.
rRNA	++	+	+
snRNA	.	.	.
vault	.	.	.

10 References

1. Gray, A. & Sharara, F. Global and regional sepsis and infectious syndrome mortality in 2019: a systematic analysis. *Lancet Glob Health* **10**, S2 (2022).
2. Rudd, K. E. *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet* **395**, 200–211 (2020).
3. Perin, J. *et al.* Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet Child Adolesc Health* **6**, 106–115 (2022).
4. Ferreras-Antolín, L., Oligbu, G., Okike, I. O. & Ladhani, S. Infection is associated with one in five childhood deaths in England and Wales: analysis of national death registrations data, 2013-15. *Arch Dis Child* **105**, 857–863 (2020).
5. Duclos, P., Okwo-Bele, J.-M., Gacic-Dobo, M. & Cherian, T. Global immunization: status, progress, challenges and future. *BMC Int Health Hum Rights* **9 Suppl 1**, S2 (2009).
6. Diray-Arce, J. *et al.* The Immune Signatures data resource, a compendium of systems vaccinology datasets. *Sci Data* **9**, 635 (2022).
7. Riese, P. *et al.* Distinct immunological and molecular signatures underpinning influenza vaccine responsiveness in the elderly. *Nat Commun* **13**, 6894 (2022).

8. Hagan, T., Nakaya, H. I., Subramaniam, S. & Pulendran, B. Systems vaccinology: Enabling rational vaccine design with systems biological approaches. *Vaccine* **33**, 5294–5301 (2015).
9. Pulendran, B., Li, S. & Nakaya, H. I. Systems vaccinology. *Immunity* **33**, 516–529 (2010).
10. Burel, J. G. *et al.* Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. *Front Immunol* **10**, 221 (2019).
11. Britto, C. *et al.* Host Respiratory Transcriptome Signature Associated with Poor Outcome in Children with Influenza–Staphylococcus aureus Pneumonia. *J Infect Dis* **226**, 1286–1294 (2022).
12. Burel, J. G. *et al.* Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. *Front Immunol* **10**, 221 (2019).
13. Lee, H. J. *et al.* Transcriptomic Studies of Malaria: a Paradigm for Investigation of Systemic Host-Pathogen Interactions. *Microbiol Mol Biol Rev* **82**, (2018).
14. Wu, D. *et al.* ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**, 2176–82 (2010).
15. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051–60 (2004).
16. Cullen, B. R. Transcription and Processing of Human microRNA Precursors. *Mol Cell* **16**, 861–865 (2004).

17. Schamberger, A., Sarkadi, B. & Orban, T. I. Human mirtrons can express functional microRNAs simultaneously from both arms in a flanking exon-independent manner. *RNA Biol* **9**, 1177–85 (2012).
18. Lee, Y., Jeon, K., Lee, J.-T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663–70 (2002).
19. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**, 509–524 (2014).
20. MiRbase version 22. http://www.mirbase.org/.
21. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140–D144 (2006).
22. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68–D73 (2014).
23. MiRbase version 22. Preprint at http://www.mirbase.org/.
24. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92–105 (2009).
25. Mehta, A. & Baltimore, D. MicroRNAs as regulatory elements in immune system logic. *Nat Rev Immunol* **16**, 279–294 (2016).
26. Chen, C.-Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs Modulate Hematopoietic Lineage Differentiation. *Science (1979)* **303**, (2004).

27. Cobb, B. S. *et al.* T cell lineage choice and differentiation in the absence of the RNase III enzyme Dicer. *J Exp Med* **201**, 1367–1373 (2005).
28. Koralov, S. B. *et al.* Dicer Ablation Affects Antibody Diversity and Cell Survival in the B Lymphocyte Lineage. *Cell* **132**, 860–874 (2008).
29. Bazzoni, F. *et al.* Induction and regulatory function of miR-9 in human monocytes and neutrophils exposed to proinflammatory signals. *Proceedings of the National Academy of Sciences* **106**, 5282–5287 (2009).
30. Curtale, G. *et al.* Negative regulation of Toll-like receptor 4 signaling by IL-10-dependent microRNA-146b. *Proceedings of the National Academy of Sciences* **110**, 11499–11504 (2013).
31. Ni, F. *et al.* MicroRNA transcriptomes of distinct human NK cell populations identify miR-362-5p as an essential regulator of NK cell function. *Sci Rep* **5**, 9993 (2015).
32. Smyth, L. A., Boardman, D. A., Tung, S. L., Lechler, R. & Lombardi, G. MicroRNAs affect dendritic cell function and phenotype. *Immunology* **144**, 197–205 (2015).
33. Steiner, D. F. *et al.* MicroRNA-29 Regulates T-Box Transcription Factors and Interferon-γ Production in Helper T Cells. *Immunity* **35**, 169–181 (2011).
34. Rossato, M. *et al.* IL-10-induced microRNA-187 negatively regulates TNF-α, IL-6, and IL-12p40 production in TLR4-stimulated monocytes. *Proc Natl Acad Sci U S A* **109**, E3101-10 (2012).
35. Drury, R. E., O'Connor, D. & Pollard, A. J. The clinical application of MicroRNAs in infectious disease. *Front Immunol* **8**, (2017).

36. WHO. Immunization. https://www.who.int/news-room/facts-in-pictures/detail/immunization.

37. Nandi, A. & Shet, A. Why vaccines matter: understanding the broader health, economic, and child development benefits of routine vaccination. *https://doi.org/10.1080/21645515.2019.1708669* **16**, 1900–1904 (2020).

38. Boylston, A. The origins of inoculation. *J R Soc Med* **105**, 309 (2012).

39. WHO. Immunization, Vaccines and Biologicals. https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases.

40. Jin, C. *et al.* Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of Salmonella Typhi: a randomised controlled, phase 2b trial. *Lancet* (2017) doi:10.1016/S0140-6736(17)32149-9.

41. Voysey, M. *et al.* Sex-dependent immune responses to infant vaccination: an individual participant data meta-analysis of antibody and memory B cells. *Vaccine* **34**, 1657–1664 (2016).

42. Klein, S. L., Jedlicka, A. & Pekosz, A. The Xs and Y of immune responses to viral vaccines. *Lancet Infect Dis* **10**, 338–349 (2010).

43. Nachtigall, I. *et al.* Effect of gender, age and vaccine on reactogenicity and incapacity to work after COVID-19 vaccination: a survey among health care workers. *BMC Infect Dis* **22**, 1–13 (2022).

44. Allen, J. C., Toapanta, F. R., Chen, W. & Tennant, S. M. Understanding immunosenescence and its impact on vaccination of older adults. *Vaccine* **38**, 8264–8272 (2020).

45. Menni, C. *et al.* Vaccine side-effects and SARS-CoV-2 infection after vaccination in users of the COVID Symptom Study app in the UK: a prospective observational study. *Lancet Infect Dis* **21**, 939–949 (2021).
46. Trzonkowski, P. *et al.* Association between cytomegalovirus infection, enhanced proinflammatory response and low level of anti-hemagglutinins during the anti-influenza vaccination - An impact of immunosenescence. *Vaccine* **21**, 3826–3836 (2003).
47. Sharpe, H. R. *et al.* CMV-associated T cell and NK cell terminal differentiation does not affect immunogenicity of ChAdOx1 vaccination. *JCI Insight* **7**, (2022).
48. Bowyer, G. *et al.* Reduced ebola vaccine responses in cmv+ young adults is associated with expansion of cd57+klrg1+ t cells. *Journal of Experimental Medicine* **217**, (2020).
49. O'Connor, D. *et al.* The effect of chronic cytomegalovirus infection on pneumococcal vaccine responses. *J Infect Dis* **209**, 1635–1641 (2014).
50. Nicoli, F. *et al.* Primary immune responses are negatively impacted by persistent herpesvirus infections in older people: results from an observational study on healthy subjects and a vaccination trial on subjects aged more than 70 years old. *EBioMedicine* **76**, (2022).
51. Lynn, D. J., Benson, S. C., Lynn, M. A. & Pulendran, B. Modulation of immune responses to vaccination by the microbiota: implications and potential mechanisms. *Nature Reviews Immunology 2021 22:1* **22**, 33–46 (2021).

52. Mentzer, A. J. *et al.* Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection. *Nature Medicine 2022* 1–1 (2022) doi:10.1038/s41591-022-02078-6.

53. Scepanovic, P. *et al.* Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med* **10**, 1–13 (2018).

54. Wen, S. *et al.* Identification of Two Novel Candidate Genetic Variants Associated With the Responsiveness to Influenza Vaccination. *Front Immunol* **12**, 2472 (2021).

55. Ovsyannikova, I. G. *et al.* Genome-Wide Association Study of Antibody Response to Smallpox Vaccine. *Vaccine* **30**, 4182 (2012).

56. Hervé, C., Laupèze, B., del Giudice, G., Didierlaurent, A. M. & Tavares Da Silva, F. The how's and what's of vaccine reactogenicity. *npj Vaccines 2019 4:1* **4**, 1–11 (2019).

57. Garçon, N., Leroux-Roels, G. & Cheng, W.-F. CHAPTER4 4 Vaccine adjuvants. *Perspectives in Vaccinology* **1**, 89–113 (2011).

58. Bemark, M. & Angeletti, D. Know your enemy or find your friend?—Induction of IgA at mucosal surfaces. *Immunol Rev* **303**, 83–102 (2021).

59. Halliley, J. L. *et al.* Long-Lived Plasma Cells Are Contained within the CD19(-)CD38(hi)CD138(+) Subset in Human Bone Marrow. *Immunity* **43**, 132–145 (2015).

60. Turner, J. S. *et al.* SARS-CoV-2 infection induces long-lived bone marrow plasma cells in humans. *Nature 2021 595:7867* **595**, 421–425 (2021).

61. Sun, L., Middleton, D. R., Wantuch, P. L., Ozdilek, A. & Avci, F. Y. Carbohydrates as T-cell antigens with implications in health and disease. *Glycobiology* **26**, 1029 (2016).

62. Allman, D., Wilmore, J. R. & Gaudette, B. T. The Continuing Story of T-cell Independent Antibodies. *Immunol Rev* **288**, 128 (2019).

63. Klein Klouwenberg, P. & Bont, L. Neonatal and infantile immune responses to encapsulated bacteria and conjugate vaccines. *Clin Dev Immunol* **2008**, (2008).

64. Johnson, P. R., Feldman, S., Thompson, J. M., Mahoney, J. D. & Wright, P. F. Comparison of long-term systemic and secretory antibody responses in children given live, attenuated, or inactivated influenza A vaccine. *J Med Virol* **17**, 325–335 (1985).

65. Gilbert, S. C. T-cell-inducing vaccines – what's the future. *Immunology* **135**, 19 (2012).

66. Hurwitz, J. L. Respiratory syncytial virus vaccine development. *Expert Rev Vaccines* **10**, 1415 (2011).

67. J, C., RL, M., LA, S., JH, S. & EH, L. Field evaluation of a respiratory syncytial virus vaccine and a trivalent parainfluenza virus vaccine in a pediatric population. *Am J Epidemiol* **89**, 449–463 (1969).

68. Cook, A. D., Christensen, A. D., Tewari, D., McMahon, S. B. & Hamilton, J. A. Immune Cytokines and Their Receptors in Inflammatory Pain. *Trends Immunol* **39**, 240–255 (2018).

69. Dinarello, C. A. Review: Infection, fever, and exogenous and endogenous pyrogens: some concepts have changed: *http://dx.doi.org/10.1177/0968051904010004030*1 **10**, 201–222 (2016).
70. Chang, R. B. Body thermal responses and the vagus nerve. *Neurosci Lett* **698**, 209 (2019).
71. Blomqvist, A. & Engblom, D. Neural Mechanisms of Inflammation-Induced Fever. *The Neuroscientist* **24**, 381 (2018).
72. Das, R. R., Panigrahi, I. & Naik, S. S. The Effect of Prophylactic Antipyretic Administration on Post-Vaccination Adverse Reactions and Antibody Response in Children: A Systematic Review. *PLoS One* **9**, (2014).
73. Burny, W. *et al.* Inflammatory parameters associated with systemic reactogenicity following vaccination with adjuvanted hepatitis B vaccines in humans. *Vaccine* **37**, 2004–2015 (2019).
74. Studholme, L. *et al.* Evaluation of the monocyte activation test for the safety testing of meningococcal B vaccine Bexsero: A collaborative study. *Vaccine* **37**, 3761–3769 (2019).
75. Dowling, D. J. *et al.* A meningococcal outer membrane vesicle vaccine incorporating genetically attenuated endotoxin dissociates inflammation from immunogenicity. *Front Immunol* **7**, 562 (2016).
76. O'Connor, D. *et al.* Gene expression profiling reveals insights into infant immunological and febrile responses to group B meningococcal vaccine. *Mol Syst Biol* **16**, 9888 (2020).
77. Zafack, J. G., Bureau, A., Skowronski, D. M. & de Serres, G. Adverse events following immunisation with four-component meningococcal

serogroup B vaccine (4CMenB): interaction with co-administration of routine infant vaccines and risk of recurrence in European randomised controlled trials. *BMJ Open* **9**, 26953 (2019).

78. Folegatti, P. M. *et al.* Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet* **396**, (2020).
79. Ramasamy, M. N. *et al.* Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *The Lancet* **396**, 1979–1993 (2020).
80. Shaw, R. H. *et al.* Heterologous prime-boost COVID-19 vaccination: initial reactogenicity data. *The Lancet* **397**, 2043–2046 (2021).
81. Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine* **384**, 403–416 (2021).
82. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine* **383**, 2603–2615 (2020).
83. Okada, C. *et al.* Reactogenicity of trivalent inactivated influenza vaccine in young children: Pronounced reactions by previous successive vaccinations. *Vaccine* **33**, 3586–3591 (2015).
84. Petousis-Harris, H. *et al.* Factors associated with reported pain on injection and reactogenicity to an OMV meningococcal B vaccine in children and adolescents. *Hum Vaccin Immunother* **11**, 1872–1877 (2015).
85. Cook, I. F. Sex differences in injection site reactions with human vaccines. *Hum Vaccin* **5**, 441–449 (2009).

86. Pittman, P. R. Aluminum-containing vaccine associated adverse events: role of route of administration and gender. *Vaccine* **20 Suppl 3**, (2002).

87. Weber, S. K. & Schlagenhauf, P. Childhood vaccination associated adverse events by sex: a literature review. *Travel Med Infect Dis* **12**, 459–480 (2014).

88. Joint committee on vaccination and immunisation. *JCVI position statement on use of Bexsero® meningococcal B vaccine in the UK*. (2014).

89. Rovin & Lisa. 22 CASE STUDIES WHERE PHASE 2 AND PHASE 3 TRIALS HAD DIVERGENT RESULTS. (2017).

90. Novavax Investor Relations - Press Releases & Statements. https://ir.novavax.com/2016-09-25-Novavax-Announces-Topline-RSV-F-Vaccine-Data-from-Two-Clinical-Trials-in-Older-Adults.

91. GE, G. *et al.* Vaccine Efficacy of ALVAC-HIV and Bivalent Subtype C gp120-MF59 in Adults. *N Engl J Med* **384**, 1089–1100 (2021).

92. Pronker, E. S., Weenen, T. C., Commandeur, H., Claassen, E. H. J. H. M. & Osterhaus, A. D. M. E. Risk in Vaccine Research and Development Quantified. *PLoS One* **8**, e57755 (2013).

93. Gouglas, D. *et al.* Estimating the cost of vaccine development against epidemic infectious diseases: a cost minimisation study. *Lancet Glob Health* **6**, e1386–e1396 (2018).

94. Win, Z. *et al.* Systematic Evaluation of Kinetics and Distribution of Muscle and Lymph Node Activation Measured by 18F-FDG- and 11C-PBR28-PET/CT Imaging, and Whole Blood and Muscle Transcriptomics After

Immunization of Healthy Humans With Adjuvanted and Unadjuvanted Vaccines. *Front Immunol* **0**, 3863 (2021).

95. I, G., EC, G. & MS, A. Human immunity to the meningococcus. I. The role of humoral antibodies. *J Exp Med* **129**, 1307–1326 (1969).
96. Measles outbreaks across England - GOV.UK. https://www.gov.uk/government/news/measles-outbreaks-across-england.
97. Amirthalingam GayatriAmirthalingam, G., Gupta, S. & Campbell, H. Pertussis immunisation and control in England and Wales, 1957 to 2012: a historical review. 1.
98. Ahmed, A. *et al.* Outbreak of vaccine-preventable diseases in Muslim majority countries. *J Infect Public Health* **11**, 153–155 (2018).
99. Nainani, V., Galal, U., Buttery, J. & Snape, M. D. An increase in accident and emergency presentations for adverse events following immunisation after introduction of the group B meningococcal vaccine: an observational study. *Arch Dis Child* **102**, 958–962 (2017).
100. Qin, L., Gilbert, P. B., Corey, L., McElrath, M. J. & Self, S. G. A framework for assessing immunological correlates of protection in vaccine trials. *J Infect Dis* **196**, 1304–1312 (2007).
101. Sadoff, J. C. & Wittes, J. Correlates, Surrogates, and Vaccines. *J Infect Dis* **196**, 1279–1281 (2007).
102. Findlow, J., Balmer, P. & Borrow, R. A review of complement sources used in serum bactericidal assays for evaluating immune responses to

meningococcal ACWY conjugate vaccines. *https://doi.org/10.1080/21645515.2019.1593082* **15**, 2491–2500 (2019).

103. Valente Pinto, M. *et al.* Understanding the reactogenicity of 4CMenB vaccine: Comparison of a novel and conventional method of assessing post-immunisation fever and correlation with pre-release in vitro pyrogen testing. *Vaccine* **38**, 7834–7841 (2020).
104. Sobolev, O. *et al.* Adjuvanted influenza-H1N1 vaccination reveals lymphoid signatures of age-dependent early responses and of clinical adverse events. *Nat Immunol* **17**, 204–213 (2016).
105. Takano, T. *et al.* Distinct immune cell dynamics correlate with the immunogenicity and reactogenicity of SARS-CoV-2 mRNA vaccine. *Cell Rep Med* **3**, (2022).
106. Blohmke, C. J. *et al.* Interferon-driven alterations of the host's amino acid metabolism in the pathogenesis of typhoid fever. *J Exp Med* **213**, 1061 (2016).
107. Nakaya, H. I. *et al.* Systems biology of vaccination for seasonal influenza in humans. *Nature Immunology 2011 12:8* **12**, 786–795 (2011).
108. Lee, A. H. *et al.* Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. *Nature Communications 2019 10:1* **10**, 1–14 (2019).
109. Ahern, D. J. *et al.* A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell* **185**, 916-938.e58 (2022).

110. M, K., JA, H., VJ, W., LJM, C. & M, L. Diagnosis of Bacterial Infection Using a 2-Transcript Host RNA Signature in Febrile Infants 60 Days or Younger. *JAMA* **317**, 1577–1578 (2017).
111. Herberg, J. A. *et al.* Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children. *JAMA* **316**, 835–845 (2016).
112. Pennisi, I. *et al.* Translation of a Host Blood RNA Signature Distinguishing Bacterial From Viral Infection Into a Platform Suitable for Development as a Point-of-Care Test. *JAMA Pediatr* **175**, 417–419 (2021).
113. Hafner, M. *et al.* RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**, 1697 (2011).
114. Tian, G. *et al.* Sequencing bias: comparison of different protocols of microRNA library construction. *BMC Biotechnol* **10**, 64 (2010).
115. Fuchs, R. T., Sun, Z., Zhuang, F. & Robb, G. B. Bias in Ligation-Based Small RNA Sequencing Library Construction Is Determined by Adaptor and RNA Structure. *PLoS One* **10**, e0126049 (2015).
116. Drury, R. E., Pollard, A. J. & O'Connor, D. The effect of H1N1 vaccination on serum miRNA expression in children: A tale of caution for microRNA microarray studies. *PLoS One* **14**, (2019).
117. Subramanian, I., Verma, S., Kumar, S., Jere, A. & Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights* **14**, (2020).
118. Gerstein, M. B. *et al.* What is a gene, post-ENCODE? History and updated definition. *Genome Res* **17**, 669–681 (2007).

119. Hodgkin, J. What does a worm want with 20,000 genes? *Genome Biol* **2**, comment2008.1 (2001).
120. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
121. Liu, G., Mattick, J. S. & Taft, R. J. A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle* **12**, 2061–2072 (2013).
122. Hrdlickova, B., de Almeida, R. C., Borek, Z. & Withoff, S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1842**, 1910–1922 (2014).
123. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–54 (1993).
124. Hu, M. *et al.* A KSHV microRNA Directly Targets G Protein-Coupled Receptor Kinase 2 to Promote the Migration and Invasion of Endothelial Cells by Inducing CXCR2 and Activating AKT Signaling. *PLoS Pathog* **11**, e1005171 (2015).
125. Liu, Y. *et al.* Ebola virus encodes a miR-155 analog to regulate importin-α5 expression. *Cell Mol Life Sci* **73**, 3733–44 (2016).
126. Sampey, G. C. *et al.* Exosomes from HIV-1-infected Cells Stimulate Production of Pro-inflammatory Cytokines through Trans-activating Response (TAR) RNA. *J Biol Chem* **291**, 1251–66 (2016).

127. Kim, S. *et al.* Human cytomegalovirus microRNA miR-US4-1 inhibits CD8+ T cell responses by targeting the aminopeptidase ERAP1. *Nat Immunol* **12**, 984–991 (2011).

128. Bellutti, F., Kauer, M., Kneidinger, D., Lion, T. & Klein, R. Identification of RISC-associated adenoviral microRNAs, a subset of their direct targets, and global changes in the targetome upon lytic adenovirus 5 infection. *J Virol* **89**, 1608–27 (2015).

129. Umbach, J. L. *et al.* MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature* **454**, 780 (2008).

130. Kawano, Y. *et al.* Plasma viral microRNA profiles reveal potential biomarkers for chronic active Epstein-Barr virus infection. *J Infect Dis* **208**, 771–9 (2013).

131. Tuddenham, L., Jung, J. S., Chane-Woon-Ming, B., Dölken, L. & Pfeffer, S. Small RNA deep sequencing identifies microRNAs and other small noncoding RNAs from human herpesvirus 6B. *J Virol* **86**, 1638–49 (2012).

132. Broekema, Nicole M1. Broekema, N. M. & Imperiale, M. J. miRNA regulation of BK polyomavirus replication during early infection. Proc. Natl. Acad. Sci. U. S. A. 110, 8200–5 (2013). & Imperiale, M. J. miRNA regulation of BK polyomavirus replication during early infection. *Proc Natl Acad Sci U S A* **110**, 8200–5 (2013).

133. Fromm, B., Zhong, X., Tarbier, M., Friedländer, M. R. & Hackenberg, M. The limits of human microRNA annotation have been met. *RNA* **28**, 781–785 (2022).

134. Lewis, B. P. *et al.* Prediction of mammalian microRNA targets. *Cell* **115**, 787–98 (2003).

135. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat Genet* **37**, 495–500 (2005).

136. Ebert, M. S. & Sharp, P. A. Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell* **149**, 515–524 (2012).

137. Lu, L. F. *et al.* A Single miRNA-mRNA Interaction Affects the Immune Response in a Context- and Cell-Type-Specific Manner. *Immunity* **43**, 52–64 (2015).

138. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).

139. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* **16**, 421–433 (2015).

140. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. *Cell* **125**, 1111–1124 (2006).

141. Schmiedel, J. M. *et al.* Gene expression. MicroRNA control of protein expression noise. *Science* **348**, 128–131 (2015).

142. Kim, D. H., Grün, D. & van Oudenaarden, A. Dampening of expression oscillations by synchronous regulation of a microRNA and its target. *Nature Genetics 2013 45:11* **45**, 1337–1344 (2013).

143. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43**, 854–859 (2011).

144. Gantier, M. P. *et al.* Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. *Nucleic Acids Res* **39**, 5692–5703 (2011).
145. Zhang, Z., Qin, Y. W., Brewer, G. & Jing, Q. MicroRNA degradation and turnover: regulating the regulators. *Wiley Interdiscip Rev RNA* **3**, 593–600 (2012).
146. Pawlica, P., Sheu-Gruttadauria, J., MacRae, I. J. & Steitz, J. A. How Complementary Targets Expose the microRNA 3′ End for Tailing and Trimming during Target-Directed microRNA Degradation. *Cold Spring Harb Symp Quant Biol* **84**, 179–183 (2019).
147. Ameres, S. L. *et al.* Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**, 1534–1539 (2010).
148. Shi, C. Y. *et al.* The ZSWIM8 ubiquitin ligase mediates target-directed microRNA degradation. *Science (1979)* **370**, (2020).
149. Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature 2014 505:7483* **505**, 344–352 (2014).
150. Liu, Y. *et al.* Competitive endogenous RNA is an intrinsic component of EMT regulatory circuits and modulates EMT. *Nature Communications 2019 10:1* **10**, 1–12 (2019).
151. de Candia, P. *et al.* Intracellular modulation, extracellular disposal and serum increase of MiR-150 mark lymphocyte activation. *PLoS One* **8**, e75348 (2013).
152. Irmak, M. K., Erdem, U. & Kubar, A. Antiviral activity of salivary microRNAs for ophthalmic herpes zoster. *Theor Biol Med Model* **9**, 21 (2012).

153. Weber, J. A. *et al.* The microRNA spectrum in 12 body fluids. *Clin Chem* **56**, 1733–41 (2010).
154. Vickers, K. C., Palmisano, B. T., Shoucri, B. M., Shamburek, R. D. & Remaley, A. T. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat Cell Biol* **13**, 423–433 (2011).
155. Arroyo, J. D. *et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proceedings of the National Academy of Sciences* **108**, 5003–5008 (2011).
156. Zernecke, A. *et al.* Delivery of microRNA-126 by apoptotic bodies induces CXCL12-dependent vascular protection. *Sci Signal* **2**, ra81 (2009).
157. Alsop, E. *et al.* A Novel Tissue Atlas and Online Tool for the Interrogation of Small RNA Expression in Human Tissues and Biofluids. *Front Cell Dev Biol* **10**, (2022).
158. Murillo, O. D. *et al.* exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell* **177**, 463-477.e15 (2019).
159. de Candia, P., Torri, A., Pagani, M. & Abrignani, S. Serum microRNAs as Biomarkers of Human Lymphocyte Activation in Health and Disease. *Front Immunol* **5**, 43 (2014).
160. Yahyaei, S. *et al.* Identification of a Specific miRNA Profile in HIV-Exposed Seronegative Individuals. *J Acquir Immune Defic Syndr* **73**, 11–9 (2016).

161. Ragusa, M. *et al.* Asymmetric RNA Distribution among Cells and Their Secreted Exosomes: Biomedical Meaning and Considerations on Diagnostic Applications. *Front Mol Biosci* **4**, 66 (2017).
162. Wang, W. *et al.* A Pilot Study of MicroRNAs Expression Profile in Serum and HBsAg Particles: Predictors of Therapeutic Vaccine Efficacy in Chronic Hepatitis B Patients. *Medicine* **95**, e2511 (2016).
163. Montecalvo, A. *et al.* Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes. *Blood* **119**, 756–766 (2012).
164. Zhang, Y. *et al.* Secreted monocytic miR-150 enhances targeted endothelial cell migration. *Mol Cell* **39**, 133–144 (2010).
165. Yang, M. *et al.* Microvesicles secreted by macrophages shuttle invasion-potentiating microRNAs into breast cancer cells. *Mol Cancer* **10**, 1–13 (2011).
166. Thomou, T. *et al.* Adipose-derived circulating miRNAs regulate gene expression in other tissues. *Nature 2017 542:7642* **542**, 450–455 (2017).
167. Liu, X. L. *et al.* Lipotoxic Hepatocyte-Derived Exosomal MicroRNA 192-5p Activates Macrophages Through Rictor/Akt/Forkhead Box Transcription Factor O1 Signaling in Nonalcoholic Fatty Liver Disease. *Hepatology* **72**, 454–469 (2020).
168. Xu, J. *et al.* Circulating Plasma Extracellular Vesicles from Septic Mice Induce Inflammation via MicroRNA- and TLR7-Dependent Mechanisms. *J Immunol* **201**, 3392–3400 (2018).
169. Mittelbrunn, M. *et al.* Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nat Commun* **2**, 282 (2011).

170. Alexander, M. *et al.* Exosome-delivered microRNAs modulate the inflammatory response to endotoxin. *Nature Communications 2015 6:1* **6**, 1–16 (2015).
171. Goodarzi, H. *et al.* Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. *Cell* **161**, 790–802 (2015).
172. Villarroya-Beltri, C. *et al.* Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. *Nat Commun* **4**, 2980 (2013).
173. Hobor, F. *et al.* A cryptic RNA-binding domain mediates Syncrip recognition and exosomal partitioning of miRNA targets. *Nat Commun* **9**, 831 (2018).
174. Lee, H. *et al.* Caveolin-1 selectively regulates microRNA sorting into microvesicles after noxious stimuli. *Journal of Experimental Medicine* **216**, 2202–2220 (2019).
175. Kosaka, N. *et al.* Neutral sphingomyelinase 2 (nSMase2)-dependent exosomal transfer of angiogenic microRNAs regulate cancer cell metastasis. *J Biol Chem* **288**, 10849–10859 (2013).
176. Xu, Y. F. *et al.* SRSF1 regulates exosome microRNA enrichment in human cancer cells. *Cell Communication and Signaling* **18**, 1–13 (2020).
177. Neilson, J. R., Zheng, G. X. Y., Burge, C. B. & Sharp, P. A. Dynamic regulation of miRNA expression in ordered stages of cellular development. *Genes Dev* **21**, 578–589 (2007).
178. Rupp, L. J. *et al.* The microRNA biogenesis machinery modulates lineage commitment during αβ T cell development. *J Immunol* **193**, 4032–42 (2014).

179. Baumjohann, D. & Ansel, K. M. MicroRNA-mediated regulation of T helper cell differentiation and plasticity. *Nat Rev Immunol* **13**, 666–78 (2013).

180. Li, J., Wan, Y., Ji, Q., Fang, Y. & Wu, Y. The role of microRNAs in B-cell development and function. *Cell Mol Immunol* **10**, 107–12 (2013).

181. Rossi, R. L. *et al.* Distinct microRNA signatures in human lymphocyte subsets and enforcement of the naive state in CD4+ T cells by the microRNA miR-125b. *Nat Immunol* **12**, 796–803 (2011).

182. Aalaei-andabili, S. H. *et al.* Toll like receptor (TLR)-induced differential expression of microRNAs (MiRs) and immune response against infection: A systematic review. *Journal of Infection* **67**, 251–264 (2013).

183. Montagner, S., Orlandi, E. M., Merante, S. & Monticelli, S. The role of miRNAs in mast cells and other innate immune cells. *Immunol Rev* **253**, 12–24 (2013).

184. Quinn, E. M., Wang, J. H., O'Callaghan, G. & Redmond, H. P. MicroRNA-146a is upregulated by and negatively regulates TLR2 signaling. *PLoS One* **8**, e62232 (2013).

185. Nahid, M. A., Satoh, M. & Chan, E. K. L. Mechanistic role of microRNA-146a in endotoxin-induced differential cross-regulation of TLR signaling. *J Immunol* **186**, 1723–34 (2011).

186. Essandoh, K., Li, Y., Huo, J. & Fan, G.-C. MiRNA-Mediated Macrophage Polarization and its Potential Role in the Regulation of Inflammatory Response. *Shock* **46**, 122–31 (2016).

187. Wang, P. *et al.* Identification of resting and type I IFN-activated human NK cell miRNomes reveals microRNA-378 and microRNA-30e as negative regulators of NK cell cytotoxicity. *J Immunol* **189**, 211–21 (2012).

188. Rossato, M. *et al.* IL-10-induced microRNA-187 negatively regulates TNF- , IL-6, and IL-12p40 production in TLR4-stimulated monocytes. *Proceedings of the National Academy of Sciences* **109**, E3101–E3110 (2012).

189. Wang, P. *et al.* Inducible microRNA-155 feedback promotes type I IFN signaling in antiviral innate immunity by targeting suppressor of cytokine signaling 1. *J Immunol* **185**, 6226–33 (2010).

190. Thounaojam, M. C. *et al.* MicroRNA 155 regulates Japanese encephalitis virus-induced inflammatory response by targeting Src homology 2-containing inositol phosphatase 1. *J Virol* **88**, 4798–810 (2014).

191. Gong, A.-Y. *et al.* MicroRNA-221 controls expression of intercellular adhesion molecule-1 in epithelial cells in response to Cryptosporidium parvum infection. *Int J Parasitol* **41**, 397–403 (2011).

192. Lam, W.-Y. *et al.* Effect of avian influenza A H5N1 infection on the expression of microRNA-141 in human respiratory epithelial cells. *BMC Microbiol* **13**, 104 (2013).

193. Zhou, R. *et al.* Histone deacetylases and NF-kB signaling coordinate expression of CX3CL1 in epithelial cells in response to microbial challenge by suppressing miR-424 and miR-503. *PLoS One* **8**, e65153 (2013).

194. Teng, Y. *et al.* The modulation of MiR-155 and MiR-23a manipulates Klebsiella pneumoniae Adhesion on Human pulmonary Epithelial cells via Integrin α5β1 Signaling. *Sci Rep* **6**, 31918 (2016).

195. Xu, H. *et al.* MicroRNA-122 supports robust innate immunity in hepatocytes by targeting the RTKs/STAT3 signaling pathway. *Elife* **8**, (2019).
196. Wang, P. *et al.* Inducible microRNA-155 Feedback Promotes Type I IFN Signaling in Antiviral Innate Immunity by Targeting Suppressor of Cytokine Signaling 1. *The Journal of Immunology* **185**, 6226–6233 (2010).
197. Lai, L. *et al.* MicroRNA-92a Negatively Regulates Toll-like Receptor (TLR)-triggered Inflammatory Response in Macrophages by Targeting MKK4 Kinase. *Journal of Biological Chemistry* **288**, 7956–7967 (2013).
198. Hou, J. *et al.* MicroRNA-146a Feedback Inhibits RIG-I-Dependent Type I IFN Production in Macrophages by Targeting TRAF6, IRAK1, and IRAK2. *The Journal of Immunology* **183**, 2150–2158 (2009).
199. Quinn, E. M., Wang, J. H., O'Callaghan, G. & Redmond, H. P. MicroRNA-146a is upregulated by and negatively regulates TLR2 signaling. *PLoS One* **8**, e62232 (2013).
200. McCoy, C. E. *et al.* IL-10 inhibits miR-155 induction by toll-like receptors. *J Biol Chem* **285**, 20492–8 (2010).
201. Xu, S., Guo, K., Zeng, Q., Huo, J. & Lam, K. P. The RNase III enzyme Dicer is essential for germinal center B-cell formation. *Blood* **119**, 767–776 (2012).
202. Yu, S. J. *et al.* Association of a microRNA-323b polymorphism with the persistence of hepatitis B virus infection by the enhancement of viral replication. *J Viral Hepat* **21**, 853–9 (2014).

203. Cezar-de-Mello, P. F. T. *et al.* Pre-miR-146a (rs2910164 G>C) single nucleotide polymorphism is genetically and functionally associated with leprosy. *PLoS Negl Trop Dis* **8**, e3099 (2014).

204. Cheong, J. Y., Shin, H. D., Kim, Y. J. & Cho, S. W. Association of polymorphism in MicroRNA 219-1 with clearance of hepatitis B virus infection. *J Med Virol* **85**, 808–14 (2013).

205. Misra, M. K. *et al.* Genetic variation in Micro-RNA genes of host genome affects clinical manifestation of symptomatic Human Cytomegalovirus infection. *Hum Immunol* **76**, 765–9 (2015).

206. Al-Qahtani, A. A. *et al.* Association of Single Nucleotide Polymorphisms in MicroRNAs with Susceptibility to Hepatitis B Virus Infection and HBV-related Liver Complications: A Study in a Saudi Arabian Population. *J Viral Hepat* (2017) doi:10.1111/jvh.12749.

207. van Loon, W., Gai, P. P., Hamann, L., Bedu-Addo, G. & Mockenhaupt, F. P. MiRNA-146a polymorphism increases the odds of malaria in pregnancy 11 Medical and Health Sciences 1107 Immunology 11 Medical and Health Sciences 1108 Medical Microbiology. *Malar J* **18**, 1–7 (2019).

208. Ellwanger, J. H., Zambra, F. M. B., Guimarães, R. L. & Chies, J. A. B. MicroRNA-related polymorphisms in infectious diseases-tiny changes with a huge impact on viral infections and potential clinical applications. *Front Immunol* **9**, 1316 (2018).

209. Kazemi, S. *et al.* Association between risk of brucellosis and genetic variations in MicroRNA-146a. *BMC Infect Dis* **21**, 1–7 (2021).

210. Nathans, R. *et al.* Cellular microRNA and P bodies modulate host-HIV-1 interactions. *Mol Cell* **34**, 696–709 (2009).
211. Lecellier, C.-H. *et al.* A cellular microRNA mediates antiviral defense in human cells. *Science* **308**, 557–60 (2005).
212. Bai, X. T. & Nicot, C. miR-28-3p is a cellular restriction factor that inhibits human T cell leukemia virus, type 1 (HTLV-1) replication and virus infection. *J Biol Chem* **290**, 5381–90 (2015).
213. He, J. *et al.* MiR-122 directly inhibits human papillomavirus E6 gene and enhances interferon signaling through blocking suppressor of cytokine signaling 1 in SiHa cells. *PLoS One* **9**, e108410 (2014).
214. Zheng, Z. *et al.* Human microRNA hsa-miR-296-5p suppresses enterovirus 71 replication by targeting the viral genome. *J Virol* **87**, 5645–56 (2013).
215. Otsuka, M. *et al.* Hypersusceptibility to Vesicular Stomatitis Virus Infection in Dicer1-Deficient Mice Is Due to Impaired miR24 and miR93 Expression. *Immunity* **27**, 123–134 (2007).
216. Siniscalchi, C., di Palo, A., Russo, A. & Potenza, N. Human MicroRNAs Interacting With SARS-CoV-2 RNA Sequences: Computational Analysis and Experimental Target Validation. *Front Genet* **12**, 760 (2021).
217. Li, C. *et al.* SARS-COV-2 as potential microRNA sponge in COVID-19 patients. *BMC Med Genomics* **15**, 1–10 (2022).
218. Jopling, C. L., Yi, M., Lancaster, A. M., Lemon, S. M. & Sarnow, P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309**, 1577–81 (2005).

219. Sarnow, P. & Sagan, S. M. Unraveling the Mysterious Interactions Between Hepatitis C Virus RNA and Liver-Specific MicroRNA-122. *Annu Rev Virol* (2016) doi:10.1146/annurev-virology-110615-042409.

220. Machlin, E. S., Sarnow, P. & Sagan, S. M. Masking the 5' terminal nucleotides of the hepatitis C virus genome by an unconventional microRNA-target RNA complex. *Proc Natl Acad Sci U S A* **108**, 3193–8 (2011).

221. Perez, J. T. *et al.* MicroRNA-mediated species-specific attenuation of influenza A virus. *Nat Biotechnol* **27**, 572–6 (2009).

222. Barnes, D., Kunitomi, M., Vignuzzi, M., Saksela, K. & Andino, R. Harnessing endogenous miRNAs to control virus tissue tropism as a strategy for developing attenuated virus vaccines. *Cell Host Microbe* **4**, 239–48 (2008).

223. Brostoff, T. *et al.* MicroRNA reduction of neuronal West Nile virus replication attenuates and affords a protective immune response in mice. *Vaccine* **34**, 5366–5375 (2016).

224. Heiss, B. L., Maximova, O. A. & Pletnev, A. G. Insertion of microRNA targets into the flavivirus genome alters its highly neurovirulent phenotype. *J Virol* **85**, 1464–72 (2011).

225. Shen, X., Sun, W., Shi, Y., Xing, Z. & Su, X. Altered viral replication and cell responses by inserting microRNA recognition element into PB1 in pandemic influenza A virus (H1N1) 2009.1. Shen, X., Sun, W., Shi, Y., Xing, Z. & Su, X. Altered viral replication and cell responses by inserting microRNA r. *Mediators Inflamm* **2015**, 976575 (2015).

226. Heiss, B. L., Maximova, O. A., Thach, D. C., Speicher, J. M. & Pletnev, A. G. MicroRNA targeting of neurotropic flavivirus: effective control of virus escape and reversion to neurovirulent phenotype. *J Virol* **86**, 5647–59 (2012).
227. Grinberg, M. *et al.* Vaccinia virus infection suppresses the cell microRNA machinery. *Arch Virol* **157**, 1719–27 (2012).
228. Moon, S. L. *et al.* Flavivirus sfRNA suppresses antiviral RNA interference in cultured cells and mosquitoes and directly interacts with the RNAi machinery. *Virology* **485**, 322–9 (2015).
229. Sanghvi, V. R. & Steel, L. F. A re-examination of global suppression of RNA interference by HIV-1. *PLoS One* **6**, e17246 (2011).
230. Chinnappan, M. *et al.* Key elements of the RNAi pathway are regulated by hepatitis B virus replication and HBx acts as a viral suppressor of RNA silencing. *Biochem J* **462**, 347–58 (2014).
231. Mansouri, S., Pan, Q., Blencowe, B. J., Claycomb, J. M. & Frappier, L. Epstein-Barr virus EBNA1 protein regulates viral latency through effects on let-7 microRNA and dicer. *J Virol* **88**, 11166–77 (2014).
232. Tang, W.-F. *et al.* Host MicroRNA miR-197 Plays a Negative Regulatory Role in the Enterovirus 71 Infectious Cycle by Targeting the RAN Protein. *J Virol* **90**, 1424–38 (2016).
233. Mukherjee, A., Shrivastava, S., Bhanja Chowdhury, J., Ray, R. B. R. & Ray, R. B. R. Transcriptional suppression of miR-181c by hepatitis C virus enhances homeobox A1 expression. *J Virol* **88**, 7929–40 (2014).

234. Li, J. *et al.* miR-34a and its novel target, NLRC5, are associated with HPV16 persistence. *Infect Genet Evol* **44**, 293–9 (2016).
235. Bandyopadhyay, S. *et al.* Hepatitis C virus infection and hepatic stellate cell activation downregulate miR-29: miR-29 overexpression reduces hepatitis C viral abundance in culture. *J Infect Dis* **203**, 1753–62 (2011).
236. Wang, Y. *et al.* Hepatitis B viral RNA directly mediates down-regulation of the tumor suppressor microRNA miR-15a/miR-16-1 in hepatocytes. *J Biol Chem* **288**, 18484–93 (2013).
237. Maudet, C. *et al.* Functional high-throughput screening identifies the miR-15 microRNA family as cellular restriction factors for Salmonella infection. *Nat Commun* **5**, 4718 (2014).
238. Fu, Y.-R. *et al.* MicroRNA miR-21 attenuates human cytomegalovirus replication in neural cells by targeting Cdc25a. *J Virol* **89**, 1070–82 (2015).
239. Cameron, J. E. *et al.* Epstein-Barr virus latent membrane protein 1 induces cellular MicroRNA miR-146a, a modulator of lymphocyte signaling pathways. *J Virol* **82**, 1946–58 (2008).
240. Mantel, P.-Y. *et al.* Infected erythrocyte-derived extracellular vesicles alter vascular function via regulatory Ago2-miRNA complexes in malaria. *Nat Commun* **7**, 12727 (2016).
241. L Ramos, T. *et al.* MSC surface markers (CD44, CD73, and CD90) can identify human MSC-derived extracellular vesicles by conventional flow cytometry. *Cell Commun Signal* **14**, 2 (2016).

242. van der Vos, K. E. *et al.* Directly visualized glioblastoma-derived extracellular vesicles transfer RNA to microglia/macrophages in the brain. *Neuro Oncol* **18**, 58–69 (2016).
243. Shen, L. M., Quan, L. & Liu, J. Tracking Exosomes in Vitro and in Vivo to Elucidate Their Physiological Functions: Implications for Diagnostic and Therapeutic Nanocarriers. *ACS Appl Nano Mater* **1**, 2438–2448 (2018).
244. Wang, H. *et al.* Four serum microRNAs identified as diagnostic biomarkers of sepsis. *J Trauma Acute Care Surg* **73**, 850–4 (2012).
245. Wang, H. *et al.* Serum miR-574-5p: a prognostic predictor of sepsis patients. *Shock* **37**, 263–7 (2012).
246. Goodwin, A. J. *et al.* Plasma levels of microRNA are altered with the development of shock in human sepsis: an observational study. *Crit Care* **19**, 440 (2015).
247. Wang, H. *et al.* Serum microRNA signatures identified by Solexa sequencing predict sepsis patients' mortality: a prospective observational study. *PLoS One* **7**, e38885 (2012).
248. Yao, L. *et al.* Clinical evaluation of circulating microRNA-25 level change in sepsis and its potential relationship with oxidative stress. *Int J Clin Exp Pathol* **8**, 7675–84 (2015).
249. The Impact of MicroRNA Gene Polymorphisms on Immune Response to Diphtheria Vaccine in Iraqi Children and Adolescents | JOURNAL OF UNIVERSITY OF BABYLON for Pure and Applied Sciences. https://www.journalofbabylon.com/index.php/JUBPAS/article/view/1356.

250. Riazalhosseini, B., Mohamed, Z., Apalasamy, Y. D., Eng, H. S. & Mohamed, R. Association between microRNA-196A2 and microRNA-146A polymorphisms and progression to cirrhosis and hepatocellular carcinoma in patients with viral hepatitis B. *Pharmacogenet Genomics* **26**, 74–9 (2016).
251. Corral-Fernández, N. E. *et al.* Induction of transcription factors, miRNAs and cytokines involved in T lymphocyte differentiation in BCG-vaccinated subjects. *Mol Immunol* **77**, 44–51 (2016).
252. Haralambieva, I. H. *et al.* Differential miRNA expression in B cells is associated with inter-individual differences in humoral immune response to measles vaccination. *PLoS One* **13**, e0191812 (2018).
253. Okamoto, M. *et al.* MicroRNA-451a in extracellular, blood-resident vesicles attenuates macrophage and dendritic cell responses to influenza whole-virus vaccine. *Journal of Biological Chemistry* **293**, 18585–18600 (2018).
254. Drury, R. E., Pollard, A. J. & O'Connor, D. The effect of H1N1 vaccination on serum miRNA expression in children: A tale of caution for microRNA microarray studies. *PLoS One* **14**, (2019).
255. Nakaya, H. I. *et al.* Systems Analysis of Immunity to Influenza Vaccination across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures. *Immunity* **43**, 1186–98 (2015).
256. Ho, B.-C. C. *et al.* Inhibition of miR-146a prevents enterovirus-induced death by restoring the production of type I interferon. *Nat Commun* **5**, 3344 (2014).

257. Zhang, T. *et al.* Salmonella enterica serovar enteritidis modulates intestinal epithelial miR-128 levels to decrease macrophage recruitment via macrophage colony-stimulating factor. *J Infect Dis* **209**, 2000–11 (2014).
258. van der Ree, M. H. *et al.* Long-term safety and efficacy of microRNA-targeted therapy in chronic hepatitis C patients. *Antiviral Res* **111**, 53–9 (2014).
259. van der Ree, M. H. *et al.* Safety, tolerability, and antiviral effect of RG-101 in patients with chronic hepatitis C: a phase 1B, double-blind, randomised controlled trial. *Lancet* (2017) doi:10.1016/S0140-6736(16)31715-9.
260. RG-101 IN COMBINATION WITH 4 WEEKS OF ORAL DIRECT ACTING ANTIVIRAL THERAPY ACHIEVES HIGH VIROLOGIC RESPONSE RATES IN TREATMENT NAÏVE GENOTYPE 1 AND 4 CHRONIC HEPATITIS C PATIENTS: INTERIM RESULTS FROM A RANDOMISED, MULTI-CENTER, PHASE 2 STUDY. https://www.natap.org/2016/EASL/EASL_14.htm.
261. Wheatley, A. K. *et al.* Co-expression of miRNA targeting the expression of PERK, but not PKR, enhances cellular immunity from an HIV-1 Env DNA vaccine. *PLoS One* **6**, e18225 (2011).
262. Bratkovič, T., Bozič, J. & Rogelj, B. Functional diversity of small nucleolar RNAs. *Nucleic Acids Res* **48**, 1627–1651 (2020).
263. Wajahat, M., Bracken, C. P. & Orang, A. Emerging Functions for snoRNAs and snoRNA-Derived Fragments. *International Journal of Molecular Sciences 2021, Vol. 22, Page 10193* **22**, 10193 (2021).
264. Wang, X., Ramat, A., Simonelig, M. & Liu, M. F. Emerging roles and functional mechanisms of PIWI-interacting RNAs. *Nature Reviews*

Molecular Cell Biology 2022 1–19 (2022) doi:10.1038/s41580-022-00528-0.

265. Saikia, M. *et al.* Angiogenin-Cleaved tRNA Halves Interact with Cytochrome c , Protecting Cells from Apoptosis during Osmotic Stress . *Mol Cell Biol* **34**, 2450–2463 (2014).
266. Ivanov, P., Emara, M. M., Villen, J., Gygi, S. P. & Anderson, P. Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol Cell* **43**, 613–623 (2011).
267. Emara, M. M. *et al.* Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J Biol Chem* **285**, 10959–10968 (2010).
268. Akiyama, Y., Tomioka, Y., Abe, T., Anderson, P. & Ivanov, P. In lysate RNA digestion provides insights into the angiogenin's specificity towards transfer RNAs. *RNA Biol* **18**, 2546–2555 (2021).
269. Zhou, J. *et al.* Identification of two novel functional tRNA-derived fragments induced in response to respiratory syncytial virus infection. *J Gen Virol* **98**, 1600–1610 (2017).
270. Deng, J. *et al.* Respiratory Syncytial Virus Utilizes a tRNA Fragment to Suppress Antiviral Responses Through a Novel Targeting Mechanism. *Mol Ther* **23**, 1622–9 (2015).
271. Wang, Q. *et al.* Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. *Mol Ther* **21**, 368–379 (2013).

272. Burroughs, A. M. *et al.* Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol* **8**, 158–177 (2011).
273. Thomson, D. W. *et al.* Assessing the gene regulatory properties of Argonaute-bound small RNAs of diverse genomic origin. *Nucleic Acids Res* **43**, 470 (2015).
274. Nechooshtan, G., Yunusov, D., Chang, K. & Gingeras, T. R. Processing by RNase 1 forms tRNA halves and distinct Y RNA fragments in the extracellular environment. *Nucleic Acids Res* **48**, 8035–8049 (2020).
275. Driedonks, T. A. P. & Nolte-T'Hoen, E. N. M. Circulating Y-RNAs in Extracellular Vesicles and Ribonucleoprotein Complexes; Implications for the Immune System. *Front Immunol* **9**, 3164 (2018).
276. Dhahbi, J. M. tRNA Halves: The Next Generation of Immune Signaling Molecules. *Front Immunol* **6**, 74 (2015).
277. Gámbaro, F. *et al.* Stable tRNA halves can be sorted into extracellular vesicles and delivered to recipient cells in a concentration-dependent manner. *RNA Biol* **17**, 1168–1182 (2020).
278. Guan, L., Karaiskos, S. & Grigoriev, A. Inferring targeting modes of Argonaute-loaded tRNA fragments. *RNA Biol* **17**, 1070–1080 (2020).
279. Terenova, J., Alagia, A., Ketley, R. F., Chelysheva, I. & Gullerova, M. Small non-coding vault RNA1-2 modulate expression of cell membrane proteins through nascent RNA silencing. *bioRxiv* 2022.08.12.503724 (2022) doi:10.1101/2022.08.12.503724.

280. Persson, H. *et al.* The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat Cell Biol* **11**, 1268–1271 (2009).
281. Hussain, S. *et al.* NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep* **4**, 255–261 (2013).
282. Chen, Z. *et al.* Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. *PLoS One* **12**, e0176458 (2017).
283. Bakre, A. A., Duffy, C., Abdullah, H., Cosby, S. L. & Tripp, R. A. Small Non-coding RNA Expression Following Respiratory Syncytial Virus or Measles Virus Infection of Neuronal Cells. *Front Microbiol* **12**, (2021).
284. Looney, M. M., Lu, Y., Karakousis, P. C. & Halushka, M. K. Mycobacterium tuberculosis Infection Drives Mitochondria-Biased Dysregulation of Host Transfer RNA–Derived Fragments. *J Infect Dis* **223**, 1796–1805 (2021).
285. Schwarzenbach, H., da Silva, A. M., Calin, G. & Pantel, K. Data Normalization Strategies for MicroRNA Quantification. *Clin Chem* **61**, 1333–1342 (2015).
286. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* **64**, 5245–5250 (2004).
287. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, 1–12 (2002).

288. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnol Lett* **26**, 509–515 (2004).

289. Das, S. *et al.* The Extracellular RNA Communication Consortium: Establishing Foundational Knowledge and Technologies for Extracellular RNA Research. *Cell* **177**, 231–242 (2019).

290. Rozowsky, J., Kitchen, R. R., Park, J. J., Subramanian, S. L. & Milosavljevic, A. exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling. (2019) doi:10.1016/j.cels.2019.03.004.

291. microRNA-seq pipeline – ENCODE. https://www.encodeproject.org/pipelines/ENCPL280YDY/.

292. Baran-Gale, J. *et al.* Addressing bias in small RNA library preparation for sequencing: A new protocol recovers microRNAs that evade capture by current methods. *Front Genet* **6**, 352 (2015).

293. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature Genetics 2007 39:10* **39**, 1278–1284 (2007).

294. Zhao, B. & Xue, B. Significant improvement of miRNA target prediction accuracy in large datasets using meta-strategy based on comprehensive voting and artificial neural networks. *BMC Genomics 2019 20:1* **20**, 1–17 (2019).

295. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, (2015).

296. McGeary, S. E. *et al.* The biochemical basis of microRNA targeting efficacy. *Science* **366**, (2019).
297. Paraskevopoulou, M. D. *et al.* DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res* **41**, W169 (2013).
298. Liu, W. & Wang, X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol* **20**, 1–10 (2019).
299. Wong, N. & Wang, X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res* **43**, D146 (2015).
300. Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* **11**, 1–14 (2010).
301. John, B. *et al.* Human MicroRNA Targets. *PLoS Biol* **2**, e363 (2004).
302. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature Genetics 2007 39:10* **39**, 1278–1284 (2007).
303. L, S., L, P., Y, T., L, K. & PP, P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353–358 (2011).
304. Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature 2014 505:7483* **505**, 344–352 (2014).
305. Huang, H. Y. *et al.* miRTarBase update 2022: an informative resource for experimentally validated miRNA–target interactions. *Nucleic Acids Res* **50**, D222–D230 (2022).

306. Xiao, F. *et al.* miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* **37**, (2009).
307. Karagkouni, D. *et al.* DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res* **46**, D239–D245 (2018).
308. Wang, X. W. *et al.* A microRNA-inducible CRISPR–Cas9 platform serves as a microRNA sensor and cell-type-specific genome regulation tool. *Nature Cell Biology 2019 21:4* **21**, 522–530 (2019).
309. Chang, H. *et al.* CRISPR/cas9, a novel genomic tool to knock down microRNA in vitro and in vivo. *Scientific Reports 2016 6:1* **6**, 1–12 (2016).
310. Park, C. Y. *et al.* A Resource for the Conditional Ablation of microRNAs in the Mouse. *Cell Rep* **1**, 385–391 (2012).
311. Lizio, M. *et al.* Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res* **47**, D752–D758 (2019).
312. Landgraf, P. *et al.* A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell* **129**, 1401 (2007).
313. Juzenas, S. *et al.* A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Res* **45**, 9290 (2017).
314. Keller, A. *et al.* miRNATissueAtlas2: an update to the human miRNA tissue atlas. *Nucleic Acids Res* **50**, D211–D221 (2022).
315. Leidinger, P., Backes, C., Meder, B., Meese, E. & Keller, A. The human miRNA repertoire of different blood compounds. *BMC Genomics* **15**, (2014).

316. R Core Team (2022). R: A language and environment for statistical . Preprint at https://www.R-project.org/.
317. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
318. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288 (2012).
319. Smyth, G. K., Michaud, J. & Scott, H. S. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067–2075 (2005).
320. Liu, R. *et al.* Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res* **43**, (2015).
321. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
322. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION. *Ann Appl Stat* **10**, 946–963 (2016).
323. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).

324. Deschamps-Francoeur, G., Simoneau, J. & Scott, M. S. Handling multi-mapped reads in RNA-seq. *Comput Struct Biotechnol J* **18**, 1569–1576 (2020).

325. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology 2010 28:5* **28**, 511–515 (2010).

326. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

327. Pantano, L., Estivill, X. & Martí, E. A non-biased framework for the annotation and classification of the non-miRNA small RNA transcriptome. *Bioinformatics* **27**, 3202–3203 (2011).

328. Zytnicki, M. mmquant: How to count multi-mapping reads? *BMC Bioinformatics* **18**, 1–6 (2017).

329. Pertea, G. gpertea/fqtrim. Preprint at https://doi.org/10.5281/ZENODO.1185412 (2018).

330. Yoshihama, M., Nakao, A. & Kenmochi, N. SnOPY: A small nucleolar RNA orthological gene database. *BMC Res Notes* **6**, 1–5 (2013).

331. Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44**, D184–D189 (2016).

332. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–D745 (2016).

333. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, 1–10 (2009).
334. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357 (2012).
335. Bowtie2 prebuilt indexes. https://bowtie-bio.sourceforge.net/bowtie2/manual.shtml.
336. Bowtie prebuilt indexes. https://bowtie-bio.sourceforge.net/manual.shtml.
337. Kern, F. *et al.* miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems. *Nucleic Acids Res* **48**, W521–W528 (2020).
338. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2021) doi:10.1101/060012.
339. Chang, L., Zhou, G., Soufan, O. & Xia, J. miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Res* **48**, W244–W251 (2020).
340. Godard, P. & van Eyll, J. Pathway analysis from lists of microRNAs: common pitfalls and alternative strategy. *Nucleic Acids Res* **43**, 3490 (2015).
341. Parikh, S. R. *et al.* Effectiveness and impact of a reduced infant schedule of 4CMenB vaccine against group B meningococcal disease in England: a national observational cohort study. *The Lancet* **388**, 2775–2782 (2016).

342. Murdoch, H., Wallace, L., Bishop, J., Robertson, C. & Claire Cameron, J. Risk of hospitalisation with fever following MenB vaccination: self-controlled case series analysis. *Arch Dis Child* **102**, 894–898 (2017).
343. Shah, J. S., Soon, P. S. & Marsh, D. J. Comparison of Methodologies to Detect Low Levels of Hemolysis in Serum for Accurate Assessment of Serum microRNAs. *PLoS One* **11**, e0153200 (2016).
344. Kirschner, M. B. *et al.* The Impact of Hemolysis on Cell-Free microRNA Biomarkers. *Front Genet* **4**, 94 (2013).
345. Kirschner, M. B. *et al.* Haemolysis during Sample Preparation Alters microRNA Content of Plasma. *PLoS One* **6**, e24145 (2011).
346. Max, K. E. A. *et al.* Human plasma and serum extracellular small RNA reference profiles and their clinical utility. *Proc Natl Acad Sci U S A* **115**, E5334–E5343 (2018).
347. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453–457 (2015).
348. Moldovan, L. *et al.* Methodological challenges in utilizing miRNAs as circulating biomarkers. *J Cell Mol Med* **18**, 371–90 (2014).
349. Murillo, O. D. *et al.* exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell* **177**, 463-477.e15 (2019).
350. Keller, A. *et al.* miRNATissueAtlas2: an update to the human miRNA tissue atlas. *Nucleic Acids Res* **50**, D211–D221 (2022).
351. Jahangiri, A. HDL and the acute phase response. *Curr Opin Endocrinol Diabetes Obes* **17**, 156 (2010).

352. Cabral, B. C. A. *et al.* Circulating microRNAs associated with liver fibrosis in chronic hepatitis C patients. *Biochem Biophys Rep* **24**, 100814 (2020).
353. Vliegenthart, A. D. B. *et al.* Comprehensive microRNA profiling in acetaminophen toxicity identifies novel circulating biomarkers for human liver and kidney injury. *Scientific Reports 2015 5:1* **5**, 1–13 (2015).
354. Ward, J. *et al.* Circulating microRNA profiles in human patients with acetaminophen hepatotoxicity or ischemic hepatitis. *Proc Natl Acad Sci U S A* **111**, 12169–12174 (2014).
355. Zhang, Z. *et al.* Serum miR-483-5p as a potential biomarker to detect hepatocellular carcinoma. *Hepatol Int* **7**, 199–207 (2013).
356. Chang, J. *et al.* RNA Biology miR-122, a Mammalian Liver-Specific microRNA, is Processed from hcr mRNA and MayDownregulate the High Affinity Cationic Amino Acid Transporter CAT-1. *RNA Biol* **1**, 106–113.
357. Jopling, C. RNA Biology Liver-specific microRNA-122: Biogenesis and function. (2012) doi:10.4161/rna.18827.
358. Rivkin, M. *et al.* Inflammation-Induced Expression and Secretion of MicroRNA 122 Leads to Reduced Blood Levels of Kidney-Derived Erythropoietin and Anemia. *Gastroenterology* **151**, 999-1010.e3 (2016).
359. Ludwig, N. *et al.* Distribution of miRNA expression across human tissues. *Nucleic Acids Res* **44**, 3865–3877 (2016).
360. Landgraf, P. *et al.* A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell* **129**, 1401 (2007).

361. Li, X. D. *et al.* Elevated plasma miRNA-122, -140-3p, -720, -2861, and -3149 during early period of acute coronary syndrome are derived from peripheral blood mononuclear cells. *PLoS One* **12**, (2017).
362. Zou, L., Feng, Y., Xu, G., Jian, W. & Chao, W. Splenic RNA and MicroRNA Mimics Promote Complement Factor B Production and Alternative Pathway Activation via Innate Immune Signaling. *The Journal of Immunology* **196**, 2788–2798 (2016).
363. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* **16**, 1–14 (2015).
364. Lizio, M. *et al.* Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res* **47**, D752–D758 (2019).
365. Béres, N. J. *et al.* Role of Altered Expression of miR-146a, miR-155, and miR-122 in Pediatric Patients with Inflammatory Bowel Disease. *Inflamm Bowel Dis* **22**, 327–335 (2016).
366. Jiang, M. *et al.* IL-22-induced miR-122-5p promotes keratinocyte proliferation by targeting Sprouty2. *Exp Dermatol* **26**, 368–374 (2017).
367. Juzenas, S. *et al.* A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Res* **45**, 9290 (2017).
368. Hunter, M. P. *et al.* Detection of microRNA expression in human peripheral blood microvesicles. *PLoS One* **3**, (2008).
369. Pritchard, C. C. *et al.* Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res (Phila)* **5**, 492–497 (2012).

370. Duttagupta, R., Jiang, R., Gollub, J., Getts, R. C. & Jones, K. W. Impact of Cellular miRNAs on Circulating miRNA Biomarker Signatures. *PLoS One* **6**, e20769 (2011).
371. Mesitov, M. V *et al.* Differential processing of small RNAs during endoplasmic reticulum stress. *Sci Rep* **7**, 46080 (2017).
372. Li, J., Zeng, X. & Wang, W. miR-122-5p downregulation attenuates lipopolysaccharide-induced acute lung injury by targeting IL1RN. *Exp Ther Med* **22**, 1–10 (2021).
373. Bros, M. *et al.* Differentially Tolerized Mouse Antigen Presenting Cells Share a Common miRNA Signature Including Enhanced mmu-miR-223-3p Expression, which Is Sufficient to Imprint a Protolerogenic State. *Front Pharmacol* **9**, (2018).
374. Shen, Y. *et al.* MicroRNA-885-3p alleviates bronchial epithelial cell injury induced by lipopolysaccharide via toll-like receptor 4. *Bioengineered* **13**, 5305–5317 (2022).
375. Wang, Y. *et al.* Injured liver-released miRNA-122 elicits acute pulmonary inflammation via activating alveolar macrophage TLR7 signaling pathway. *Proc Natl Acad Sci U S A* **116**, 6162–6171 (2019).
376. Xu, H. *et al.* MicroRNA-122 supports robust innate immunity in hepatocytes by targeting the RTKs/STAT3 signaling pathway. *Elife* **8**, (2019).
377. Shi, L. *et al.* The contribution of miR-122 to the innate immunity by regulating toll-like receptor 4 in hepatoma cells. *BMC Gastroenterol* **19**, 130 (2019).

378. Wang, Y. *et al.* Injured liver-released miRNA-122 elicits acute pulmonary inflammation via activating alveolar macrophage TLR7 signaling pathway. *Proc Natl Acad Sci U S A* **116**, 6162–6171 (2019).
379. Qi, Y. *et al.* Dysregulated microRNA expression in serum of non-vaccinated children with varicella. *Viruses* **6**, 1823–36 (2014).
380. Wang, H.-J. *et al.* Serum miR-122 levels are related to coagulation disorders in sepsis patients. *Clin Chem Lab Med* **52**, 927–33 (2014).
381. Wang, H., Yu, B., Deng, J., Jin, Y. & Xie, L. Serum miR-122 correlates with short-term mortality in sepsis patients. *Crit Care* **18**, 704 (2014).
382. Collison, A. M. *et al.* miR-122 promotes virus-induced lung disease by targeting SOCS1. *JCI Insight* **6**, (2021).
383. Feng, Y. *et al.* Extracellular MicroRNAs Induce Potent Innate Immune Responses via TLR7/MyD88-Dependent Mechanisms. *J Immunol* **199**, 2106–2117 (2017).
384. Tesovnik, T. *et al.* Extracellular Vesicles Derived Human-miRNAs Modulate the Immune System in Type 1 Diabetes. *Front Cell Dev Biol* **8**, 202 (2020).
385. Nakamura, M. *et al.* MicroRNA-122 Inhibits the Production of Inflammatory Cytokines by Targeting the PKR Activator PACT in Human Hepatic Stellate Cells. *PLoS One* **10**, (2015).
386. Zhou, Z., Xu, M. J. & Gao, B. Hepatocytes: a key cell type for innate immunity. *Cellular & Molecular Immunology 2016 13:3* **13**, 301–315 (2015).

387. Li, Y., Tan, J., Miao, Y. & Zhang, Q. MicroRNA in extracellular vesicles regulates inflammation through macrophages under hypoxia. *Cell Death Discovery 2021 7:1* **7**, 1–12 (2021).

388. Drury, R. E., O'Connor, D. & Pollard, A. J. The Clinical Application of MicroRNAs in Infectious Disease. *Front Immunol* **8**, 1182 (2017).

389. Chiou, N. T., Kageyama, R. & Ansel, K. M. Selective Export into Extracellular Vesicles and Function of tRNA Fragments during T Cell Activation. *Cell Rep* **25**, 3356-3370.e4 (2018).

390. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology 2008 26:12* **26**, 1367–1372 (2008).

391. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods 2019 17:1* **17**, 41–44 (2019).

392. Dunkelberger, J. R. & Song, W. C. Complement and its role in innate and adaptive immune responses. *Cell Research 2010 20:1* **20**, 34–50 (2009).

393. Agrawal, A., Singh, P. P., Bottazzi, B., Garlanda, C. & Mantovani, A. Pattern Recognition by Pentraxins. *Adv Exp Med Biol* **653**, 98 (2009).

394. Lu, J., Mold, C., Du Clos, T. W. & Sun, P. D. Pentraxins and Fc receptor-mediated immune responses. *Front Immunol* **9**, 2607 (2018).

395. Stephensen, C. B. VITAMIN A, INFECTION, AND IMMUNE FUNCTION*. *http://dx.doi.org/10.1146/annurev.nutr.21.1.167* **21**, 167–192 (2003).

396. Raverdeau, M. & Mills, K. H. G. Modulation of T Cell and Innate Immune Responses by Retinoic Acid. *J Immunol References* **192**, 2953–2958 (2022).

397. Vellozo, N. S. *et al.* All- Trans Retinoic Acid Promotes an M1- to M2-Phenotype Shift and Inhibits Macrophage-Mediated Immunity to Leishmania major. *Front Immunol* **8**, (2017).

398. Han, R. Plasma lipoproteins are important components of the immune system. *Microbiol Immunol* **54**, 246–253 (2010).

399. Pirillo, A., Catapano, A. L. & Norata, G. D. HDL in infectious diseases and sepsis. *Handb Exp Pharmacol* **224**, 483–508 (2015).

400. Tsai, M. Y. *et al.* Effect of influenza vaccine on markers of inflammation and lipid profile. *Journal of Laboratory and Clinical Medicine* **145**, 323–327 (2005).

401. Fda & Cber. Package Insert and Patient Information Sheet - Fluzone Quadrivalent.

402. Herberg, J. A. *et al.* Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children. *JAMA* **316**, 835–45.

403. Sanchez, J. *et al.* Immune Profiles Identification by Vaccinomics After MVA Immunization in Randomized Clinical Study. *Front Immunol* **11**, 2866 (2020).

404. Bergamaschi, C. *et al.* Systemic IL-15, IFN-γ, and IP-10/CXCL10 signature associated with effective immune response to SARS-CoV-2 in BNT162b2 mRNA vaccine recipients. *Cell Rep* **36**, 109504 (2021).

405. Korczowski, B. Procalcitonin and C-reactive protein in vaccination-associated adverse reactions [3]. *Pediatric Infectious Disease Journal* **23**, 283 (2004).

406. Paget, S. BestBets: Does primary immunisation cause C-Reactive Protein to rise in premature infants? *BestBets* https://bestbets.org/bets/bet.php?id=1637 (2008).

407. Andersen-Nissen, E. *et al.* Innate immune signatures to a partially-efficacious HIV vaccine predict correlates of HIV-1 infection risk. *PLoS Pathog* **17**, (2021).

408. Zak, D. E. *et al.* Merck Ad5/HIV induces broad innate immune activation that predicts $CD8^+$ T-cell responses but is attenuated by preexisting Ad5 immunity. *Proc Natl Acad Sci U S A* **109**, E3503–E3512 (2012).

409. Querec, T. D. *et al.* Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nature Immunology 2008 10:1* **10**, 116–125 (2008).

410. Weiner, J. *et al.* Characterization of potential biomarkers of reactogenicity of licensed antiviral vaccines: randomized controlled clinical trials conducted by the BIOVACSAFE consortium. *Sci Rep* **9**, (2019).

411. Sobolev, O. *et al.* Adjuvanted influenza-H1N1 vaccination reveals lymphoid signatures of age-dependent early responses and of clinical adverse events. *Nat Immunol* **17**, 204–213 (2016).

412. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

413. Kern, F. *et al.* miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems. *Nucleic Acids Res* **48**, W521–W528 (2020).
414. Brad Efron & R. Tibshirani. CRAN - Package GSA. *GSA: Gene Set Analysis.* Preprint at https://CRAN.R-project.org/package=GSA (2022).
415. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605–D612 (2021).
416. Rohart, F., Gautier, B., Singh, A. & le Cao, K.-A. mixOmics: an R package for 'omics feature selection and multiple data integration. *bioRxiv* (2017).
417. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. (2006).
418. Abugessaisa, I. *et al.* FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. *Sci Data* **4**, (2017).
419. Dataset - ENCODE Transcription Factor Targets. https://maayanlab.cloud/Harmonizome/dataset/ENCODE+Transcription+Factor+Targets.
420. Tong, Z., Cui, Q., Wang, J. & Zhou, Y. TransmiR v2.0: an updated transcription factor-microRNA regulation database. *Nucleic Acids Res* **47**, D253–D258 (2019).
421. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature Communications 2021 12:1* **12**, 1–20 (2021).
422. Home < IMEx. https://www.imexconsortium.org/.

423. Fukuhara, T. *et al.* Expression of microRNA miR-122 facilitates an efficient replication in nonhepatic cells upon infection with hepatitis C virus. *J Virol* **86**, 7918–33 (2012).

424. Chen, C. L. *et al.* Baculovirus-mediated miRNA regulation to suppress hepatocellular carcinoma tumorigenicity and metastasis. *Mol Ther* **23**, 79–88 (2015).

425. Werner, A. *et al.* Contribution of natural antisense transcription to an endogenous siRNA signature in human cells. *BMC Genomics* **15**, (2014).

426. Gossger, N. *et al.* Immunogenicity and tolerability of recombinant serogroup B meningococcal vaccine administered with or without routine infant vaccinations according to different immunization schedules: a randomized controlled trial. *JAMA* **307**, 573–582 (2012).

427. Wang, N. *et al.* Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. *Nature Communications 2019 10:1* **10**, 1–12 (2019).

428. Georgountzou, A. & Papadopoulos, N. G. Postnatal innate immune development: From birth to adulthood. *Front Immunol* **8**, 957 (2017).

429. Yasui, K. *et al.* An increase in polymorphonuclear leucocyte chemotaxis accompanied by a change in the membrane fluidity with age during childhood. *Clin Exp Immunol* **81**, 156 (1990).

430. Belderbos, M. E. *et al.* Skewed pattern of Toll-like receptor 4-mediated cytokine production in human neonatal blood: low LPS-induced IL-12p70 and high IL-10 persist throughout the first month of life. *Clin Immunol* **133**, 228–237 (2009).

431. Nguyen, M. *et al.* Acquisition of Adult-Like TLR4 and TLR9 Responses during the First Year of Life. *PLoS One* **5**, e10407 (2010).
432. Yerkovich, S. T. *et al.* Postnatal development of monocyte cytokine responses to bacterial lipopolysaccharide. *Pediatr Res* **62**, 547–552 (2007).
433. Corbett, N. P. *et al.* Ontogeny of Toll-Like Receptor Mediated Cytokine Responses of Human Blood Mononuclear Cells. *PLoS One* **5**, e15041 (2010).
434. Dicker, D. *et al.* Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1684–1735 (2018).
435. Kollmann, T. R., Kampmann, B., Mazmanian, S. K., Marchant, A. & Levy, O. Protecting the Newborn and Young Infant from Infectious Diseases: Lessons from Immune Ontogeny. *Immunity* **46**, 350–363 (2017).
436. Drury, R. E., O'Connor, D. & Pollard, A. J. The Clinical Application of MicroRNAs in Infectious Disease. *Front Immunol* **8**, 1182 (2017).
437. Ö, E. & Ş, K. Approach to Fever in the Returning Traveler. *N Engl J Med* **376**, 1797–1798 (2017).
438. Jensenius, M. *et al.* Acute and Potentially Life-Threatening Tropical Diseases in Western Travelers—A GeoSentinel Multicenter Study, 1996–2011. *Am J Trop Med Hyg* **88**, 397 (2013).
439. Stanaway, J. D. *et al.* The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect Dis* **19**, 369–381 (2019).

440. WHO. Typhoid. *World Health Organisation Fact Sheet* https://www.who.int/news-room/fact-sheets/detail/typhoid.
441. Levine, M. M. Typhoid fever. *Bacterial Infections of Humans: Epidemiology and Control* 913–937 (2009) doi:10.1007/978-0-387-09843-2_43/COVER.
442. JS, G. *et al.* Salmonella chronic carriage: epidemiology, diagnosis, and gallbladder persistence. *Trends Microbiol* **22**, 648–655 (2014).
443. MM, L., RE, B. & C, L. Precise estimation of the numbers of chronic carriers of Salmonella typhi in Santiago, Chile, an endemic area. *J Infect Dis* **146**, 724–726 (1982).
444. Basnyat, B., Qamar, F. N., Rupali, P., Ahmed, T. & Parry, C. M. Enteric fever. *BMJ* **372**, (2021).
445. Treatment Guidelines of Typhoid and Paratyphoid (Enteric fever) in Adults.
446. Typhoid Management Guidelines – 2019 – MMIDSP. https://www.mmidsp.com/typhoid-management-guidelines-2019/.
447. de Jong, H. K., Parry, C. M., van der Poll, T. & Wiersinga, W. J. Host–Pathogen Interaction in Invasive Salmonellosis. *PLoS Pathog* **8**, e1002933 (2012).
448. RP, W. *et al.* The Vi capsular polysaccharide prevents complement receptor 3-mediated clearance of Salmonella enterica serotype Typhi. *Infect Immun* **79**, 830–837 (2011).
449. RB, H., chemotherapy, F. M. J.-A. agents and & 1963, undefined. ASYMPTOMATIC TYPHOIDAL BACTEREMIA IN VOLUNTEERS. *europepmc.org*.

450. M, A., NJ, S., S, B., AJ, P. & VE, P. The Relationship Between Blood Sample Volume and Diagnostic Sensitivity of Blood Culture for Typhoid and Paratyphoid Fever: A Systematic Review and Meta-Analysis. *J Infect Dis* **218**, S255–S267 (2018).
451. Näsström, E. *et al.* Reproducible diagnostic metabolites in plasma from typhoid fever patients in Asia and Africa. *Elife* **6**, (2017).
452. Blohmke, C. J. *et al.* Diagnostic host gene signature for distinguishing enteric fever from other febrile diseases. *EMBO Mol Med* **11**, e10431 (2019).
453. Britto, C. D., Wong, V. K., Dougan, G. & Pollard, A. J. A systematic review of antimicrobial resistance in Salmonella enterica serovar Typhi, the etiological agent of typhoid. *PLoS Negl Trop Dis* **12**, (2018).
454. Klemm, E. J. *et al.* Emergence of an Extensively Drug-Resistant Salmonella enterica Serovar Typhi Clone Harboring a Promiscuous Plasmid Encoding Resistance to Fluoroquinolones and Third-Generation Cephalosporins. *mBio* **9**, (2018).
455. Dahora, L. C. *et al.* IgA and IgG1 Specific to Vi Polysaccharide of Salmonella Typhi Correlate With Protection Status in a Typhoid Fever Controlled Human Infection Model. *Front Immunol* **0**, 2582 (2019).
456. Cross, D. Signatures of cellular immunity after vaccination and challenge that associate with clinical outcome of Salmonella Typhi infection.
457. Nakaya, H. I. *et al.* Systems Analysis of Immunity to Influenza Vaccination across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures. *Immunity* **43**, 1186–1198 (2015).

458. Jin, C. Immunogenicity and efficacy of conjugated and unconjugated Vi-polysaccharide vaccines in the prevention of typhoid fever.
459. Jin, C. *et al.* Vi-specific serological correlates of protection for typhoid fever. *J Exp Med* **218**, (2021).
460. Jin, C. Immunogenicity and efficacy of conjugated and unconjugated Vi-polysaccharide vaccines in the prevention of typhoid fever. (2018).
461. Dahora, L. C. *et al.* IgA and IgG1 Specific to Vi Polysaccharide of Salmonella Typhi Correlate With Protection Status in a Typhoid Fever Controlled Human Infection Model. *Front Immunol* **0**, 2582 (2019).
462. Tawfik, D. M. *et al.* Comparison of host immune responses to LPS in human using an immune profiling panel, in vivo endotoxemia versus ex vivo stimulation. *Sci Rep* **10**, (2020).
463. Lai, N. Y. *et al.* Gut-innervating nociceptor neurons regulate Peyer's Patch Microfold cells and SFB levels to mediate Salmonella host defense. *Cell* **180**, 33 (2020).
464. Chen, Y. C. *et al.* Next generation sequencing reveals miR-431-3p/miR-1303 as immune-regulating microRNAs for active tuberculosis. *J Infect* **85**, 519–533 (2022).
465. Schulte, L. N., Eulalio, A., Mollenkopf, H. J., Reinhardt, R. & Vogel, J. Analysis of the host microRNA response to Salmonella uncovers the control of major cytokines by the let-7 family. *EMBO J* **30**, 1977 (2011).
466. Maudet, C. *et al.* Functional high-throughput screening identifies the miR-15 microRNA family as cellular restriction factors for Salmonella infection. *Nature Communications 2014 5:1* **5**, 1–13 (2014).

467. Aguilar, C. *et al.* Functional screenings reveal different requirements for host microRNAs in Salmonella and Shigella infection. *Nature Microbiology 2019 5:1* **5**, 192–205 (2019).
468. Aguilar, C. *et al.* Reprogramming of microRNA expression via E2F1 downregulation promotes Salmonella infection both in infected and bystander cells. *Nat Commun* **12**, (2021).
469. M, S. *et al.* Phase 3 Efficacy Analysis of a Typhoid Conjugate Vaccine Trial in Nepal. *N Engl J Med* **381**, 2209–2218 (2019).
470. Yousafzai, M. T. *et al.* Effectiveness of typhoid conjugate vaccine against culture-confirmed Salmonella enterica serotype Typhi in an extensively drug-resistant outbreak setting of Hyderabad, Pakistan: a cohort study. *Lancet Glob Health* **9**, e1154–e1162 (2021).
471. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
472. Weiss, S. R. & Navas-Martin, S. Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. *Microbiology and Molecular Biology Reviews* **69**, 635 (2005).
473. Lytras, S. *et al.* Exploring the Natural Origins of SARS-CoV-2 in the Light of Recombination. *Genome Biol Evol* **14**, (2022).
474. Transmission of SARS-CoV-2: implications for infection prevention precautions. https://www.who.int/news-

room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.

475. Wilmes, P. *et al.* SARS-CoV-2 transmission risk from asymptomatic carriers: Results from a mass screening programme in Luxembourg. *The Lancet regional health. Europe* **4**, (2021).

476. Iacobucci, G. Covid-19: Runny nose, headache, and fatigue are commonest symptoms of omicron, early data show. *BMJ* **375**, n3103 (2021).

477. Menni, C. *et al.* Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine 2020 26:7* **26**, 1037–1040 (2020).

478. Docherty, A. B. *et al.* Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* **369**, (2020).

479. Raleigh, V. S. Ethnic differences in covid-19 death rates. *BMJ* **376**, (2022).

480. Woodward, M. *et al.* Social deprivation as a risk factor for COVID-19 mortality among women and men in the UK Biobank: nature of risk and context suggests that social interventions are essential to mitigate the effects of future pandemics. *J Epidemiol Community Health* **75**, 1050–1055 (2021).

481. COVID-19 Data Explorer - Our World in Data. https://ourworldindata.org/explorers/coronavirus-data-explorer?facet=none&uniformYAxis=0&Metric=Confirmed+deaths&Interval=Cumulative&Relative+to+Population=false&Color+by+test+positivity=false&country=~OWID_WRL.

482. The impact of COVID-19 on child abuse in the UK | NSPCC Learning. https://learning.nspcc.org.uk/research-resources/2020/social-isolation-risk-child-abuse-during-and-after-coronavirus-pandemic.

483. Pieh, C. *et al.* Mental Health During COVID-19 Lockdown in the United Kingdom. *Psychosom Med* **83**, 328–337 (2021).

484. Badarinza, C., Balasubramaniam, V. & Ramadorai, T. The Household Finance Landscape in Emerging Economies. *Annual Review of Financial Economics* **11**, 109–129 (2019).

485. Folegatti, P. M. *et al.* Safety and immunogenicity of a candidate Middle East respiratory syndrome coronavirus viral-vectored vaccine: a dose-escalation, open-label, non-randomised, uncontrolled, phase 1 trial. *Lancet Infect Dis* **20**, 816–826 (2020).

486. Airfinity | AstraZeneca and Pfizer/BioNTech saved over 12 million lives in the first year of vaccination. https://www.airfinity.com/articles/astrazeneca-and-pfizer-biontech-saved-over-12-million-lives-in-the-first.

487. Honda-Okubo, Y. *et al.* Severe Acute Respiratory Syndrome-Associated Coronavirus Vaccines Formulated with Delta Inulin Adjuvants Provide Enhanced Protection while Ameliorating Lung Eosinophilic Immunopathology. *J Virol* **89**, 2995–3007 (2015).

488. Liu, L. *et al.* Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. *JCI Insight* **4**, (2019).

489. Ewer, K. J. *et al.* T cell and antibody responses induced by a single dose of ChAdOx1 nCoV-19 (AZD1222) vaccine in a phase 1/2 clinical trial. *Nature Medicine 2020 27:2* **27**, 270–278 (2020).

490. van Doremalen, N. *et al.* ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature 2020 586:7830* **586**, 578–582 (2020).

491. Li, C., Hu, X., Li, L. & Li, J. hui. Differential microRNA expression in the peripheral blood from human patients with COVID-19. *J Clin Lab Anal* **34**, (2020).

492. Togami, Y. *et al.* Significance of interferon signaling based on mRNA-microRNA integration and plasma protein analyses in critically ill COVID-19 patients. *Mol Ther Nucleic Acids* **29**, 343–353 (2022).

493. Najafipour, R. *et al.* Screening for differentially expressed microRNAs in BALF and blood samples of infected COVID-19 ARDS patients by small RNA deep sequencing. *J Clin Lab Anal* (2022) doi:10.1002/JCLA.24672.

494. Chen, Y.-M. *et al.* Blood molecular markers associated with COVID-19 immunopathology and multi-organ damage. *EMBO J* **39**, e105896 (2020).

495. Duecker, R. P. *et al.* The MiR-320 Family Is Strongly Downregulated in Patients with COVID-19 Induced Severe Respiratory Failure. *Int J Mol Sci* **22**, (2021).

496. Liu, X. *et al.* SARS-CoV-2 causes a significant stress response mediated by small RNAs in the blood of COVID-19 patients. *Mol Ther Nucleic Acids* **27**, 751–762 (2022).

497. Tang, H. *et al.* The noncoding and coding transcriptional landscape of the peripheral immune response in patients with COVID-19. *Clin Transl Med* **10**, (2020).

498. Pollet, K. *et al.* Host miRNAs as biomarkers of SARS-CoV-2 infection: a critical review. *Sensors & Diagnostics* (2022) doi:10.1039/D2SD00140C.

499. Kokkotis, G. *et al.* Systematic review with meta-analysis: COVID-19 outcomes in patients receiving anti-TNF treatments. *Aliment Pharmacol Ther* **55**, 154–167 (2022).

500. Spinelli, S. v. *et al.* miR-30c is specifically repressed in patients with active pulmonary tuberculosis. *Tuberculosis* **105**, 73–79 (2017).

501. Larabi, A., Dalmasso, G., Delmas, J., Barnich, N. & Nguyen, H. T. T. Exosomes transfer miRNAs from cell-to-cell to inhibit autophagy during infection with Crohn's disease-associated adherent-invasive E. coli. *Gut Microbes* **11**, 1677–1694 (2020).

502. Jin, Y. *et al.* MiR-30c-5p mediates inflammatory responses and promotes microglia survival by targeting eIF2α during Cryptococcus neoformans infection. *Microb Pathog* **141**, (2020).

503. Polonio, C. M. *et al.* microRNAs Control Antiviral Immune Response, Cell Death and Chemotaxis Pathways in Human Neuronal Precursor Cells (NPCs) during Zika Virus Infection. *Int J Mol Sci* **23**, (2022).

504. Cai, Y., Chen, H., Jin, L., You, Y. & Shen, J. STAT3-dependent transactivation of miRNA genes following Toxoplasma gondii infection in macrophage. *Parasit Vectors* **6**, (2013).

505. Li, J. *et al.* The Reaction Pathway of miR-30c-5p Activates Lipopolysaccharide Promoting the Course of Traumatic and Hemorrhagic Shock Acute Lung Injury. *Biomed Res Int* **2022**, (2022).
506. Harper, K. L. *et al.* Dysregulation of the miR-30c/DLL4 axis by circHIPK3 is essential for KSHV lytic replication. *EMBO Rep* **23**, (2022).
507. Siniscalchi, C., di Palo, A., Russo, A. & Potenza, N. Human MicroRNAs Interacting With SARS-CoV-2 RNA Sequences: Computational Analysis and Experimental Target Validation. *Front Genet* **12**, 760 (2021).
508. Kaur, H. *et al.* Screening and identification of potential novel biomarker for diagnosis of complicated Plasmodium vivax malaria. *J Transl Med* **16**, (2018).
509. Farr, R. J. *et al.* Machine Learning Identifies Cellular and Exosomal MicroRNA Signatures of Lyssavirus Infection in Human Stem Cell-Derived Neurons. *Front Cell Infect Microbiol* **11**, (2021).
510. de la Rosa, I. A. *et al.* Impaired microRNA processing in neutrophils from rheumatoid arthritis patients confers their pathogenic profile. Modulation by biological therapies. *Haematologica* **105**, 2250 (2020).
511. Bronevetsky, Y. *et al.* T cell activation induces proteasomal degradation of Argonaute and rapid remodeling of the microRNA repertoire. *J Exp Med* **210**, 417–432 (2013).
512. Wilson, J. C. *et al.* Integrated miRNA/cytokine/chemokine profiling reveals severity-associated step changes and principal correlates of fatality in COVID-19. *iScience* **25**, (2022).

513. Zheleznyakova, G. Y. *et al.* Small noncoding RNA profiling across cellular and biofluid compartments and their implications for multiple sclerosis immunopathology. *Proc Natl Acad Sci U S A* **118**, (2021).

514. Zhuravlev, E. *et al.* Analysis of Expression Pattern of snoRNAs in Human Cells A549 Infected by Influenza A Virus. *International Journal of Molecular Sciences 2022, Vol. 23, Page 13666* **23**, 13666 (2022).

515. Burel, J. G. *et al.* Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. *Front Immunol* **10**, 221 (2019).

516. Blood ATlas Consortium, M. *et al.* Authors COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium Correspondence In brief ll Resource A blood atlas of COVID-19 defines hallmarks of disease severity and specificity COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium. *Cell* **185**, 916–938 (2022).

517. Faes, C. *et al.* Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients. *Int J Environ Res Public Health* **17**, 1–18 (2020).

518. Alaa, A., Qian, Z., Rashbass, J., Benger, J. & van der Schaar, M. Retrospective cohort study of admission timing and mortality following COVID-19 infection in England. *BMJ Open* **10**, e042712 (2020).

519. Ismail, N. *et al.* Macrophage microvesicles induce macrophage differentiation and miR-223 transfer. *Blood* **121**, 984–995 (2013).

520. Wang, J. *et al.* Serum miR-146a and miR-223 as potential new biomarkers for sepsis. *Biochem Biophys Res Commun* **394**, 184–8 (2010).

521. Wang, X. *et al.* Exosomal miR-223 Contributes to Mesenchymal Stem Cell-Elicited Cardioprotection in Polymicrobial Sepsis. *Scientific Reports 2015 5:1* **5**, 1–16 (2015).
522. Neudecker, V. *et al.* Neutrophil transfer of miR-223 to lung epithelial cells dampens acute lung injury in mice. *Sci Transl Med* **9**, eaah5360 (2017).
523. Wu, N. *et al.* miR-223 inhibits dengue virus replication by negatively regulating the microtubule-destabilizing protein STMN1 in EAhy926 cells. *Microbes and infection / Institut Pasteur* **16**, 911–22 (2014).
524. Benz, F. *et al.* Circulating microRNA-223 serum levels do not predict sepsis or survival in patients with critical illness. *Dis Markers* **2015**, 384208 (2015).
525. Murray, D. D. *et al.* Circulating microRNAs in Sera Correlate with Soluble Biomarkers of Immune Activation but Do Not Predict Mortality in ART Treated Individuals with HIV-1 Infection: A Case Control Study. *PLoS One* **10**, e0139981 (2015).
526. Wong, J. J.-L. *et al.* RBM3 regulates temperature sensitive miR-142-5p and miR-143 (thermomiRs), which target immune genes and control fever. *Nucleic Acids Res* **44**, 2888–2897 (2016).
527. Haralambieva, I. H. *et al.* Differential miRNA expression in B cells is associated with inter-individual differences in humoral immune response to measles vaccination. *PLoS One* **13**, e0191812 (2018).
528. Stelma, F. *et al.* Immune phenotype and function of natural killer and T cells in chronic hepatitis C patients who received a single dose of anti-MicroRNA-122, RG-101. *Hepatology* (2017) doi:10.1002/hep.29148.

529. Wright, C. *et al.* Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics* **20**, 1–21 (2019).
530. Huang, R. S. *et al.* Population differences in microRNA expression and biological implications. *RNA Biol* **8**, 692 (2011).
531. Flowers, E., Kanaya, A. M., Zhang, L. & Aouizerat, B. E. The Role of Racial and Ethnic Factors in MicroRNA Expression and Risk for Type 2 Diabetes. *Front Genet* **13**, 638 (2022).
532. De Buck, M. *et al.* The cytokine-serum amyloid A-chemokine network. *Cytokine Growth Factor Rev* **30**, 55 (2016).
533. Kruzel, M. L., Zimecki, M. & Actor, J. K. Lactoferrin in a Context of Inflammation-Induced Pathology. *Front Immunol* **8**, (2017).
534. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–D452 (2015).
535. Razanakolona, M. *et al.* Anti-inflammatory Activity of the Protein Z-Dependent Protease Inhibitor. *TH Open* **5**, e220 (2021).
536. Bonacina, F. *et al.* Myeloid apolipoprotein E controls dendritic cell antigen presentation and T cell activation. *Nature Communications 2018 9:1* **9**, 1–15 (2018).
537. Sack, G. H. Serum amyloid A – a review. *Molecular Medicine 2018 24:1* **24**, 1–27 (2018).
538. Huntoon, K. M. *et al.* The acute phase protein haptoglobin regulates host immunity. *J Leukoc Biol* **84**, 170 (2008).

539. Wang, Y., Kinzie, E., Berger, F. G., Lim, S.-K. & Baumann, H. Redox Report Communications in Free Radical Research Haptoglobin, an inflammation-inducible plasma protein. (2013) doi:10.1179/135100001101536580.

540. Arredouani, M., Matthys, P., Kasran, A., Baumann, H. & Ceuppens, J. L. Haptoglobin and the Th1/Th2 balance: hints from in vitro and in vivo studies. *Redox Rep* **6**, 369–371 (2001).

541. Luyendyk, J. P., Schoenecker, J. G. & Flick, M. J. The multifaceted role of fibrinogen in tissue injury and inflammation. *Blood* **133**, 511 (2019).

542. Dupont, H. L. *et al.* Studies of immunity in typhoid fever: Protection Induced by Killed Oral Antigens or by Primary Infection*. *Bull World Health Organ* **44**, 667 (1971).

543. Gibani, M. M. *et al.* Homologous and heterologous re-challenge with Salmonella Typhi and Salmonella Paratyphi A in a randomised controlled human infection model. *PLoS Negl Trop Dis* **14**, e0008783 (2020).

544. Hosoglu, S. *et al.* A simple validated prediction rule to diagnose typhoid fever in Turkey. *Trans R Soc Trop Med Hyg* **100**, 1068–1074 (2006).

545. L, W., S, M. & CM, P. Rapid diagnostic tests for typhoid and paratyphoid (enteric) fever. *Cochrane Database Syst Rev* **5**, (2017).

546. LA, O. & AL, K. Widal agglutination test - 100 years later: still plagued by controversy. *Postgrad Med J* **76**, 80–84 (2000).

547. Giannella, R. A. Salmonella. *Medical Microbiology* (1996).

548. J. Barton, A., Hill, J., J. Blohmke, C. & J. Pollard, A. Host restriction, pathogenesis and chronic carriage of typhoidal Salmonella. *FEMS Microbiol Rev* **014**, 1–12 (2021).

549. M, Y. *et al.* Allelic variation contributes to bacterial host specificity. *Nat Commun* **6**, (2015).

550. Gonzales, A. M., Wilde, S. & Roland, K. L. New Insights into the Roles of Long Polar Fimbriae and Stg Fimbriae in Salmonella Interactions with Enterocytes and M Cells. *Infect Immun* **85**, (2017).

551. Juel, H. B. *et al.* Salmonella Typhi bactericidal antibodies reduce disease severity but do not protect against typhoid fever in a controlled human infection model. *Front Immunol* **8**, (2018).

552. Gibani, M. M. *et al.* Investigation of the role of typhoid toxin in acute typhoid fever in a human challenge model. *Nature Medicine 2019 25:7* **25**, 1082–1088 (2019).

553. Adcox, H. E. *et al.* Salmonella extracellular matrix components influence biofilm formation and gallbladder colonization. *Infect Immun* **84**, 3243–3251 (2016).

554. Marathe, S. A., Lahiri, A., Negi, V. D. & Chakravortty, D. Typhoid fever & vaccine development: a partially answered question. *Indian J Med Res* **135**, 161 (2012).

555. Syed, K. A. *et al.* Review on the Recent Advances on Typhoid Vaccine Development and Challenges Ahead. *Clinical Infectious Diseases* **71**, S141–S150 (2020).

556. Bodhidatta, L., Taylor, D. N., Thisyakom, U. & Echeverria, P. Control of Typhoid Fever in Bangkok, Thailand, by Annual Immunization of Schoolchildren with Parenteral Typhoid Vaccine. *Rev Infect Dis* **9**, 841–845 (1987).

557. Cross, D. Signatures of cellular immunity after vaccination and challenge that associate with clinical outcome of Salmonella Typhi infection.

558. Jin, C. *et al.* Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of Salmonella Typhi: a randomised controlled, phase 2b trial. *The Lancet* (2017) doi:10.1016/S0140-6736(17)32149-9.

559. Jin, C. Immunogenicity and efficacy of conjugated and unconjugated Vi-polysaccharide vaccines in the prevention of typhoid fever. (2018).

560. Saad, N. J. *et al.* Seasonal dynamics of typhoid and paratyphoid fever. *Scientific Reports 2018 8:1* **8**, 1–9 (2018).

561. Voysey, M. & Pollard, A. J. Seroefficacy of Vi Polysaccharide–Tetanus Toxoid Typhoid Conjugate Vaccine (Typbar TCV). *Clinical Infectious Diseases* **67**, 18–24 (2018).

562. Jin, C. *et al.* Vi-specific serological correlates of protection for typhoid fever. *J Exp Med* **218**, (2021).

563. CS, W. *et al.* An outpatient, ambulant-design, controlled human infection model using escalating doses of Salmonella Typhi challenge delivered in sodium bicarbonate solution. *Clin Infect Dis* **58**, 1230–1240 (2014).

564. Szu, S. C., Klugman, K. P. & Hunt, S. Re-examination of immune response and estimation of anti-Vi IgG protective threshold against typhoid fever-

based on the efficacy trial of Vi conjugate in young children. *Vaccine* **32**, 2359 (2014).

565. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* **6**, e21800 (2011).
566. Voysey, M. *et al.* Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *The Lancet* **397**, 99–111 (2021).
567. Bernal, J. L. *et al.* Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ* **373**, (2021).
568. Coronavirus Vaccines-summary of Yellow Card reporting. (2022).